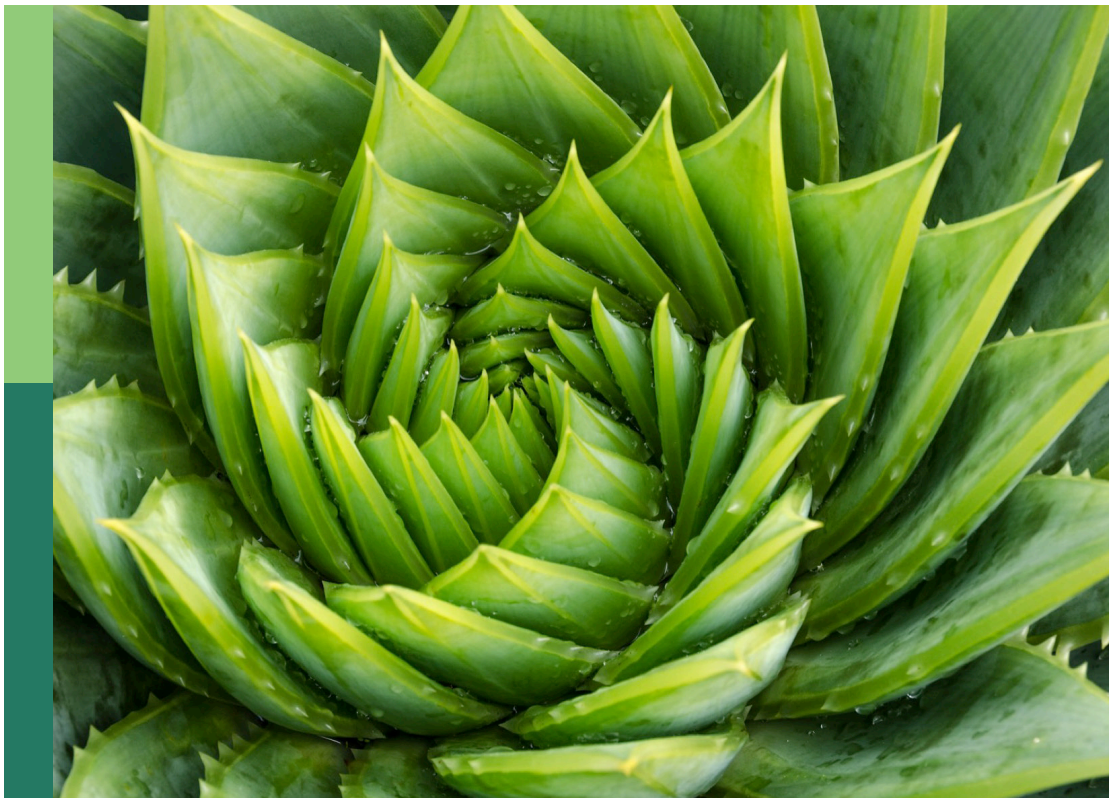
frontiers | Research Topics

Leveraging phenotyping and crop modeling in smart agriculture

Edited by
Ting Sun, Wenyu Zhang, Syed Tahir Ata-Ul-Karim, Yuntao Ma and Liujun Xiao

Published in
Frontiers in Plant Science



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-6619-0
DOI 10.3389/978-2-8325-6619-0

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Leveraging phenotyping and crop modeling in smart agriculture

Topic editors

Ting Sun — Jiangsu Academy of Agricultural Sciences Wuxi Branch, China
Wenyu Zhang — Jiangsu Academy of Agricultural Sciences Wuxi Branch, China
Syed Tahir Ata-Ul-Karim — Aarhus University, Denmark
Yuntao Ma — China Agricultural University, China
Liujun Xiao — Nanjing Agricultural University, China

Citation

Sun, T., Zhang, W., Ata-Ul-Karim, S. T., Ma, Y., Xiao, L., eds. (2025). *Leveraging phenotyping and crop modeling in smart agriculture*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6619-0

Table of contents

06 Editorial: Leveraging phenotyping and crop modeling in smart agriculture
Ting Sun, Liujun Xiao, Syed Tahir Ata-Ul-Karim, Yuntao Ma and Wenyu Zhang

12 Estimation of the rice aboveground biomass based on the first derivative spectrum and Boruta algorithm
Ying Nian, Xiangxiang Su, Hu Yue, Yongji Zhu, Jun Li, Weiqiang Wang, Yali Sheng, Qiang Ma, Jikai Liu and Xinwei Li

25 Combining features selection strategy and features fusion strategy for SPAD estimation of winter wheat based on UAV multispectral imagery
Xiangxiang Su, Ying Nian, Hiba Shaghaleh, Amar Hamad, Hu Yue, Yongji Zhu, Jun Li, Weiqiang Wang, Hong Wang, Qiang Ma, Jikai Liu, Xinwei Li and Yousef Alhaj Hamoud

43 The grading detection model for fingered citron slices (citrus medica 'fingered') based on YOLOv8-FCS
Lingtao Zhang, Pu Luo, Shaoyun Ding, Tingxuan Li, Kebei Qin and Jiong Mu

59 Improved YOLO-FastestV2 wheat spike detection model based on a multi-stage attention mechanism with a LightFPN detection head
Shunhao Qing, Zhaomei Qiu, Weili Wang, Fei Wang, Xin Jin, Jiangtao Ji, Long Zhao and Yi Shi

72 CTHNet: a network for wheat ear counting with local-global features fusion based on hybrid architecture
Qingqing Hong, Wei Liu, Yue Zhu, Tianyu Ren, Changrong Shi, Zhixin Lu, Yunqin Yang, Ruiting Deng, Jing Qian and Changwei Tan

86 Lightweight tomato ripeness detection algorithm based on the improved RT-DETR
Sen Wang, Huiping Jiang, Jixiang Yang, Xuan Ma, Jiamin Chen, Zhongjie Li and Xingqun Tang

105 Enhancing kiwifruit flower pollination detection through frequency domain feature fusion: a novel approach to agricultural monitoring
Fei Pan, Mengdie Hu, Xuliang Duan, Boda Zhang, Pengjun Xiang, Lan Jia, Xiaoyu Zhao and Dawei He

121 A lightweight grape detection model in natural environments based on an enhanced YOLOv8 framework
Xinyu Wu, Rong Tang, Jiong Mu, Yupeng Niu, Zihan Xu and Ziao Chen

134 Advances in the tea plants phenotyping using hyperspectral imaging technology
Baidong Luo, Hongwei Sun, Leilei Zhang, Fengnong Chen and Kaihua Wu

153 Cotton morphological traits tracking through spatiotemporal registration of terrestrial laser scanning time-series data
Javier Rodriguez-Sanchez, John L. Snider, Kyle Johnsen and Changying Li

176 Establishing a knowledge structure for yield prediction in cereal crops using unmanned aerial vehicles
Ghulam Mustafa, Yuhong Liu, Imran Haider Khan, Sarfraz Hussain, Yuhan Jiang, Jiayuan Liu, Saeed Arshad and Raheel Osman

199 An integrated method for phenotypic analysis of wheat based on multi-view image sequences: from seedling to grain filling stages
Shengxuan Sun, Yeping Zhu, Shengping Liu, Yongkuai Chen, Yihan Zhang and Shijuan Li

215 CFIHL: a variety of chlorophyll a fluorescence transient image datasets of hydroponic lettuce
Yiyu Jiang, Yu Tan, Fang Ji, Daobilige Su, Shuo Wang, Lina Zhang and Qing Zhou

223 FIDMT-GhostNet: a lightweight density estimation model for wheat ear counting
Baohua Yang, Runchao Chen, Zhiwei Gao and Hongbo Zhi

239 Barrier-free tomato fruit selection and location based on optimized semantic segmentation and obstacle perception algorithm
Lingli Zhou, Anqi Hu, Yawen Cheng, Wenxiang Zhang, Bingyuan Zhang, Xinyu Lu, Qian Wu and Ni Ren

257 Identification of sweetpotato virus disease-infected leaves from field images using deep learning
Ziyu Ding, Fanguo Zeng, Haifeng Li, Jianyu Zheng, Junzhi Chen, Biao Chen, Wenshan Zhong, Xuantian Li, Zhangying Wang, Lifei Huang and Xuejun Yue

271 Predicting alfalfa leaf area index by non-linear models and deep learning models
Songtao Yang, Yongqi Ge, Jing Wang, Rui Liu and Li Fu

285 YOLO-SDL: a lightweight wheat grain detection technology based on an improved YOLOv8n model
Zhaomei Qiu, Fei Wang, Weili Wang, Tingting Li, Xin Jin, Shunhao Qing and Yi Shi

297 Estimation of potato leaf area index based on spectral information and Haralick textures from UAV hyperspectral images
Jiejie Fan, Yang Liu, Yiguang Fan, Yihan Yao, Riqiang Chen, Mingbo Bian, Yanpeng Ma, Huifang Wang and Haikuan Feng

313 Mapping rapeseed (*Brassica napus* L.) aboveground biomass in different periods using optical and phenotypic metrics derived from UAV hyperspectral and RGB imagery
Chuanliang Sun, Weixin Zhang, Genping Zhao, Qian Wu, Wanjie Liang, Ni Ren, Hongxin Cao and Lidong Zou

329 Enhancing Nitrogen Nutrition Index estimation in rice using multi-leaf SPAD values and machine learning approaches
Yuan Wang, Peihua Shi, Yinfei Qian, Gui Chen, Jiang Xie, Xianjiao Guan, Weiming Shi and Haitao Xiang

345 Mitigating saturation effects in rice nitrogen estimation using Dualex measurements and machine learning
Peihua Shi, Yuan Wang, Congfei Yin, Kaiqing Fan, Yinfei Qian and Gui Chen

360 Quantitative relationship model between soil profile salinity and soil depth in cotton fields based on data assimilation algorithm: forecasting cotton field yields and profits
Yang Gao, Lin Chang, Mei Zeng, Quanze Hu, Jiaojiao Hui and Qingsong Jiang

376 Development of a vegetation canopy reflectance sensor and its diurnal applicability under clear sky conditions
Naisen Liu, Jingyu Guo, Fuxia Liu, Xuedong Zha, Jing Cao, Yuezhen Chen, Haixia Yan, Chenggong Du, Xuqi Wang, Jiping Li and Yongzhen Zhao

390 Hyperspectral estimation of chlorophyll density in winter wheat using fractional-order derivative combined with machine learning
Chenbo Yang, Meichen Feng, Juan Bai, Hui Sun, Rutian Bi, Lifang Song, Chao Wang, Yu Zhao, Wude Yang, Lujie Xiao, Meijun Zhang and Xiaoyan Song

407 Fruit size prediction of tomato cultivars using machine learning algorithms
Masaaki Takahashi, Yasushi Kawasaki, Hiroki Naito, Unseok Lee and Koichi Yoshi

416 Two-dimensional semantic morphological feature extraction and atlas construction of maize ear leaves
Hongli Song, Weiliang Wen, Ying Zhang, Yanxin Zhao, Xinyu Guo and Chunjiang Zhao

433 A 3D reconstruction platform for complex plants using OB-NeRF
Sixiao Wu, Changhao Hu, Boyuan Tian, Yuan Huang, Shuo Yang, Shanjun Li and Shengyong Xu

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 03 July 2025
DOI 10.3389/fpls.2025.1626622

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Ruslan Kalendar,
University of Helsinki, Finland

*CORRESPONDENCE
Wenyu Zhang
research@wwery.cn

RECEIVED 11 May 2025
ACCEPTED 21 May 2025
PUBLISHED 03 July 2025

CITATION
Sun T, Xiao L, Ata-Ul-Karim ST, Ma Y and Zhang W (2025) Editorial: Leveraging phenotyping and crop modeling in smart agriculture. *Front. Plant Sci.* 16:1626622. doi: 10.3389/fpls.2025.1626622

COPYRIGHT
© 2025 Sun, Xiao, Ata-Ul-Karim, Ma and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Leveraging phenotyping and crop modeling in smart agriculture

Ting Sun[1], Liujun Xiao[2], Syed Tahir Ata-Ul-Karim[3], Yuntao Ma[4] and Wenyu Zhang[1]*

[1]Jiangsu Academy of Agricultural Sciences Wuxi Branch, Wuxi, Jiangsu, China, [2]National Engineering and Technology Center for Information Agriculture, Key Laboratory for Crop System Analysis and Decision Making, Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing Agricultural University, Nanjing, Jiangsu, China, [3]Department of Agroecology, Climate and Water, Aarhus University, Tjele, Denmark, [4]College of Land Science and Technology, China Agricultural University, Beijing, China

KEYWORDS

phenotyping, modeling, smart agriculture, functional-structural plant models, environment, genotype

Editorial on the Research Topic
Leveraging phenotyping and crop modeling in smart agriculture

In recent years, the agricultural sector has witnessed a significant transformation driven by the integration of sensing technologies, big data analytics, and artificial intelligence (Ahmed and Shakoor, 2025). Cutting-edge innovations, notably high-throughput phenotyping and crop modeling, have fundamentally altered our understanding and management of crop systems (Keating and Thorburn, 2018; Yang et al., 2020). In many cases, phenotyping and modeling are closely intertwined: phenotyping provides accurate characterization of plant traits, forming the basis for reliable crop models, while modeling elucidates interactions among phenotypes, genotypes, and the environment, and enables prediction of phenotypic outcomes (Yu et al., 2023; Zhang et al., 2023b). Despite their natural synergy, phenotyping and modeling are still frequently treated as separate domains, limiting their full potential. This Research Topic aims to close that gap by promoting the development of integrated phenotyping-modeling frameworks to advance smart agriculture. The following sections provide a categorized overview of the contributions to this Research Topic (https://www.frontiersin.org/research-topics/62521/leveraging-phenotyping-and-crop-modeling-in-smart-agriculture), highlighting key findings and identifying future directions for this rapidly advancing field.

1 Crop phenotyping

Crop phenotyping, which plays a vital role in gene function analysis, plant breeding, and smart agriculture, can be broadly categorized based on the traits measured. Morphological and structural traits include leaf length, leaf width, leaf area, and leaf angle, while physiological and biological traits encompass chlorophyll content, nitrogen levels, transpiration, and photosynthetic parameters.

1.1 Morphological and structural phenotypes

2D imaging combined with machine vision remains the most widely adopted technique for acquiring plant morphological and structural phenotypes. In this topic, a range of studies have explored deep learning-based approaches tailored for specific plant phenotyping applications, with a particular focus on refining model architectures and technical strategies to enhance detection accuracy, computational efficiency, and adaptability to complex field conditions. Among them, semantic segmentation frameworks have been effectively employed for fine-grained plant disease identification. Ding et al. employed DeepLabV3+ with an integrated Attention Pyramid Fusion (APF) module, achieving rapid and accurate segmentation of sweetpotato virus disease (SPVD) lesions in field images. Building on the success of YOLO-based object detection, several teams developed tailored variants to tackle specific phenotyping challenges. Pan et al. developed KIWI-YOLO, a kiwifruit flower pollination detection model based on YOLOv5, incorporating a frequency domain feature fusion (FDFF) module and a Bi-Level Routing Attention (BRA) mechanism to improve feature focus and detection performance. Qiu et al. introduced YOLO-SDL, which combines the lightweight YOLOv8n framework with a ShuffleNetV2 backbone and integrates depthwise separable convolutions (DWConv) and the large separable kernel attention (LSKA) module in the neck to enhance wheat grain classification while ensuring compactness and computational efficiency. Wu et al. proposed TiGra-YOLOv8, which integrates an Attentional Scale Fusion (ASF) module, Adaptive Training Sample Selection (ATSS), and Wise-IoU loss, while employing channel pruning to optimize model size and inference speed for grapefruit detection in dense orchard environments. Wang et al. presented PDSI-RTDETR for tomato ripeness detection, improving RT-DETR by replacing the Basic_Block with PConv_Block and integrating deformable attention with intra-scale feature interaction. The model also introduces a slimneck-SSFF fusion structure and replaces GIoU with Inner-EIoU loss to accelerate convergence and improve small object detection accuracy. Zhang et al. proposed YOLOv8-FCS for grading fingered citron slices, enhancing YOLOv8n by substituting its backbone with the Fasternet module and redesigning the PAN-FPN structure using BiFPN to improve computational efficiency and multi-scale feature utilization. Qing et al. also contributed an improved YOLO-FastestV2 model for wheat spike detection, incorporating a multi-stage attention mechanism and LightFPN detection head to optimize detection under variable field conditions. Beyond object detection, several studies addressed plant organ counting through CNN-Transformer hybrid models and lightweight convolutional networks. Hong et al. introduced CTHNet, which combines a CSP-based CNN for multi-scale local feature extraction with a Pyramid Pooling Transformer for global context learning, further enhanced by a feature fusion module to improve wheat ear counting performance. Yang et al. developed a wheat ear positioning and counting approach based on FIDMT-GhostNet, leveraging GhostNet for multi-scale feature extraction, a dense upsampling module for improved image resolution, and a local maximum detection strategy to reduce background noise and enhance counting accuracy in dense environments. Complementing these RGB-based imaging approaches, UAV-based hyperspectral imaging has also been utilized for monitoring canopy structural traits. Fan et al. proposed a potato leaf area index (LAI) estimation model by integrating spectral and textural features selected via the Successive Projection Algorithm (SPA) with machine learning, finding that spectral data exhibited higher sensitivity to LAI than Haralick texture features. Collectively, these studies highlight the diverse technical strategies that are accelerating plant phenotyping. Through architectural refinements and targeted module innovations, they offer efficient, accurate, and scalable solutions suited to real-world, high-throughput agricultural applications.

2D image-based methods are inherently constrained by issues such as occlusion, perspective distortion, and the loss of spatial information. In contrast, 3D approaches are capable of capturing the complete spatial geometry and topological structure of plants, offering a richer and more precise foundation for the analysis of complex phenotypic traits. Nevertheless, reconstructing and analyzing plant 3D point clouds remains challenging due to dense occlusions, intricate structural details, complex organ overlaps, and variations in lighting and background conditions. Overcoming these challenges is critical for advancing high-throughput, *in-situ* plant phenotyping and digital twin modeling in agricultural research. Song et al. gathered 1,431 ear leaves from 518 maize inbred lines at the silking stage using a 3D digitizer. Area-preserving 2D leaf models were generated through mesh subdivision and planar parameterization. Eleven semantic features were identified via clustering and correlation analysis. A 2D leaf shape indicator (L2D) and an atlas were developed, allowing precise identification of inbred lines based on 2D leaf shape. Rodriguez-Sanchez et al. developed a spatiotemporal registration approach for time-series terrestrial laser scanning data, enabling continuous 4D monitoring of cotton canopy traits with high spatial accuracy, and used the registered models to track growth dynamics and assess genotype differences throughout the season. Although 3D digitizers, LiDAR, and similar technologies have been widely applied for crop 3D reconstruction, their high costs and limited precision have significantly constrained broader agricultural applications. In recent years, with the decreasing cost of industrial-grade cameras, multi-view stereo (MVS) based 3D reconstruction has emerged as a mainstream solution. Wu et al. developed a fast and accurate 3D reconstruction platform for the mandarin orange based on Object-Based NeRF (OB-NeRF). By integrating optimized camera pose calibration, efficient ray sampling, and exposure adjustment, the platform reconstructs high-quality neural radiance fields from videos within 250 seconds. Sun et al. employed a multi-view imaging platform to capture wheat plant images, generating high-quality point clouds through Structure-from-Motion and Multi-View Stereo (SfM-MVS) using Euclidean clustering, color filtering, and statistical methods. A region-growing algorithm was used for stem and leaf segmentation, though substantial leaf overlap during the tillering, jointing, and booting stages made the process particularly challenging. Plant height, convex hull volume, plant

surface area, and crown area were extracted, enabling a detailed analysis of dynamic changes in wheat throughout its growth cycle. In recent years, ultra-low-altitude UAV-based cross-circling oblique imaging has become a more efficient and cost-effective approach for in-field 3D reconstruction (Fei et al., 2025; Sun et al., 2024). Unlike indoor multi-view imaging systems, 3D phenotyping conducted directly in the field more accurately reflects real-world agricultural conditions and population-level dynamics.

1.2 Physiological and biological phenotypes

Physiological and biological phenotypes are typically assessed rapidly and non-destructively using hyperspectral or multispectral techniques. Yang et al. combined fractional-order derivatives (FOD) with machine learning techniques to estimate chlorophyll density in winter wheat using hyperspectral imagery. Three FOD methods and eight machine learning models were tested with both full-spectrum data and CARS-selected bands. The Riemann-Liouville FOD (RL-FOD) showed superior performance in model construction. The highest accuracy was achieved by combining 0.3-order RL-FOD, CARS-based band selection, and extra-trees regression (ETsR). Su et al. used UAV multispectral sensors to capture winter wheat canopy images, extracting spectral and texture features. Feature selection methods (Boruta and Recursive Feature Elimination) were applied to identify key features, and a feature fusion strategy combined with Support Vector Machine Regression was used to develop the SPAD estimation model. The results indicated that combining NIR spectral features with other bands, along with red and NIR texture features, effectively captured SPAD variations during the reproductive growth stage. Jiang et al. presented a ChlF dataset of hydroponic lettuce seedlings, consisting of transient images captured under different cultural conditions. The effectiveness of the threshold segmentation algorithm and the Deeplabv3+ algorithm for extracting the seedling canopy was compared. Sun et al. utilized UAV hyperspectral and ultra-high-resolution RGB images to derive vegetation indices, texture features, and structural characteristics for estimating rapeseed aboveground biomass. Various models, including deep neural networks, random forests, and support vector regression, were tested with different feature combinations. Models that incorporated all three feature types delivered higher accuracy compared to those using individual feature sets, with deep neural networks consistently outperforming the other algorithms. Luo et al. reviewed the advancements in applying hyperspectral imaging technology to obtain information on tea plant phenotypes, growth conditions, and quality indicators under environmental stress. Wang et al. applied multi-leaf SPAD measurements combined with machine learning to improve nitrogen diagnostics in rice. Integrating SPAD data with models like Random Forest and Extreme Gradient Boosting enhanced the estimation accuracy of Leaf Nitrogen Concentration (LNC) and Nitrogen Nutrition Index (NNI). The second leaf from the top was most important for predicting LNC, while the third leaf was key for NNI. Shi et al. evaluated flavonoid content (Flav) and the Nitrogen Balance Index (NBI), measured by a Dualex sensor, alongside machine learning models for nitrogen status assessment. Data from 15 rice varieties under varying nitrogen rates showed chlorophyll saturation at high nitrogen levels, while Flav and NBI remained reliable. Random Forest and Extreme Gradient Boosting achieved high prediction accuracy, with SHAP analysis identifying NBI and Flav from the top two leaves as critical predictors. In recent years, these technologies have been widely applied to precision farmland management. For example, on farms in Brazil, Castilho Silva et al. (2025) used UAV-based multispectral remote sensing to monitor leaf nitrogen content in maize and applied variable-rate fertilization accordingly. Compared to conventional methods, this approach reduced nitrogen input by 6.6% to 35% without compromising yield.

1.3 Phenotyping equipment

Phenotyping equipment is essential for the precise monitoring of plant traits and environmental growth conditions. Liu et al. developed a portable vegetation canopy reflectance (VCR) sensor for continuous operation throughout the day, featuring optical bands at 710 nm and 870 nm. The sensor was calibrated using an integrating sphere and a solar altitude correction model, with validation against a standard reflectance gray scale board. Field measurements taken at 14 sites using both the VCR sensor and an ASD spectroradiometer showed closely aligned reflectance values. In Bermuda grass measurements, the intra-day reflectance range narrowed and the coefficient of variation decreased after solar altitude correction, demonstrating the sensor's effectiveness for precise vegetation monitoring. Compared to remote sensing, recent developments in flexible sensors enable direct, continuous, and high-resolution monitoring of plant physiological traits and environmental conditions (Zhang et al., 2024). These innovative sensing technologies are poised to significantly enhance phenotyping applications.

2 Crop modeling

While various models for the direct extraction or inversion of crop phenotypes have been explored in the crop phenotyping section, crop modeling in this context specifically refers to growth modeling designed to predict crop development and growth. Depending on the approach, crop growth models may be data-based, incorporating machine learning techniques, or mechanistic, based on process-based simulations of crop physiological processes (Maestrini et al., 2022). In this topic, process-based models are limited, with more researchers focusing on simpler modeling approaches.

2.1 Data-based models

Takahashi et al. proposed a machine learning approach for early prediction of tomato fruit size at harvest, comparing Ridge Regression, Extra Tree Regression, and CatBoost Regression

models. Estimated fruit weight, derived from diameter measurements at various cumulative temperatures after anthesis, was used as the explanatory variable, with final harvest weight as the target. Results indicated that incorporating estimates from multiple cumulative temperature points improved prediction accuracy, particularly for cultivars with stable growth patterns. Including average temperature as a variable further enhanced model performance. Yang et al. compared classical non-linear models and deep learning methods for predicting alfalfa leaf area index (LAI). Logistic, Gompertz, and Richards models were developed based on growth days, while a time-series model integrating environmental factors was proposed using a mutation point detection method and an encoder-attention-decoder BiLSTM network (TMEAD-BiLSTM). Results showed that the TMEAD-BiLSTM model outperformed non-linear models in prediction accuracy and effectively integrated environmental factors. Nian et al. estimated the rice aboveground biomass based on the first derivative spectrum and Boruta algorithm. Mustafa et al. developed a knowledge framework for yield prediction in cereal crops by leveraging UAVs.

2.2 Data assimilation between phenotyping and crop models

Gao et al. explored the quantitative relationship between soil profile salinity and soil depth in drip-irrigated cotton fields in southern Xinjiang using a multivariate linear regression model combined with a Kalman filter algorithm. The model effectively captured the dynamic changes in soil salinity across different growth stages and improved prediction accuracy after data assimilation. Based on the calibrated model and predicted soil conductivity data, the total cotton yield and income in the study area were estimated. The results demonstrate that the Kalman filter can enhance model reliability, providing a practical tool for monitoring soil salinity dynamics, assessing the relationship between soil salinity and cotton yield, and supporting efficient saline soil management in cotton fields. Zhou et al. propose a method for fruit selection and location in harvesting robots that accounts for obstacle perception. Synthetic data were generated using a 3D tomato greenhouse model and pixel-level segmentation labels. An attention-based feature extraction module (SFM) was designed to enhance the DeepLab v3+ segmentation network, improving detection of linear obstructions like stems and wires. An adaptive K-means clustering method was used to identify individual fruits. The barrier-free fruit selection algorithm identifies the largest, non-occluded fruit as the optimal target. This approach effectively detects and locates barrier-free fruits, providing a reliable solution for harvesting robots, applicable to other fruits and vegetables as well.

3 Perspectives

In conclusion, we propose an integrated framework that links plant phenotype, genotype, and environment (Figure 1), aiming to better synthesize current research efforts. Environmental parameters are commonly obtained via *in-situ* sensing, where sensors capture electrical signals (e.g., capacitance, resistance) and convert them into quantitative data such as air temperature,

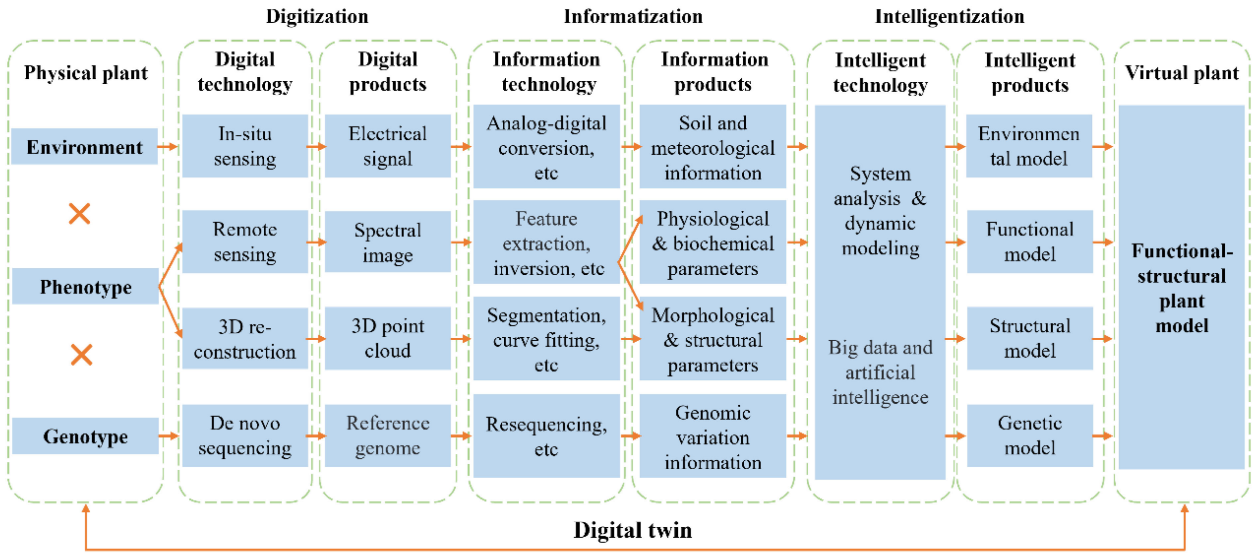


FIGURE 1
An integrated framework linking plant phenotype, genotype, and environment. The diagram outlines the collection of physiological, biochemical, morphological, structural, genomic, soil, and meteorological data using rapidly evolving technologies in recent years. These data are integrated through functional–structural plant models to link the physical and virtual representations of the plant, enabling the creation of a digital twin.

humidity, atmospheric pressure, photosynthetically active radiation, and soil temperature and moisture. These parameters facilitate the development of microclimate models, which can be further coupled with other simulation models. Phenotypic information is generally acquired through remote sensing and 3D reconstruction. Multispectral or hyperspectral imagery is processed through feature extraction and inversion to retrieve physiological and biochemical traits, while RGB imagery enables extraction of morphological and structural features at 2D level. Additionally, 3D point clouds derived from LiDAR or multi-view image reconstruction are processed through segmentation and surface modeling to obtain 3D structural traits. Therefore, functional and structural models are established through system analysis and dynamic modeling based on these phenotypes. In recent years, such models have been widely applied to investigate the impacts of climate change on crop productivity, identify potential yield gaps, and explore targeted improvement pathways (Gavasso-Rita et al., 2023). On the genetic level, reference genomes from *de novo* sequencing and genomic variations from resequencing support the development of genotype-based models (Zhang et al., 2023a).The integration of big data and artificial intelligence further enables hybrid modeling approaches—such as knowledge-guided machine learning (KGML) (Li et al., 2025) and improved phenotype–model data assimilation techniques (Jin et al., 2018). KGML leverages mechanistic knowledge of biological processes to guide the learning process, enhancing model interpretability and generalization capacity. Meanwhile, data assimilation techniques dynamically update model states and parameters using real-time phenotypic observations, thereby allowing high-throughput phenotyping data acquired by modern sensing technologies to be effectively integrated into the modeling framework. This unified phenotyping-modeling framework creates a digital twin by linking physical plants to their virtual counterparts, offering a promising pathway to integrate phenotyping with modeling for intelligent breeding and smart agriculture. However, the current framework remains incomplete, as it primarily emphasizes the virtual simulation of physical plants. Achieving a true digital twin requires establishing reverse control mechanisms that enable real-time feedback from the virtual twin to the physical system—a process that depends on further advancements in intelligent agricultural equipment and the seamless integration of agronomic practices with agricultural machinery.

Author contributions

TS: Investigation, Validation, Writing – original draft, Writing – review & editing. LX: Investigation, Validation, Writing – review & editing. SA-U-K: Investigation, Validation, Writing – review & editing. YM: Investigation, Validation, Writing – review & editing. WZ: Funding acquisition, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. National Key R&D Program of China (2022YFD2001001, 2022YFD2001005, 2017YFD0300409), Jiangsu Agricultural Science and Technology Innovation Fund (CX(21)1006), Jiangsu Provincial Key R&D Program (Modern Agriculture) (BE2023323), Second Division & Tiemenguan City Science and Technology Programme Project (2024HZ0107), Wuxi Municipal Financial Fund (33212303).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmed, N., and Shakoor, N. (2025). Advancing agriculture through IoT, Big Data, and AI: A review of smart technologies enabling sustainability. *Smart Agric. Technology.* 10, 100848. doi: 10.1016/j.atech.2025.100848

Castilho Silva, D., Madari, B. E., Santana Carvalho, M., Costa, J. V. S., and Ferreira, M. E. (2025). Planning and optimization of nitrogen fertilization in corn based on multispectral images and leaf nitrogen content using unmanned aerial vehicle (UAV). *Precis. Agric.* 26, 30. doi: 10.1007/s11119-025-10221-9

Fei, S., Jia, Y., Li, L., Xiao, S., Song, J., Yang, S., et al. (2025). Genetic resolution of multi-level plant height in common wheat using the 3D canopy model from ultra-low altitude unmanned aerial vehicle imagery. *Plant Phenomics* 7, 100017. doi: 10.1016/j.plaphe.2025.100017

Gavasso-Rita, Y. L., Papalexiou, S. M., Li, Y., Elshorbagy, A., Li, Z., and Schuster-Wallace, C. (2023). Crop models and their use in assessing crop production and food security: A review. *Food Energy Secur.* 13, e503. doi: 10.1002/fes3.503

Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., et al. (2018). A review of data assimilation of remote sensing and crop models. *Eur. J. Agron.* 92, 141–152. doi: 10.1016/j.eja.2017.11.002

Keating, B. A., and Thorburn, P. J. (2018). Modelling crops and cropping systems—Evolving purpose, practice and prospects. *Eur. J. Agron.* 100, 163–176. doi: 10.1016/j.eja.2018.04.007

Li, L., He, Q., Harrison, M. T., Shi, Y., Feng, P., Wang, B., et al. (2025). Knowledge-guided machine learning for improving crop yield projections of waterlogging effects

under climate change. *Resources Environ. Sustainability.* 19, 100185. doi: 10.1016/j.resenv.2024.100185

Maestrini, B., Mimić, G., van Oort, P. A. J., Jindo, K., Brdar, S., Athanasiadis, I. N., et al. (2022). Mixing process-based and data-driven approaches in yield prediction. *Eur. J. Agron.* 139, 126569. doi: 10.1016/j.eja.2022.126569

Sun, G., Zhang, Y., Chen, H., Wang, L., Li, M., Sun, X., et al. (2024). Improving soybean yield prediction by integrating UAV nadir and cross-circling oblique imaging. *Eur. J. Agronomy.* 155, 127134. doi: 10.1016/j.eja.2024.127134

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., et al. (2020). Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. *Mol. Plant* 13, 187–214. doi: 10.1016/j.molp.2020.01.008

Yu, Y., Cheng, Q., Wang, F., Zhu, Y., Shang, X., Jones, A., et al. (2023). Crop/plant modeling supports plant breeding: I. Optimization of environmental factors in accelerating crop growth and development for speed breeding. *Plant Phenomics.* 5, 99. doi: 10.34133/plantphenomics.0099

Zhang, P., Huang, J., Ma, Y., Wang, X., Kang, M., and Song, Y. (2023b). Crop/plant modeling supports plant breeding: II. Guidance of functional plant phenotyping for trait discovery. *Plant Phenomics.* 5, 91. doi: 10.34133/plantphenomics.0091

Zhang, C., Jiang, S., Tian, Y., Dong, X., Xiao, J., Lu, Y., et al. (2023a). Smart breeding driven by advances in sequencing technology. *Modern Agriculture.* 1, 43–56. doi: 10.1002/moda.8

Zhang, C., Kong, J., Wang, Z., Tu, C., Li, Y., Wu, D., et al. (2024). Origami-inspired highly stretchable and breathable 3D wearable sensors for *in-situ* and online monitoring of plant growth and microclimate. *Biosensors Bioelectronics.* 259, 116379. doi: 10.1016/j.bios.2024.116379

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 25 April 2024
DOI 10.3389/fpls.2024.1396183

Check for updates

OPEN ACCESS

EDITED BY
Ting Sun,
China Jiliang University, China

REVIEWED BY
Muhammad Ahmad Hassan,
Anhui Academy of Agricultural Sciences, China
Pan Gao,
Shihezi University, China
Haibing He,
Anhui Agricultural University, China

*CORRESPONDENCE
Jikai Liu
liujk@ahstu.edu.cn
Xinwei Li
lixw@ahstu.edu.cn

†These authors contributed equally to this work

RECEIVED 05 March 2024
ACCEPTED 11 April 2024
PUBLISHED 25 April 2024

CITATION
Nian Y, Su X, Yue H, Zhu Y, Li J, Wang W, Sheng Y, Ma Q, Liu J and Li X (2024) Estimation of the rice aboveground biomass based on the first derivative spectrum and Boruta algorithm.
Front. Plant Sci. 15:1396183.
doi: 10.3389/fpls.2024.1396183

COPYRIGHT
© 2024 Nian, Su, Yue, Zhu, Li, Wang, Sheng, Ma, Liu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Estimation of the rice aboveground biomass based on the first derivative spectrum and Boruta algorithm

Ying Nian[1], Xiangxiang Su[1], Hu Yue[1], Yongji Zhu[1], Jun Li[1], Weiqiang Wang[1], Yali Sheng[1], Qiang Ma[1], Jikai Liu[1,2*†] and Xinwei Li[1,2,3*†]

[1]College of Resource and Environment, Anhui Science and Technology University, Chuzhou, China, [2]Anhui Province Crop Intelligent Planting and Processing Technology Engineering Research Center, Anhui Science and Technology University, Chuzhou, Anhui, China, [3]Anhui Province Agricultural Waste Fertilizer Utilization and Cultivated Land Quality Improvement Engineering Research Center, Anhui Science and Technology University, Chuzhou, China

Aboveground biomass (AGB) is regarded as a critical variable in monitoring crop growth and yield. The use of hyperspectral remote sensing has emerged as a viable method for the rapid and precise monitoring of AGB. Due to the extensive dimensionality and volume of hyperspectral data, it is crucial to effectively reduce data dimensionality and select sensitive spectral features to enhance the accuracy of rice AGB estimation models. At present, derivative transform and feature selection algorithms have become important means to solve this problem. However, few studies have systematically evaluated the impact of derivative spectrum combined with feature selection algorithm on rice AGB estimation. To this end, at the Xiaogang Village (Chuzhou City, China) Experimental Base in 2020, this study used an ASD FieldSpec handheld 2 ground spectrometer (Analytical Spectroscopy Devices, Boulder, Colorado, USA) to obtain canopy spectral data at the critical growth stage (tillering, jointing, booting, heading, and maturity stages) of rice, and evaluated the performance of the recursive feature elimination (RFE) and Boruta feature selection algorithm through partial least squares regression (PLSR), principal component regression (PCR), support vector machine (SVM) and ridge regression (RR). Moreover, we analyzed the importance of the optimal derivative spectrum. The findings indicate that (1) as the growth stage progresses, the correlation between rice canopy spectrum and AGB shows a trend from high to low, among which the first derivative spectrum (FD) has the strongest correlation with AGB. (2) The number of feature bands selected by the Boruta algorithm is 19~35, which has a good dimensionality reduction effect. (3) The combination of FD-Boruta-PCR (FB-PCR) demonstrated the best performance in estimating rice AGB, with an increase in R^2 of approximately 10% ~ 20% and a decrease in RMSE of approximately 0.08% ~ 14%. (4) The best estimation stage is the booting stage, with R^2 values between 0.60 and 0.74 and RMSE values between 1288.23 and 1554.82 kg/hm^2. This study confirms the accuracy of hyperspectral remote sensing in estimating vegetation biomass and further explores the theoretical foundation and future direction for monitoring rice growth dynamics.

KEYWORDS

rice, AGB, remote sensing, hyperspectral data, derivative transform, machine learning algorithm

1 Introduction

Rice (*Oryza sativa*), a herbaceous plant belonging to the genus *Oryza*, has a long history of cultivation and a broad planting region (Krishnan et al., 2011). Aboveground biomass (AGB) serves as a critical indicator for describing crop growth (Wang et al., 2016). Applying scientific methods to estimate rice AGB to improve crop yields is crucial for human food security and increased economic benefits (Spiertz and Ewert, 2009; Devia et al., 2019; Cao et al., 2021b). The conventional method for acquiring AGB primarily involves field sampling, which is both time-consuming and laborious (Gnyp et al., 2014). In contrast, remote sensing technology provides a fast and nondestructive method, providing a scientific basis and technical support for accurate estimation of AGB (Schino et al., 2003). Among them, satellite remote sensing platforms is commonly used for AGB estimation in large areas, such as forests (Lucas et al., 2020), grasslands (Fang et al., 2018; Yang et al., 2018), However, it is easily affected by spatial conditions and it is difficult to obtain satisfactory results at small field scales (Yue et al., 2019). UAV remote sensing platforms can perform low-altitude operations and obtain high-resolution images of crops, but the spectral resolution is low, which is not conducive to in-depth exploration of the spectral response of crops (Bansod et al., 2017). In comparison, near-ground hyperspectral remote sensing is not only suitable for small-scale crop monitoring, but can also acquire fine spectral data with high resolution and a broad spectrum of wavelengths, encapsulating abundant crop phenotype information and thereby offering increased opportunities for estimating the physical and chemical parameters of vegetation (Galvão et al., 2005; Atzberger et al., 2015).

To enhance the AGB estimation performance via hyperspectral remote sensing, the original spectral data are typically preprocessed to reduce noise and increase the accuracy of the data (Mishra et al., 2020). The Savitzky–Golay (SG) smoothing algorithm and derivative transform are the principal methods for preprocessing. The SG smoothing algorithm helps to mitigate interference signals within spectral data and minimize random errors, thereby improving data reliability (Savitzky and Golay, 1964; Liu et al., 2014). Derivative transform spectrum emphasizes the absorption and reflection characteristics of the subject matter and augments the differentiation in spectral information (Gerretzen et al., 2015). For instance, Wang et al. (2004) demonstrated that the derivative spectrum can lessen the impact of noise, exhibits a high correlation with rice AGB, and performs effectively in estimating rice AGB. However, spectral data still face challenges such as excessive information redundancy and high dimensionality after preprocessing. There is a need for a reliable method that can reduce the dimensionality of spectral data and eliminate high correlations between bands.

In previous research, Some researchers have used a single vegetation index to estimate rice AGB and achieved good results. However, the vegetation indices is usually a combination of several sensitive wavelengths, ignoring a large volume of spectral information. In the 2020s, some scholars adopted complex dimensionality reduction methods to improve the utilization efficiency of rice spectral data. For example, Jia et al. (2022) used the continuous projection algorithm (SPA) to reduce the hyperspectral dimension and identified 12 characteristic bands to estimate the soil plant analysis development (SPAD) content of rice. Yu et al. (2020) utilized principal component analysis (PCA) to extract five principal components for the inversion of rice leaf nitrogen deficiency. The above studies used different dimensionality reduction methods to reduce spectral data redundancy, which is of positive significance for applying hyperspectral remote sensing technology to estimate rice AGB accurately. However, they present limitations. The objective of SPA is to eliminate combinations of variables that exhibit low collinearity and include distinctive information; however, this approach might omit some initial characteristics in the spectral data (Cao et al., 2021a). PCA is capable of using several independent principal components to depict the original data, yet this technique is confined to linear projection and exhibits inadequate performance in managing nonlinear data relationships (Moore, 1981; Abdi and Williams, 2010). Consequently, the effectiveness of the monitoring model in practical scenarios may be affected by different dimensionality reduction methods (Fu et al., 2020). Given these considerations, selecting a suitable dimensionality reduction method is essential for accurately identifying sensitive bands and enhancing the accuracy of the estimation model.

As a data dimensionality reduction method, the core of Boruta algorithm is to find all feature bands related to the dependent variable. Compared with other commonly used feature screening algorithms, Compared with other commonly used feature screening algorithms (Kursa and Rudnicki, 2010; Kursa et al., 2010). The algorithm was originally used in biological and medical fields and has since been used in agricultural research as well. For instance, Li et al. (2022) demonstrated that filtering spectral indices from winter wheat canopy hyperspectral data using the Boruta algorithm enhances data validity and establishes an accurate yield estimation model. The Boruta algorithm has great potential in physical and chemical parameter estimation. To further evaluate the ability of the Boruta algorithm to estimate rice AGB, this study also selected the RFE feature selection algorithm for comparison. RFE, an adaptive feature selection algorithm, iteratively removes the least important feature variables and ranks feature importance until the optimal feature subset is identified. This approach minimizes the effects of random fluctuations and interference information (Tunca et al., 2023). Yoosefzadeh-Najafabadi et al. (2021) applied RFE to determine the optimal wavelengths from hyperspectral data for soybean yield prediction. This indicates that these feature selection methods are effective in improving estimation models' accuracy. However, relatively few studies have utilized the Boruta algorithm and the RFE algorithm for feature selection of derivative spectra to construct rice AGB estimation models.

Therefore, this study used the Boruta and RFE algorithms to eliminate interference information, identify sensitive bands, and integrate partial least squares regression (PLSR), principal component regression (PCR), support vector machine (SVM) and ridge regression (RR) to develop an AGB estimation model. Additionally, this study compares the estimation effects of various combinations of dimensionality reduction algorithms and machine learning models to identify the most accurate estimation results.

The research objectives were to (1) compare the correlations between derivative spectrum of different orders and rice AGB to determine the optimal order; (2) evaluate the effectiveness of the Boruta algorithm and RFE algorithm in selecting feature bands to determine the best dimensionality reduction method; (3) evaluate the estimation capability of machine learning algorithms combined with feature selection algorithms on rice AGB; and (4) identify the optimal combination of models and the most suitable estimation stage.

2 Materials and methods

2.1 Experimental design

The study was conducted in Xiaogang village, Fengyang County, Chuzhou city, Anhui Province, China (longitude 117°46′7″E, latitude 32°48′52″N) (Figure 1A), which is characterized by a subtropical monsoon climate, an average annual temperature of 15.4°C, and an average annual precipitation of 1179.2 mm. The soil predominantly consists of clay with medium fertility on flat terrain. The experimental area included 36 plots, measuring 2 m × 8 m each. The rice varieties tested were Runzhuxiangzhan (V1), Runzhuyinzhan (V2), and Hongxiangnuo (V3), with each plot replicated three times. Four nitrogen gradient treatments were applied with nitrogen application rates of N0 (0 kg/hm^2), N1 (100 kg/hm^2), N2 (200 kg/hm^2), and N3 (300 kg/hm^2). Phosphorus (90 kg/hm^2) and potash (135 kg/hm^2) fertilizers were applied once as basal fertilizer (Figure 1B). The experiment spanned from June 2020 to October 2020. The plants experienced no drought or flooding during the seedling stage and favorable conditions of abundant sunshine and moderate temperatures during the mid-growth stage, which are conducive to rice growth and development. Field management practices followed local cultivation methods and pest control technologies.

2.2 Data acquisition

2.2.1 Ground data acquisition

Rice samples were collected on July 25 (tillering stage), August 13 (jointing stage), August 23 (booting stage), August 31 (heading stage) and September 20 (maturity stage). The sampling time was consistent with the hyperspectral data acquisition time. Rice samples were collected from three holes in each plot, and the locations of the sampling points were marked after each sampling to avoid duplicate sampling.

Rice samples were subsequently sent to the laboratory, after which stem, leaf, and ear separation was performed (Figure 1C). The sample was placed in an oven, dried for 0.5 h after the temperature increased to 105°C, and then adjusted to 75°C for more than 24 h until the mass was constant. The dry mass of the sample was obtained by summing the dry mass of the stem, leaves and ears of the plant. Finally, the aboveground biomass of the rice plants in each plot was obtained through the dry mass of the sample and the row spacing during rice planting.

2.2.2 Hyperspectral data collection

Hyperspectral data on the rice canopy spectra were collected using a handheld ground spectrometer (ASD FieldSpec® HandHeld^{TM}2) covering a wavelength range from 325 to 1075 nm with a resolution of 1 nm. Measurements were conducted between 10:00 a.m. and 2:00 p.m. for each test. Prior to each measurement session, the instrument was calibrated against a standard radiation white plate and recalibrated after every two blocks to ensure accuracy. The data were collected at three evenly distributed points along the diagonal of each plot, with the spectrometer positioned vertically 0.5 m above the rice canopy (Figure 1C). The spectral data from the same plot were processed collectively, and the average spectral reflectance of the plot was calculated.

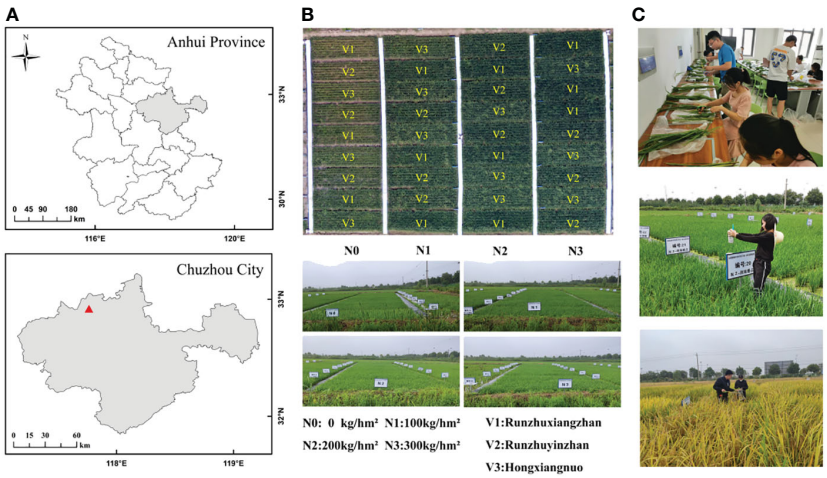


FIGURE 1
Rice experimental area **(A, B)** and data collection **(C)**.

2.3 Hyperspectral data processing

2.3.1 SG smoothing and derivative transform

The Savitzky–Golay (SG) smoothing algorithm, a digital signal processing technique proposed by Savitzky and Golay, is employed to enhance signal frequency and eliminate data noise (Savitzky and Golay, 1964). The smoothing effect of the algorithm, which varies with the window length, preserves the original shape and peak heights of the wave signal (Schafer, 2011). To minimize data fluctuations, capture the overall trend of the spectral curves, and accurately analyze the spectral characteristics of rice, this study applied SG smoothing to the canopy spectral curves from 325 to 1075 nm recorded during the growth stage (Chen et al., 2020; Shi et al., 2021). However, due to the influence of noise, only the 400 to 900 nm range was selected for analysis.

Derivative transform serves to diminish background signals, with the FD highlighting absorption peaks and shoulders in the original spectrum (OS) (Demetriades-Shah et al., 1990; Becker et al., 2005). By differentiating peaks and valleys, more precise spectral reflectance data can be obtained (Ji et al., 2022; Zhao et al., 2022). The second derivative spectrum (SD) facilitates the accurate identification of absorption peak and shoulder positions within the OS, enhancing spectral reflectance differentiation and eliminating baseline offsets in spectral reflectance data (Yang et al., 2011). This study utilized the derivative function within the Origin software's data processing menu (Origin 2018) for these transformations.

2.3.2 Feature selection

Feature selection algorithms play a pivotal role in optimizing datasets prior to model construction. In this investigation, both the Boruta algorithm and the RFE algorithm were employed for feature selection.

The Boruta algorithm (Prasad et al., 2019) is an innovative feature selection method derived from the random forest approach. It identifies all characteristic bands in spectral data that are related to dependent variables, thereby elucidating the relationship between spectral characteristics and rice AGB. The foundational concept involves shuffling the original parameters to create shadow parameters, which are then combined with the original parameters into a feature matrix for training. Based on the importance scores from the random forest, the shadow parameters are ranked by importance, with the highest value designated the Z score. Original parameters more significant than the Z score are labeled "most important", while those below are deemed "unimportant" and thus eliminated. Ultimately, the most important original parameters are selected as the optimal combination for constructing the inversion model.

Recursive feature elimination (RFE) (Guyon et al., 2002) is a wrapper-based selection method that starts with the construction of an initial model, ranking all feature bands by their importance, and iteratively removing the least significant features. By retraining the model with the remaining features and repeating this process, it gradually identifies the most critical subset of features. This algorithm excels by iteratively evaluating the contribution of each feature band to the model, thereby identifying the optimal feature subset to reveal the underlying structure and patterns within the data.

2.4 Model construction method

To thoroughly investigate the relationships between the independent and dependent variables, four machine learning models were selected for comparison: PLSR, PCR, SVM, and RR.

PLSR (Geladi and Kowalski, 1986) is a multivariate linear statistical method that focuses on finding a hyperplane that minimizes the variance between the response and independent variables. It projects predictor and observed variables into a new space to derive a linear regression model, known as a bilinear factor model, due to the presence of both data X and Y in this new space.

PCR (Kawano et al., 2018) employs PCA for predictive data mining, transforming original variables into principal components that are linear combinations of the original variables and mutually independent. Regression analysis is then performed on these principal components to derive the regression equation, which is subsequently applied to the original variables.

SVM (Suykens and Vandewalle, 1999) operates as a supervised learning model utilized for analyzing data in classification and regression analysis. The core concept involves identifying the optimal classification hyperplane for two types of samples in the original space when they are linearly separable. In instances where linear separability is not achievable, the samples are projected into a high-dimensional feature space where an optimal hyperplane is determined. This hyperplane classifies the samples, aiming to minimize the distance within similar classes and maximize the distance between distinct classes.

RR (Ivanda et al., 2021) serves as a technique for estimating the coefficients of a multiple regression model in scenarios where the independent variables exhibit high correlation. This is achieved by incorporating an L2 regularization term into the OLS loss function. The L2 regularization term is calculated as the product of the sum of the squares of the coefficients and a regularization parameter λ (lambda).

2.5 Accuracy evaluation

In this study, 70% of the sample data (n=25) were selected as the modeling set for each growth stage, and 30% of the sample data (n=11) were used as the validation set to construct the rice AGB estimation model. The coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) were used to evaluate the model accuracy. The closer R^2 is to 1, the lower the RMSE and MAE are, and the higher the accuracy of the estimation model is. R^2, RMSE and MAE are calculated using Equations 1–3, respectively:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \hat{x})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \quad (3)$$

Note: x_i, y_i, and $\hat{x}$ are the actual measured value, predicted value and average value of the measured data, respectively; n is the number of samples.

3 Results

SG smoothing was applied to the spectral curves within the 400 to 900 nm range, The processing results of OS, FD and SD at different growth stages are shown in Figure 2.

The processing results for the OS reveal the characteristic spectral profile of the rice canopy, marked by typical green plant reflectance features, with spectral reflectance values ranging from 0 to 0.5. Notably, a reflection peak and an absorption valley are observed near wavelengths of 550 nm and 680 nm, respectively, both of which are situated within the visible light spectrum. The spectral reflectance of rice rapidly increases within the 680 to 750 nm range, creating a "red edge" phenomenon (Gitelson et al., 1996). The first peak of the FD curve appears at 500 ~ 550 nm. The band range of 680 ~ 750 nm shows a drastic change that first rises and then falls. The SD curve exhibits continuous fluctuations across the 500 to 690 nm band range, with two notable peaks within the 700 to 800 nm band range.

3.1 Correlations between different orders of derivative spectrum and rice AGB

To fully explore the sensitivity of the different orders of rice canopy spectra and to compare the effects of spectral transformations on AGB, FD and SD transformations were carried out on the basis of OS, and their correlations with rice AGB were investigated (Figure 3).

During the full growth stage, the correlation between OS and AGB showed a consistent trend from negative to positive as the band increased, with the maximum correlation at 720 nm. The correlation curve between FD and AGB shows a continuous fluctuation trend in the 400 ~ 900 nm band, among which the 680 ~ 750 nm band has the highest correlation, with a correlation coefficient between 0.60 ~ 0.85. The correlation curve between SD and AGB was similar to that between FD and AGB, but the fluctuations were more severe.

In summary, the derivative spectrum of the different orders were compared with the maximum absolute value of the AGB correlation coefficient (Figure 4). Except for the tillering stage, the maximum values in the other four periods all occurred at FD. Comprehensive analysis revealed that the FD could better reflect the growth status of rice, so the FD was used as the input variable of the feature screening algorithm for subsequent construction of the AGB estimation model.

3.2 Feature band selection

An AGB estimation model utilizing PLSR, PCR, SVM, and RR was developed to evaluate the effectiveness of the band screening algorithms. This model incorporated characteristic bands selected by both the Boruta algorithm and the RFE algorithm, as well as the full band as input variables. The diversity and scope of the characteristic bands selected by these feature screening algorithms during different growth stages are summarized in Table 1. Detailed feature selection results are shown in Appendix 2.

The Boruta algorithm demonstrated remarkable stability, with the number of characteristic wavelengths in each growth stage not exceeding 40. In contrast, the RFE algorithm showed more significant fluctuations, with 3 characteristic wavelengths identified at both the tillering and booting stages and 439 at the maturity stage, indicating substantial variance in the number of wavelengths identified across different growth stages. Notably, the characteristic wavelengths identified by both screening algorithms predominantly fall within the "red edge" spectral range, highlighting their importance in estimating rice AGB.

3.3 Establishment and evaluation of the rice AGB estimation model based on FD

This study employs three types of model input variables: first derivative full band (FD-FB), the characteristic bands selected by

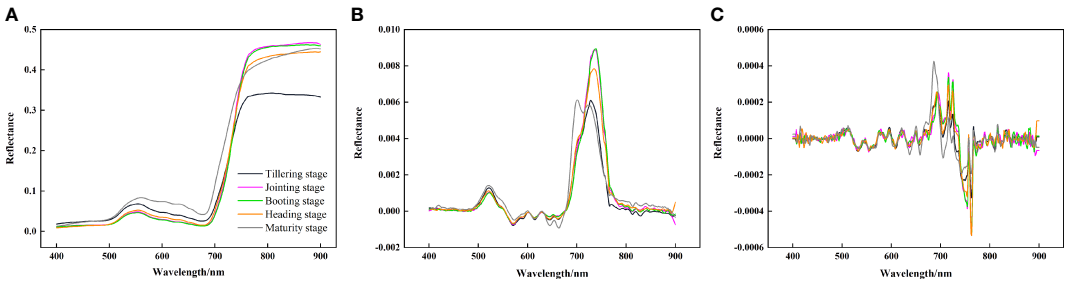


FIGURE 2
Derivative transformation spectral curves of rice canopy at different growth stages. **(A)** original spectrum, **(B)** first derivative spectrum, **(C)** second derivative spectrum.

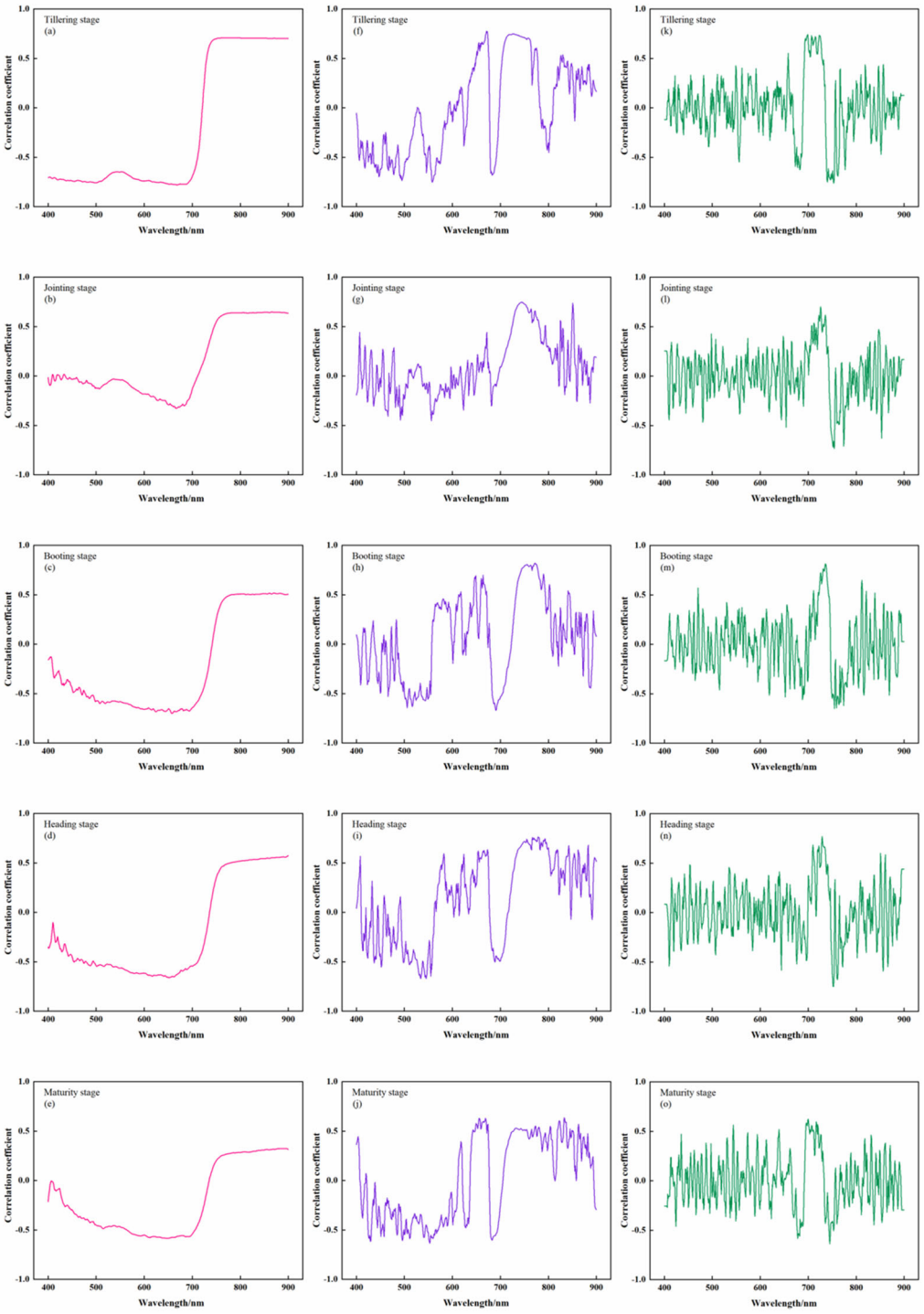


FIGURE 3
Correlation between rice canopy spectra and AGB at different growth stages. **(A–E)** The correlation between OS and AGB; **(F–J)** the correlation between FD and AGB; **(K–O)** the correlation between SD and AGB.

the Boruta algorithm (FD-Boruta), and the characteristic bands selected by the RFE algorithm (FD-RFE). Sixty AGB estimation models were developed for various growth stages using four machine learning algorithms (PLSR, PCR, SVM, RR), with R^2 (Figure 5), RMSE (Table 2) and MAE (Appendix 1) serving as the model evaluation metrics.

Initially, the AGB estimation abilities of different feature selection algorithms were compared within the same model framework. In the PLSR model, the estimation precision for FD-Boruta and FD-RFE was found to be comparable, with R^2 values ranging from 0.43 to 0.70, and RMSE values between 464.34 and 2515.77 kg/hm^2 and 493.81 and 2407.61 kg/hm^2, respectively. These

10.3389/fpls.2024.1396183

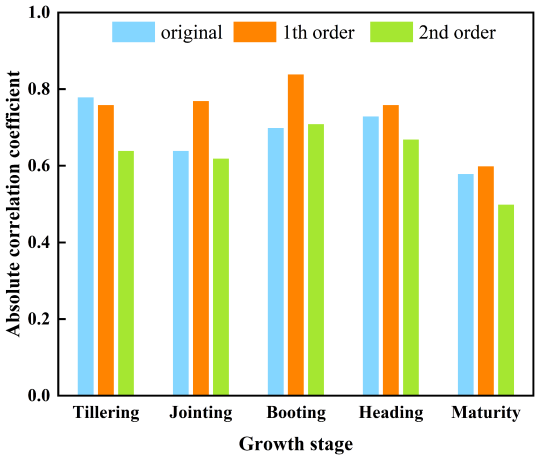


FIGURE 4
The maximum absolute value of the correlation coefficient between different order derivative spectra and AGB.

results are more accurate than those obtained using FD-FB, with R^2 values improving by 0 to 0.06 and RMSE values decreasing by 29.47 to 148.91 kg/hm^2. Among the PCR models, FD-Boruta exhibited the highest estimation accuracy, followed by FD-RFE, with FD-FB showing the least accuracy. The R^2 accuracy of FD-Boruta improved by 0.03, 0.19, 0.02, 0.01, and 0.04, respectively, compared to FD-RFE. In contrast, the R^2 accuracy of FD-RFE improved by 0.02, 0.05, 0.07, and 0.01 compared to FD-FB. The outcomes for the SVM and RR models were analogous, although the performances of the three algorithms varied significantly. R^2 ranged from 0.58 to 0.71 for FD-Boruta, 0.22 to 0.74 for FD-RFE, and 0.19 to 0.60 for FD-FB.

Second, the estimation accuracy of AGB by different models under the same feature selection algorithm was compared. For the model constructed with FD-Boruta inputs, when the estimation accuracy was similar, FD-Boruta-PCR demonstrated more stable performance. The estimation outcomes of models based on FD-RFE varied significantly. FD-RFE-PLSR and FD-RFE-PCR provided more precise estimates of rice AGB, whereas the FD-RFE-SVM and FD-RFE-RR models exhibited overfitting in certain stages. The models based on FD-RFE generally showed poor performance, with model R^2 values below 0.65, and the accuracy of the FD-FB-SVM and FD-FB-RR models varied greatly.

Finally, the performance of all model combinations was assessed for estimating AGB effects across different growth stages. In the tillering stage, FD-Boruta-SVM yielded the best estimation, with an R^2 of 0.66 and an RMSE of 34.33 kg/hm^2. During the jointing stage, FD-Boruta-SVM achieved the highest estimation accuracy, with an R^2 of 0.65 and an RMSE of 985.55 kg/hm^2, while FD-FB-SVM showed the lowest accuracy, with an R^2 of 0.32 and an RMSE of 1412.37 kg/hm2. In the booting stage, the estimation accuracies R^2 of models using FD-FB and FD-RFE inputs were both above 0.7, with FD-RFE-SVM and FD-Boruta-PCR performing the best, achieving R^2 values of 0.74 and 0.72, respectively, and RMSE values of 1311.04 kg/hm^2 and 1290.24 kg/hm^2, respectively. At the heading stage, FD-Boruta-PCR had the highest estimation accuracy, with an R^2 of 0.71, while FD-FB-SVM had the lowest accuracy, with an R^2 of 0.30. In the maturity stage, the AGB estimation accuracy of all the models decreased, with R^2 values ranging from 0.19 to 0.53.

In conclusion, models built using the Boruta algorithm are more reliable and exhibit greater stability, among which FD-Boruta-PCR is the most stable, followed by FD-Boruta-PLSR. Considering the model estimation accuracy, the booting period is the most accurately estimated, with R^2 values exceeding 0.7.

TABLE 1 Comparison of feature selection results between the Boruta algorithm and RFE algorithm.

Feature selection algorithm	Growth Stage	Number	Feature band range (nm)
Boruta	Tillering	19	724 ~ 761
	Jointing	35	644, 721 ~ 771, 816, 817
	Booting	27	664 ~ 694, 717 ~ 774
	Heading	24	414, 416, 699 ~ 777,817, 873
	Maturity	30	420, 498, 686 ~ 759, 842
RFE	Tillering	3	742, 751, 761
	Jointing	392	402 ~ 900
	Booting	3	753, 754, 764
	Heading	127	402 ~ 495, 507 ~ 594, 614 ~ 699, 702 ~ 900
	Maturity	439	400 ~ 900

4 Discussion

4.1 Correlations between OS, FD, SD and AGB

An in-depth analysis of the correlation between spectral information and AGB will facilitate a comprehensive understanding of the growth status of rice. This study performed a correlation analysis between different orders of derivative spectrum and AGB at different growth stages of rice. Generally, the correlation between the derivative spectrum of various orders and AGB tended to increase and then decrease throughout the growth stage, potentially due to the influence of the rice growth cycle (Xu et al., 2023). From the tillering stage to heading stage, as the chlorophyll content in rice leaves increases and the vegetation canopy structure becomes fully developed, the canopy becomes richer in spectral information, which can more accurately reflect crop characteristics (Ren et al., 2018). In the mature stage, as stems and leaves progressively wither and yellow, the chlorophyll content drops sharply, the spectral characteristics obtained are less able to accurately represent crop growth, leading to a decrease in the correlation between the canopy spectrum and AGB (Yang and Chen, 2004; He et al., 2019; Zhang et al., 2019).

Within the 680~750 nm band range, the correlation between the canopy spectrum and AGB reached its maximum value in different

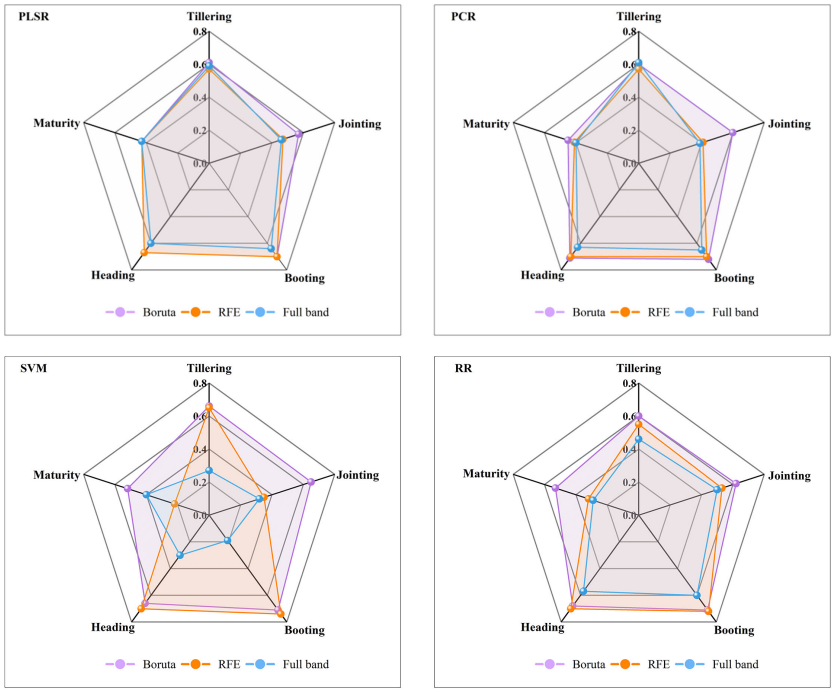


FIGURE 5
Comparison of the R^2 values of the different estimation models for the validation sets. PLSR, partial least squares regression; PCR, principal component regression; SVM, support vector machine; RR, ridge regression.

stages (Figure 3). This occurs because this band serves as the transition region between the infrared and near-infrared bands, where spectral reflectance transitions rapidly from a low negative correlation to a high positive correlation. This shift is attributed to strong absorption and reflection (Kanke et al., 2016; Xu et al., 2022). The correlation between FD and AGB was generally greater than that between OS and SD at the different fertility stages (Figure 4). This observation aligns with the findings of Liang et al. (2018) and Tong et al. (2023). As a method for analyzing spectral information, derivative transform can diminish noise and enhance data accuracy (Li and Xie, 2015). FD reflects the slope of the spectrum, while SD represents the change in the slope of the reflection spectrum. While SD identifies more absorption peaks, it also introduces noise and may result in errors (Shen et al., 2022). Therefore, FD is strongly correlated with AGB and can be effectively utilized for estimating rice AGB.

4.2 Rice AGB estimation based on different feature selection algorithms

In the realm of feature selection, prior research has demonstrated that employing screened feature variables for model construction enhances the estimation power, inversion accuracy, and utility of the original models (Gao et al., 2019; Sun et al., 2021). For example, Yu et al. (2023) used the SPA algorithm to extract sensitive bands in rice canopy spectral data, and combined with PROSAIL to establish a rice AGB estimation model, achieving high accuracy (R^2=0.69). This study yielded similar findings, the characteristic band had a greater estimation ability than the original band. The difference is that previous research is usually based on OS and does not involve derivative transform. Therefore, whether the estimation model constructed by derivative transformed spectrum is better than OS remains to be determined. This study performs SG smoothing and derivative transform on OS and selects characteristic bands based on the optimal derivative spectrum. The results show that the Boruta algorithm is more robust than the RFE algorithm, and the selected feature bands are more sensitive, which is conducive to constructing subsequent rice AGB estimation model.

Under identical conditions, the most accurate AGB estimates were achieved using feature bands selected by the Boruta algorithm. This superiority of the Boruta algorithm is attributed to its ability to identify bands of features that are genuinely relevant to the dependent variable, thereby enhancing the prediction accuracy of the model (Kursa and Rudnicki, 2010), a conclusion that aligns with the findings of Degenhardt et al. (2019). However, the RFE algorithm exhibited suboptimal performance in the jointing, heading, and maturity stages. This could be due to the RFE algorithm generates feature subsets with corresponding accuracy by continuously building models. This process may result in retaining a large number of features, leading to significant collinearity between bands and diminishing the model's estimation accuracy (Paul et al., 2015; Chen et al., 2018), echoing the research results of Lin et al. (2023). The poorest performance was observed when the entire band was utilized as an input variable, attributed to the presence of redundant information and increased collinearity among bands, which impaired the model's estimation capability (Hansen and Schjoerring, 2003; Pan et al., 2017).

TABLE 2 Rice AGB estimation model under different machine learning algorithms (R^2, RMSE).

machine learning algorithm	Bandselection algorithm	Tillering				Jointing				Booting				Heading				Maturity			
		Modeling Set		Validation Set		Modeling Set		Validation Set		Modeling Set		Validation Set		Modeling Set		ValidationSet		Modeling Set		Validation Set	
		R^2	RMSE (kg/hm^2)	R^2	RMSE (kg/hm^2)	R^2	RMSE (kg/hm^2)	R^2	RMSE (kg/hm^2)	R^2	RMSE (kg/hm^2)	R^2	RMSE (kg/hm^2)	R^2	RMSE (kg/hm^2)	R^2	RMSE (kg/hm^2)	R^2	RMSE (kg/hm^2)	R^2	RMSE (kg/hm^2)
PLSR	Boruta	0.54	427.85	0.61	464.34	0.68	801.67	0.57	1072.49	0.72	1096.66	0.70	1351.62	0.72	1125.80	0.67	1405.25	0.49	2017.00	0.43	2515.77
	RFE	0.53	434.66	0.57	493.81	0.78	635.91	0.47	1226.82	0.70	1148.63	0.70	1336.75	0.80	908.01	0.67	1398.94	0.63	1736.85	0.43	2407.61
	FB	0.72	324.48	0.59	473.29	0.77	644.07	0.46	1245.08	0.81	912.34	0.64	1474.44	0.79	948.73	0.60	1547.85	0.63	1744.52	0.43	2406.05
PCR	Boruta	0.54	427.33	0.60	469.15	0.64	853.54	0.60	1057.88	0.72	1101.43	0.72	1290.24	0.70	1171.49	0.71	1307.01	0.45	2120.01	0.45	2392.65
	RFE	0.53	435.68	0.57	493.77	0.43	1067.73	0.41	1311.00	0.70	1150.33	0.70	1338.99	0.70	1180.64	0.70	1321.14	0.43	2178.91	0.41	2412.86
	FB	0.55	423.94	0.61	477.47	0.41	1087.71	0.39	1332.82	0.68	1186.97	0.65	1465.01	0.59	1377.74	0.63	1541.66	0.41	2205.87	0.40	2430.66
SVM	Boruta	0.67	368.29	0.66	434.33	0.80	663.56	0.65	985.55	0.86	800.08	0.71	1349.87	0.82	902.99	0.66	1411.43	0.85	1135.73	0.52	2189.95
	RFE	0.67	367.30	0.65	445.05	0.97	329.32	0.35	1370.60	0.75	1073.13	0.74	1311.04	0.99	216.76	0.70	1430.80	0.96	716.15	0.22	2964.74
	FB	0.80	481.27	0.27	701.52	0.98	276.86	0.32	1412.37	0.97	486.02	0.19	2316.48	0.98	386.45	0.30	2266.78	0.95	852.86	0.19	2980.00
RR	Boruta	0.58	421.84	0.60	504.74	0.72	764.35	0.62	1009.78	0.76	1036.43	0.71	1301.00	0.81	955.61	0.68	1370.52	0.63	1827.13	0.53	2217.62
	RFE	0.53	435.50	0.55	513.04	0.90	513.16	0.53	1120.75	0.70	1144.30	0.72	1288.23	0.94	579.73	0.70	1318.70	0.91	1102.65	0.32	2628.54
	FB	0.95	192.63	0.46	514.86	0.90	518.87	0.50	1160.32	0.94	627.07	0.60	1554.82	0.95	612.77	0.57	1588.79	0.92	1095.07	0.29	2648.03

PLSR, partial least squares regression; PCR, principal component regression; SVM, support vector machine; RR, ridge regression.

10.3389/fpls.2024.1396183

Compared with the RFE algorithm, the Boruta algorithm identified fewer feature bands (Table 1); however, the resulting estimation model was more stable. This stability may be due to the characteristic bands determined by the Boruta algorithm being more effective and independent (Poona et al., 2016). Additionally, when comparing feature bands screened by both the Boruta and RFE algorithms, it was noted that their intersection predominantly occurred in the "red edge" region. This could be due to the band encapsulates rich crop information and is highly correlated with AGB (Horler et al., 1983; Filella and Penuelas, 1994), consistent with the findings of Cheng et al. (2017), who assessed the estimation effects of eight vegetation indices on rice canopy composition AGB and found the red edge index to be particularly sensitive to leaf AGB, achieving the highest estimation accuracy.

Another aspect of interest is that the characteristic bands identified by the Boruta algorithm, which are relatively evenly distributed across the spectrum, are predominantly found at wavelengths of 420 nm, 644 nm, 750 nm, and 842 nm. These characteristic bands are mainly located in areas sensitive to crop AGB response and are relatively evenly distributed. During the tillering stage, the characteristic bands are primarily located in the "red edge" region, which may be due to the low vegetation coverage at this time and the spectrum response is not obvious (Kokaly and Clark, 1999; Clevers et al., 2001). In the jointing stage, characteristic wavelengths are observed in the near-infrared region, likely due to the increase in rice canopy biomass, making the wavelength response in this region more pronounced and easier to detect during the selection process (Datt, 1998). From the heading to the maturity stage, a few characteristic wavelengths in the blue light region are identified, possibly because the chlorophyll content in the leaves gradually decreases, leading to increased spectral reflectance in the blue light region (Yan-lin et al., 2004). Thus, the Boruta algorithm effectively mines the spectral information of the rice canopy, and the selected characteristic bands align with the spectral response changes in the rice canopy, thereby enhancing the accuracy of AGB estimation.

4.3 Rice AGB estimation based on different machine learning algorithms

The performances of various models were meticulously compared, with the PCR model emerging as the most reliable and stable for estimating AGB. The differences between the R^2 values for the modeling set and the validation set are minimal, ranging from 0 to 0.1. Both the RMSE and MAE are lower for the PCR model than for the other models. This superior performance of the PCR model can be attributed to its foundation in predictive data mining technology through principal component analysis (PCA), which efficiently reduces the dimensionality of independent variables, mitigates the effects of multicollinearity among variables, and enhances the model's adaptability (Suryakala and Prince, 2019). In the PLSR model. Models constructed with three different input variables all demonstrated the ability to estimate rice AGB, particularly at the booting stage, when the R^2 of the validation set exceeded 0.6. This performance is likely due to the robust adaptability of the PLSR algorithm, its ability to diminish the dimensionality of spectral data, its ability to effectively handle complex collinear relationships between independent variables, and its overall enhancement of the model's estimation accuracy (Helland, 2001). In the SVM model, estimation models built on FD-RFE and FD-FB inputs exhibited overfitting at the jointing, heading, and maturity stages. Due to the excessive number of FD-RFE and FD-FB, when fitting as sample data, the noise interference in the model is large, resulting in an increase in estimation error. Even some noise may be recognized as characteristic frequency bands by machine learning algorithms, destroying the preset classification rules and significantly affecting the accuracy of model estimation (Cherkassky, 1997; Shafaei and Kisi, 2017; Raja et al., 2022). FD-Boruta has a small number but high estimation potential and can solve collinearity problems well when combined with the SVM algorithm. Similar outcomes were also observed in the RR model. This may be because the RR algorithm cannot automatically select important feature variables when fitting sample data. Therefore, when the amount of input sample data is large, it is easily affected by outliers, thereby reducing the model's predictive ability. It may also be that because determining optimal ridge parameters is critical when building an estimation model, poor parameter selection can lead to overfitting of the model, thereby reducing its interpretability (Smith and Campbell, 1980; Forrester and Kalivas, 2004). Moreover, this study demonstrates that the model achieves the highest estimation accuracy during the booting stage. This can be ascribed to the booting stage, which is the primary period for increases in rice chlorophyll and is characterized by high vegetation coverage and the absence of panicles. At this juncture, noise interference is minimized, and the spectral information collected is purer and more precise, providing advantageous conditions for estimating rice AGB (Chang et al., 2005; Takai et al., 2006; Kawamura et al., 2018).

5 Conclusion

This study implements SG smoothing and derivative transform for the preprocessing of spectral data, aiming to diminish fluctuations and noise interference. Subsequently, the Boruta and RFE algorithms are applied to select features from the first-order spectrum, thereby reducing the data dimensions and information redundancy and enhancing the data accuracy. A rice AGB estimation model was developed from the tillering stage to the maturity stage by integrating PLSR, PCR, SVM, and ridge regression machine learning algorithms. The key findings of the research are as follows:

(1) Except during the booting stage, FD exhibited the strongest correlation with AGB across the other four growth stages, followed by OS and then SD. As the rice growth stage progresses, the correlation shows a trend of first increasing and then decreasing.

(2) The performance of the Boruta algorithm is more stable, the selected characteristic bands are more sensitive, and effective dimensionality reduction of spectrum data is

achieved. The number of feature strips provided by the RFE algorithm ranges from 3 to 439, and the number of feature strips selected by the Boruta algorithm is within 40.

(3) Among all model combinations, the characteristic band selected by the Boruta algorithm performs best, and FD-Boruta-PCR emerged as the superior model for estimating rice AGB, with R^2 values between 0.45 and 0.72 and RMSE values between 469.15 and 2392.65 kg/hm^2. The worst performing model is the FD-FB-SVM model, with R^2 values between 0.19 and 0.32 and RMSE values between 701.52 and 2980.00 kg/hm^2.

(4) As the progression from the tillering stage to the mature stage occurs, the accuracy of the AGB estimation model decreases. Among them, the booting stage was determined to be the most accurate prediction stage, as evidenced by an R^2 of 0.72 and an RMSE of 1290.24 kg/hm^2.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

YN: Conceptualization, Writing – original draft, Writing – review & editing, Data curation, Formal Analysis. XS: Writing – review & editing. HY: Investigation, Writing – review & editing. YZ: Methodology, Supervision, Writing – review & editing. JL: Formal Analysis, Writing – review & editing. WW: Project administration, Writing – review & editing. YS: Validation, Writing – review & editing. QM: Visualization, Writing – review & editing. JKL: Resources, Writing – review & editing. XL: Funding acquisition, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by Scientific research projects in higher education institutions of Anhui Province (no. 2023AH051855; 2022AH051623); Anhui Province Crop Intelligent Planting and Processing Technology Engineering Research Center Open Research Project (no. ZHKF202303); Natural Science Foundation of Hebei Province (no. C2023408010), and College Students ' Innovation and Entrepreneur ship Training Project (no. 202210879043; 202310879114; 202310879115).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1396183/full#supplementary-material

References

Abdi, H., and Williams, L. J. (2010). Principal component analysis. *WIREs Comput. Stat* 2, 433–459. doi: 10.1002/wics.101

Atzberger, C., Darvishzadeh, R., Immitzer, M., Schlerf, M., Skidmore, A., and le Maire, G. (2015). Comparative analysis of different retrieval methods for mapping grassland leaf area index using airborne imaging spectroscopy. *Int. J. Appl. Earth Observation Geoinformation* 43, 19–31. doi: 10.1016/j.jag.2015.01.009

Bansod, B., Singh, R., Thakur, R., and Singhal, G. (2017). A comparision between satellite based and drone based remote sensing technology to achieve sustainable development: a review. *J. Agric. Environ. Int. Dev. (JAEID)* 111, 383–407. doi: 10.12895/jaeid.20172.690

Becker, B. L., Lusch, D. P., and Qi, J. (2005). Identifying optimal spectral bands from in *situ* measurements of Great Lakes coastal wetlands using second-derivative analysis. *Remote Sens. Environ.* 97, 238–248. doi: 10.1016/j.rse.2005.04.020

Cao, C., Wang, T., Gao, M., Li, Y., Li, D., and Zhang, H. (2021a). Hyperspectral inversion of nitrogen content in maize leaves based on different dimensionality reduction algorithms. *Comput. Electron. Agric.* 190, 106461. doi: 10.1016/j.compag.2021.106461

Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Zhang, J., et al. (2021b). Integrating Multi-Source Data for Rice Yield Prediction across China using Machine Learning and Deep Learning Approaches. *Agric. For. Meteorol.* 297, 108275. doi: 10.1016/j.agrformet.2020.108275

Chang, K.-W., Shen, Y., and Lo, J.-C. (2005). Predicting rice yield using canopy reflectance measured at booting stage. *Agron. J.* 97, 872–878. doi: 10.2134/agronj2004.0162

Chen, Q., Meng, Z., Liu, X., Jin, Q., and Su, R. (2018). Decision variants for the automatic determination of optimal feature subset in RF-RFE. *Genes* 9, 301. doi: 10.3390/genes9060301

Chen, S., Hu, T., Luo, L., He, Q., Zhang, S., Li, M., et al. (2020). Rapid estimation of leaf nitrogen content in apple-trees based on canopy hyperspectral reflectance using multivariate methods. *Infrared Phys. Technol.* 111, 103542. doi: 10.1016/j.infrared.2020.103542

Cheng, T., Song, R., Li, D., Zhou, K., Zheng, H., Yao, X., et al. (2017). Spectroscopic estimation of biomass in canopy components of paddy rice using dry matter and chlorophyll indices. *Remote Sens.* 9, 319. doi: 10.3390/rs9040319

Cherkassky, V. (1997). The nature of statistical learning theory~. *IEEE Trans. Neural Netw.* 8, 1564–1564. doi: 10.1109/TNN.72

Clevers, J. G. P. W., de Jong, S. M., Epema, G. F., van der Meer, F., Bakker, W. H., Skidmore, A. K., et al. (2001). MERIS and the red-edge position. *Int. J. Appl. Earth Observation Geoinformation* 3, 313–320. doi: 10.1016/S0303-2434(01)85038-8

Datt, B. (1998). Remote Sensing of Chlorophyll a, Chlorophyll b, Chlorophyll a+b, and Total Carotenoid Content in Eucalyptus Leaves. *Remote Sens. Environ.* 66, 111–121. doi: 10.1016/S0034-4257(98)00046-7

Degenhardt, F., Seifert, S., and Szymczak, S. (2019). Evaluation of variable selection methods for random forests and omics data sets. *Briefings Bioinf.* 20, 492–503. doi: 10.1093/bib/bbx124

Demetriades-Shah, T. H., Steven, M. D., and Clark, J. A. (1990). High resolution derivative spectra in remote sensing. *Remote Sens. Environ.* 33, 55–64. doi: 10.1016/0034-4257(90)90055-Q

Devia, C. A., Rojas, J. P., Petro, E., Martinez, C., Mondragon, I. F., Patino, D., et al. (2019). High-throughput biomass estimation in rice crops using UAV multispectral imagery. *J. Intell. Robot Syst.* 96, 573–589. doi: 10.1007/s10846-019-01001-5

Fang, Q., Wang, G., Liu, T., Xue, B.-L., and Yinglan, A. (2018). Controls of carbon flux in a semi-arid grassland ecosystem experiencing wetland loss: Vegetation patterns and environmental variables. *Agric. For. Meteorol.* 259, 196–210. doi: 10.1016/j.agrformet.2018.05.002

Filella, I., and Penuelas, J. (1994). The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *Int. J. Remote Sens.* 15, 1459–1470. doi: 10.1080/01431169408954177

Forrester, J. B., and Kalivas, J. H. (2004). Ridge regression optimization using a harmonious approach. *J. Chemometrics* 18, 372–384. doi: 10.1002/cem.883

Fu, Y., Yang, G., Li, Z., Li, H., Li, Z., Xu, X., et al. (2020). Progress of hyperspectral data processing and modelling for cereal crop nitrogen monitoring. *Comput. Electron. Agric.* 172, 105321. doi: 10.1016/j.compag.2020.105321

Galvão, L. S., Formaggio, A. R., and Tisot, D. A. (2005). Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data. *Remote Sens. Environ.* 94, 523–534. doi: 10.1016/j.rse.2004.11.012

Gao, J., Meng, B., Liang, T., Feng, Q., Ge, J., Yin, J., et al. (2019). Modeling alpine grassland forage phosphorus based on hyperspectral remote sensing and a multi-factor machine learning algorithm in the east of Tibetan Plateau, China. *ISPRS J. Photogrammetry Remote Sens.* 147, 104–117. doi: 10.1016/j.isprsjprs.2018.11.015

Geladi, P., and Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta* 185, 1–17. doi: 10.1016/0003-2670(86)80028-9

Gerretzen, J., Szymańska, E., Jansen, J. J., Bart, J., Van Manen, H.-J., Van Den Heuvel, E. R., et al. (2015). Simple and effective way for data preprocessing selection based on design of experiments. *Anal. Chem.* 87, 12096–12103. doi: 10.1021/acs.analchem.5b02832

Gitelson, A. A., Merzlyak, M. N., and Lichtenthaler, H. K. (1996). Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. *J. Plant Physiol.* 148, 501–508. doi: 10.1016/S0176-1617(96)80285-9

Gnyp, M. L., Miao, Y., Yuan, F., Ustin, S. L., Yu, K., Yao, Y., et al. (2014). Hyperspectral canopy sensing of paddy rice aboveground biomass at different growth stages. *Field Crops Res.* 155, 42–55. doi: 10.1016/j.fcr.2013.09.023

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422. doi: 10.1023/A:1012487302797

Hansen, P. M., and Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens. Environ.* 86, 542–553. doi: 10.1016/S0034-4257(03)00131-7

He, J., Zhang, N., Su, X., Lu, J., Yao, X., Cheng, T., et al. (2019). Estimating leaf area index with a new vegetation index considering the influence of rice panicles. *Remote Sens.* 11, 1809. doi: 10.3390/rs11151809

Helland, I. S. (2001). Some theoretical aspects of partial least squares regression. *Chemometrics Intelligent Lab. Syst.* 58, 97–107. doi: 10.1016/S0169-7439(01)00154-X

Horler, D. N. H., Dockray, M., and Barber, J. (1983). The red edge of plant leaf reflectance. *Int. J. Remote Sens.* 4, 273–288. doi: 10.1080/01431168308948546

Ivanda, A., Šerić, L., Bugarić, M., and Braović, M. (2021). Mapping chlorophyll-a concentrations in the kaštela bay and brač Channel using ridge regression and sentinel-2 satellite images. *Electronics* 10, 3004. doi: 10.3390/electronics10233004

Ji, S., Gu, C., Xi, X., Zhang, Z., Hong, Q., Huo, Z., et al. (2022). Quantitative monitoring of leaf area index in rice based on hyperspectral feature bands and ridge regression algorithm. *Remote Sens.* 14, 2777. doi: 10.3390/rs14122777

Jia, Y., Zhang, H., Zhang, X., and Su, Z. (2022). Quantitative analysis and hyperspectral remote sensing inversion of rice canopy spad in A cold region. *Eng. Agríc.* 42, e20220030. doi: 10.1590/1809-4430-eng.agric.v42n4e20220030/2022

Kanke, Y., Tubaña, B., Dalen, M., and Harrell, D. (2016). Evaluation of red and red-edge reflectance-based vegetation indices for rice biomass and grain yield prediction models in paddy fields. *Precis. Agric.* 17, 507–530. doi: 10.1007/s11119-016-9433-1

Kawamura, K., Ikeura, H., Phongchanmaixay, S., and Khanthavong, P. (2018). Canopy Hyperspectral Sensing of Paddy Fields at the Booting Stage and PLS Regression can Assess Grain Yield. *Remote Sens.* 10, 1249. doi: 10.3390/rs10081249

Kawano, S., Fujisawa, H., Takada, T., and Shiroishi, T. (2018). Sparse principal component regression for generalized linear models. *Comput. Stat Data Anal.* 124, 180–196. doi: 10.1016/j.csda.2018.03.008

Kokaly, R. F., and Clark, R. N. (1999). Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sens. Environ.* 67, 267–287. doi: 10.1016/S0034-4257(98)00084-4

Krishnan, P., Ramakrishnan, B., Reddy, K. R., and Reddy, V. R. (2011). Chapter three - high-temperature effects on rice growth, yield, and grain quality. *Adv. Agron.* 111, 87–206. doi: 10.1016/B978-0-12-387689-8.00004-7

Kursa, M. B., Jankowski, A., and Rudnicki, W. R. (2010). Boruta – A system for feature selection. *Fundamenta Informaticae* 101, 271–285. doi: 10.3233/FI-2010-288

Kursa, M. B., and Rudnicki, W. R. (2010). Feature selection with the boruta package. *J. Stat. Software* 36, 1–13. doi: 10.18637/jss.v036.i11

Li, Z., Chen, Z., Cheng, Q., Duan, F., Sui, R., Huang, X., et al. (2022). UAV-based hyperspectral and ensemble machine learning for predicting yield in winter wheat. *Agronomy* 12, 202. doi: 10.3390/agronomy12010202

Li, B., and Xie, W. (2015). Adaptive fractional differential approach and its application to medical image enhancement. *Comput. Electrical Eng.* 45, 324–335. doi: 10.1016/j.compeleceng.2015.02.013

Liang, L., Di, L., Huang, T., Wang, J., Lin, L., Wang, L., et al. (2018). Estimation of leaf nitrogen content in wheat using new hyperspectral indices and a random forest regression algorithm. *Remote Sens.* 10, 1940. doi: 10.3390/rs10121940

Lin, X., Ren, C., Li, Y., Yue, W., Liang, J., and Yin, A. (2023). Eucalyptus plantation area extraction based on SLPSO-RFE feature selection and multi-temporal sentinel-1/2 data. *Forests* 14, 1864. doi: 10.3390/f14091864

Liu, G.-S., Guo, H.-S., Pan, T., Wang, J.-H., and Cao, G. (2014). Vis-NIR spectroscopic pattern recognition combined with SG smoothing applied to breed screening of transgenic sugarcane. *Guang Pu Xue Yu Guang Pu Fen Xi* 34, 2701–2706.

Lucas, R., Van De Kerchove, R., Otero, V., Lagomasino, D., Fatoyinbo, L., Omar, H., et al. (2020). Structural characterisation of mangrove forests achieved through combining multiple sources of remote sensing data. *Remote Sens. Environ.* 237, 111543. doi: 10.1016/j.rse.2019.111543

Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., and Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends Analytical Chem.* 132, 116045. doi: 10.1016/j.trac.2020.116045

Moore, B. (1981). Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Trans. Automatic Control* 26, 17–32. doi: 10.1109/TAC.1981.1102568

Pan, L., Li, H.-C., Deng, Y.-J., Zhang, F., Chen, X.-D., and Du, Q. (2017). Hyperspectral dimensionality reduction by tensor sparse and low-rank graph-based discriminant analysis. *Remote Sens.* 9, 452. doi: 10.3390/rs9050452

Paul, J., D'Ambrosio, R., and Dupont, P. (2015). Kernel methods for heterogeneous feature selection. *Neurocomputing* 169, 187–195. doi: 10.1016/j.neucom.2014.12.098

Poona, N. K., Niekerk, A., Nadel, R. L., and Ismail, R. (2016). Random forest (RF) wrappers for waveband selection and classification of hyperspectral data. *Appl. Spectrosc. AS* 70, 322–333. doi: 10.1177/0003702815620545

Prasad, R., Deo, R. C., Li, Y., and Maraseni, T. (2019). Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. *CATENA* 177, 149–166. doi: 10.1016/j.catena.2019.02.012

Raja, S. P., Sawicka, B., Stamenkovic, Z., and Mariammal, G. (2022). Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers. *IEEE Access* 10, 23625–23641. doi: 10.1109/ACCESS.2022.3154350

Ren, H., Zhou, G., and Zhang, F. (2018). Using negative soil adjustment factor in soil-adjusted vegetation index (SAVI) for aboveground living biomass estimation in arid grasslands. *Remote Sens. Environ.* 209, 439–445. doi: 10.1016/j.rse.2018.02.068

Savitzky, A., and Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36, 1627–1639. doi: 10.1021/ac60214a047

Schafer, R. W. (2011). What is a savitzky-golay filter? [Lecture notes]. *IEEE Signal Process. Magazine* 28, 111–117. doi: 10.1109/MSP.2011.941097

Schino, G., Borfecchia, F., De Cecco, L., Dibari, C., Iannetta, M., Martini, S., et al. (2003). Satellite estimate of grass biomass in a mountainous range in central Italy. *Agroforestry Syst.* 59, 157–162. doi: 10.1023/A:1026308928874

Shafaei, M., and Kisi, O. (2017). Predicting river daily flow using wavelet-artificial neural networks based on regression analyses in comparison with artificial neural networks and support vector machine models. *Neural Comput. Applic* 28, 15–28. doi: 10.1007/s00521-016-2293-9

Shen, L., Gao, M., Yan, J., Wang, Q., and Shen, H. (2022). Winter wheat SPAD value inversion based on multiple pretreatment methods. *Remote Sens.* 14, 4660. doi: 10.3390/rs14184660

Shi, X., Yao, L., and Pan, T. (2021). Visible and near-infrared spectroscopy with multi-parameters optimization of savitzky-golay smoothing applied to rapid analysis of soil cr content of pearl river delta. *J. Geosci. Environ. Prot.* 09, 75. doi: 10.4236/gep.2021.93006

Smith, G., and Campbell, F. (1980). A critique of some ridge regression methods. *J. Am. Stat. Assoc.* 75, 74–81. doi: 10.1080/01621459.1980.10477428

Spiertz, J. H. J., and Ewert, F. (2009). Crop production and resource use to meet the growing demand for food, feed and fuel: opportunities and constraints. *NJAS: Wageningen J. Life Sci.* 56, 281–300. doi: 10.1016/S1573-5214(09)80001-8

Sun, H., Feng, M., Xiao, L., Yang, W., Ding, G., Wang, C., et al. (2021). Potential of multivariate statistical technique based on the effective spectra bands to estimate the

plant water content of wheat under different irrigation regimes. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.631573

Suryakala, S. V., and Prince, S. (2019). Investigation of goodness of model data fit using PLSR and PCR regression models to determine informative wavelength band in NIR region for non-invasive blood glucose prediction. *Opt Quant Electron* 51, 271. doi: 10.1007/s11082-019-1985-7

Suykens, J. A. K., and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Process. Lett.* 9, 293–300. doi: 10.1023/A:1018628609742

Takai, T., Matsuura, S., Nishio, T., Ohsumi, A., Shiraiwa, T., and Horie, T. (2006). Rice yield potential is closely related to crop growth rate during late reproductive period. *Field Crops Res.* 96, 328–335. doi: 10.1016/j.fcr.2005.08.001

Tong, X., Duan, L., Liu, T., Yang, Z., Wang, Y., and Singh, V. P. (2023). Estimation of grassland aboveground biomass combining optimal derivative and raw reflectance vegetation indices at peak productive growth stage. *Geocarto Int.* 38, 2186497. doi: 10.1080/10106049.2023.2186497

Tunca, E., Köksal, E. S., Öztürk, E., Akay, H., and Çetin Taner, S. (2023). Accurate estimation of sorghum crop water content under different water stress levels using machine learning and hyperspectral data. *Environ. Monit Assess.* 195, 877. doi: 10.1007/s10661-023-11536-8

Wang, X., Huang, J., Li, Y., and Wang, R. (2004). Study on hyperspectral remote sensing estimation models about aboveground fresh biomass of rice. *Remote Sens. Modeling Ecosyst. Sustainability (SPIE)*, 336–345. doi: 10.1117/12.557403

Wang, L., Zhou, X., Zhu, X., Dong, Z., and Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Crop J.* 4, 212–219. doi: 10.1016/j.cj.2016.01.008

Xu, T., Wang, F., Shi, Z., Xie, L., and Yao, X. (2023). Dynamic estimation of rice aboveground biomass based on spectral and spatial information extracted from hyperspectral remote sensing images at different combinations of growth stages. *ISPRS J. Photogrammetry Remote Sens.* 202, 169–183. doi: 10.1016/j.isprsjprs.2023.05.021

Xu, T., Wang, F., Xie, L., Yao, X., Zheng, J., Li, J., et al. (2022). Integrating the textural and spectral information of UAV hyperspectral images for the improved estimation of rice aboveground biomass. *Remote Sens.* 14, 2534. doi: 10.3390/rs14112534

Yang, C.-M., and Chen, R.-K. (2004). Modeling rice growth with hyperspectral reflectance data. *Crop Sci.* 44, 1283–1290. doi: 10.2135/cropsci2004.1283

Yang, S., Feng, Q., Liang, T., Liu, B., Zhang, W., and Xie, H. (2018). Modeling grassland above-ground biomass based on artificial neural network and remote sensing in the Three-River Headwaters Region. *Remote Sens. Environ.* 204, 448–455. doi: 10.1016/j.rse.2017.10.011

Yang, X., Huang, J., Wu, Y., Wang, J., Wang, P., Wang, X., et al. (2011). Estimating biophysical parameters of rice with remote sensing data using support vector machines. *Sci. China Life Sci.* 54, 272–281. doi: 10.1007/s11427-011-4135-4

Yan-lin, T., Jing-feng, H., and Ren-chao, W. (2004). Change law of hyperspectral data in related with chlorophyll and carotenoid in rice at different developmental stages. *Rice Sci.* 11, 274–282.

Yoosefzadeh-Najafabadi, M., Earl, H. J., Tulpan, D., Sulik, J., and Eskandari, M. (2021). Application of machine learning algorithms in plant breeding: predicting yield from hyperspectral reflectance in soybean. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.624273

Yu, F., Bai, J., Jin, Z., Zhang, H., and Xu, T. (2023). Inversion of rice leaf biomass based on PROSAIL model optimization. *J. Huazhong Agric. Univ. (Natural Sci. Edition) (English Edition)* 42, 187–194. doi: 10.13300/j.cnki.hnlkxb.2023.03.022

Yu, F., Feng, S., Du, W., Wang, D., Guo, Z., Xing, S., et al. (2020). A study of nitrogen deficiency inversion in rice leaves based on the hyperspectral reflectance differential. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.573272

Yue, J., Yang, G., Tian, Q., Feng, H., Xu, K., and Zhou, C. (2019). Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices. *ISPRS J. Photogrammetry Remote Sens.* 150, 226–244. doi: 10.1016/j.isprsjprs.2019.02.022

Zhang, K., Ge, X., Shen, P., Li, W., Liu, X., Cao, Q., et al. (2019). Predicting rice grain yield based on dynamic changes in vegetation indexes during early to mid-growth stages. *Remote Sens.* 11, 387. doi: 10.3390/rs11040387

Zhao, M., Gao, Y., Lu, Y., and Wang, S. (2022). Hyperspectral modeling of soil organic matter based on characteristic wavelength in east China. *Sustainability* 14, 8455. doi: 10.3390/su14148455

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 10 May 2024
DOI 10.3389/fpls.2024.1404238

Check for updates

OPEN ACCESS

EDITED BY
Liujun Xiao,
Nanjing Agricultural University, China

REVIEWED BY
Min Kang,
Nanjing Agricultural University, China
Jinling Zhao,
Anhui University, China

*CORRESPONDENCE
Xinwei Li
lixw@ahstu.edu.cn
Yousef Alhaj Hamoud
Yousef-hamoud11@hhu.edu.cn

RECEIVED 20 March 2024
ACCEPTED 17 April 2024
PUBLISHED 10 May 2024

CITATION
Su X, Nian Y, Shaghaleh H, Hamad A, Yue H, Zhu Y, Li J, Wang W, Wang H, Ma Q, Liu J, Li X and Alhaj Hamoud Y (2024) Combining features selection strategy and features fusion strategy for SPAD estimation of winter wheat based on UAV multispectral imagery. *Front. Plant Sci.* 15:1404238. doi: 10.3389/fpls.2024.1404238

COPYRIGHT
© 2024 Su, Nian, Shaghaleh, Hamad, Yue, Zhu, Li, Wang, Wang, Ma, Liu, Li and Alhaj Hamoud. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Combining features selection strategy and features fusion strategy for SPAD estimation of winter wheat based on UAV multispectral imagery

Xiangxiang Su[1], Ying Nian[1], Hiba Shaghaleh[2], Amar Hamad[2], Hu Yue[3], Yongji Zhu[1], Jun Li[1], Weiqiang Wang[1], Hong Wang[1], Qiang Ma[1], Jikai Liu[1,3], Xinwei Li[1,3,4]* and Yousef Alhaj Hamoud[5]*

[1]College of Resource and Environment, Anhui Science and Technology University, Fengyang, China, [2]College of Environmental, Hohai University, Nanjing, China, [3]Anhui Engineering Research Center of Smart Crop Planting and Processing Technology, Fengyang, China, [4]Anhui Province Agricultural Waste Fertilizer Utilization and Cultivated Land Quality Improvement Engineering Research Center, Anhui Science and Technology University, Fengyang, China, [5]College of Hydrology and Water Resources, Hohai University, Nanjing, China

The Soil Plant Analysis Development (SPAD) is a vital index for evaluating crop nutritional status and serves as an essential parameter characterizing the reproductive growth status of winter wheat. Non-destructive and accurate monitorin3g of winter wheat SPAD plays a crucial role in guiding precise management of crop nutrition. In recent years, the spectral saturation problem occurring in the later stage of crop growth has become a major factor restricting the accuracy of SPAD estimation. Therefore, the purpose of this study is to use features selection strategy to optimize sensitive remote sensing information, combined with features fusion strategy to integrate multiple characteristic features, in order to improve the accuracy of estimating wheat SPAD. This study conducted field experiments of winter wheat with different varieties and nitrogen treatments, utilized UAV multispectral sensors to obtain canopy images of winter wheat during the heading, flowering, and late filling stages, extracted spectral features and texture features from multispectral images, and employed features selection strategy (Boruta and Recursive Feature Elimination) to prioritize sensitive remote sensing features. The features fusion strategy and the Support Vector Machine Regression algorithm are applied to construct the SPAD estimation model for winter wheat. The results showed that the spectral features of NIR band combined with other bands can fully capture the spectral differences of winter wheat SPAD during the reproductive growth stage, and texture features of the red and NIR band are more sensitive to SPAD. During the heading, flowering, and late filling stages, the stability and estimation accuracy of the SPAD model constructed using both features selection strategy and features fusion strategy are superior to models using only a single feature strategy or no strategy. The enhancement of model accuracy by this method becomes more significant, with the greatest improvement observed during the late filling stage, with R^2 increasing by 0.092-0.202, root mean squared error (RMSE) decreasing

by 0.076-4.916, and ratio of performance to deviation (RPD) increasing by 0.237-0.960. In conclusion, this method has excellent application potential in estimating SPAD during the later stages of crop growth, providing theoretical basis and technical support for precision nutrient management of field crops.

KEYWORDS

UAV, winter wheat, SPAD, features selection strategy, features fusion strategy

1 Introduction

As a principal food crop worldwide, wheat provides vital carbohydrate sustenance to millions of people, holding an essential position in addressing global food security (Asseng et al., 2011). The Soil Plant Analysis Development (SPAD) indicator, which represents the relative chlorophyll content in leaves (Netto et al., 2005), is intimately connected with crop nutritional status, serving as a significant index of crop growth and development (Gitelson et al., 2005; Zhang et al., 2019; Liu et al., 2020). Consequently, the accurate and efficient acquisition of SPAD information for winter wheat is of paramount importance for field management decisions and crop nutrition monitoring. The conventional approaches for obtaining crop chlorophyll content primarily rely on destructive field sampling and laborious chemical analysis in the laboratory. These approaches are not only time-consuming and labor-intensive but also incur significant costs (Sampson et al., 2003). Existing research has demonstrated a close correlation between data obtained from the handheld SPAD-502 chlorophyll meter and the chlorophyll content analyzed chemically in a laboratory setting (Rigon et al., 2016). Although the SPAD acquisition method offers more convenience than traditional chlorophyll measurement techniques and allows for non-destructive sampling (Uddling et al., 2007), its limitations in measurement points entail certain labor costs, resulting in small operational scales, low efficiency, and difficulty in meeting the demands for large-scale acquisition of crop chlorophyll content information (Wang J et al., 2021; Wu et al., 2023). Therefore, there is a pressing need for a method that can rapidly, non-destructively, and accurately obtain crop SPAD information.

High-throughput phenotyping (HTP) technology has made this work possible, especially the unmanned aerial vehicle (UAV) phenotyping platform equipped with high-performance image sensors (Guo et al., 2021). Its advantages of high spatio-temporal resolution, flexibility, and easy operation have been widely applied to crop growth monitoring (Liu J et al., 2022; Yin et al., 2022; Fan et al., 2023). With the development of sensor technology, RGB sensors (Vélez et al., 2022), multispectral sensors (Impollonia et al., 2022), hyperspectral sensors (Zhou X et al., 2023), and lidar (Jimenez-Berni et al., 2018) have provided substantial data support for crop phenotyping, intelligent breeding, and more (Jin et al., 2015). Although RGB sensors are cost-effective, the spectral information they can acquire is limited, and the limited number of spectral bands often struggle to accurately capture the complete physiological and biochemical information of crops (Feng et al., 2022). Hyperspectral sensors and lidar offer high data accuracy, but due to their high cost, they are not suitable for widespread application in field crop monitoring (Madec et al., 2017; Fan et al., 2022). Therefore, the use of a lightweight, cost-effective UAV equipped with a multispectral sensor to capture remote sensing information of the winter wheat canopy for SPAD estimation becomes particularly essential.

Currently, spectral features of UAV imagery (vegetation indices, bands reflectance) have been extensively utilized in the field of precision agriculture for estimating crop physiological parameters (Luo et al., 2022), such as Leaf Area Index (LAI) (Qiao et al., 2022; Zhou et al., 2022), Above-Ground Biomass (AGB) (Han et al., 2022; Yue et al., 2023), SPAD (Wang J et al., 2021;Yang et al., 2021), and nitrogen content (Yang et al., 2019; Xu et al., 2023). However, under high canopy cover in the later stages of crop growth, vegetation indices can be affected by spectral saturation effects (Mutanga and Skidmore, 2004; Yue et al., 2019; Mutanga et al., 2023). Moreover, during the reproductive growth stage of crops, the emergence of crop spikes further increases the spectral mixing effect, reducing the sensitivity of vegetation indices (Jay et al., 2019; Wang W et al., 2021). It is challenging to establish a reliable SPAD estimation model using only spectral features during the reproductive growth stage. Texture features, as remote sensing information reflecting image grayscale attributes, color, and spatial structure, can not only characterize differences in canopy image spatial structure arrangement caused by crop growth and development (Liu et al., 2022a) but also reflect changes in crop leaf color due to different treatments such as varieties, fertilizers, and density (Li R. et al., 2022). Texture features have the potential to be combined with spectral features (Fu et al., 2021), and features fusion can reduce the impact of spectral saturation effects to enhance the accuracy of estimation models (Li et al., 2020). However, with the increase of features, while improving model accuracy, there may be information redundancy (Zheng et al., 2020), increasing data complexity. Therefore, to eliminate this redundancy, selecting appropriate features is extremely necessary for optimizing model structure and accuracy.

Features selection plays a pivotal role in machine learning, as it directly determines the input features of the prediction model, influencing its performance and effectiveness. RFE method is widely used in features selection nowadays (Zhu et al., 2020; Shao et al., 2022; Yang et al., 2022). This method is based on multiple rounds of training using a specific algorithm model. It constructs new feature subsets by removing particular variables and conducts the next round of training based on the new feature set (Chen et al., 2021; Rasel et al., 2021). This process is repeated until the model performs optimally. In recent years, the Boruta features selection method has quickly gained popularity due to its low operational cost and speed (Shu et al., 2021; Tatsumi et al., 2021; Shu et al., 2023). The goal of Boruta is to filter out all feature sets related to the target parameter. It allows a more comprehensive understanding of the factors influencing the target parameter, thereby effectively performing features selection (Kursa and Rudnicki, 2010).

Considering the impact of spectral saturation in the later stages of crop growth, features selection and features fusion have the potential to enhance the performance of prediction models. Identifying the remote sensing features that influence the SPAD of winter wheat in the later stages of growth and jointly constructing a high-accuracy SPAD model will be a key focus of future research. Therefore, this study utilizes a UAV equipped with a multispectral sensor to capture high-resolution images of the winter wheat canopy, and employs image processing technology to extract the spectral and texture features of the wheat, and utilizes two features selection methods (Boruta and RFE) to explore the potential of spectral and texture features in predicting SPAD during the later stages of winter wheat growth. We also develop a winter wheat SPAD monitoring model that combines features selection and features fusion strategy. The objectives of this study are: (1) to identify the spectral and texture features sensitive to SPAD during the reproductive growth stage of winter wheat; (2) to evaluate the performance of the winter wheat SPAD prediction model under features selection strategy; and (3) to explore the potential of combining features selection and features fusion strategy to estimate SPAD in the later stages of winter wheat growth.

2 Materials and methods

2.1 Field experimental design

The winter wheat experiment took place in Chuzhou City, located in Anhui Province, China (32°48′52″N,117°46′7″E) from 2020 to 2021 (Figure 1A). This region, positioned in the mid-lower reaches of the Yangtze River, experiences a subtropical monsoon climate. It is characterized by a moist weather pattern, distinct four seasons, an average annual temperature of 15.4°C, and an annual rainfall between 1000-1100mm. The region records an average of 144 rainy days per year and approximately 210 frost-free days annually.

The field experiment consisted of 36 plots, each measuring 16m² (2×8 m). It involved three winter wheat varieties with stable high-yield potential (V1: Huaimai 44, V2: Yannong 999, and V3: Ningmai 13) and four levels of nitrogen fertilizer application (N0-N3: 0, 100, 200, 300 kg/ha), with each treatment replicated three times (Figure 1C). The fertilization scheme involved the application of phosphorus (P=90kg/ha) and potassium (K=135 kg/ha) as base fertilizers before sowing, with a 6:4 ratio of nitrogen fertilizer applied pre-sowing and at the jointing stage. The sowing date was November 7, 2020, with the winter wheat manually strip-sown at a row spacing of 30 cm. The harvest took place on June 3, 2021, and

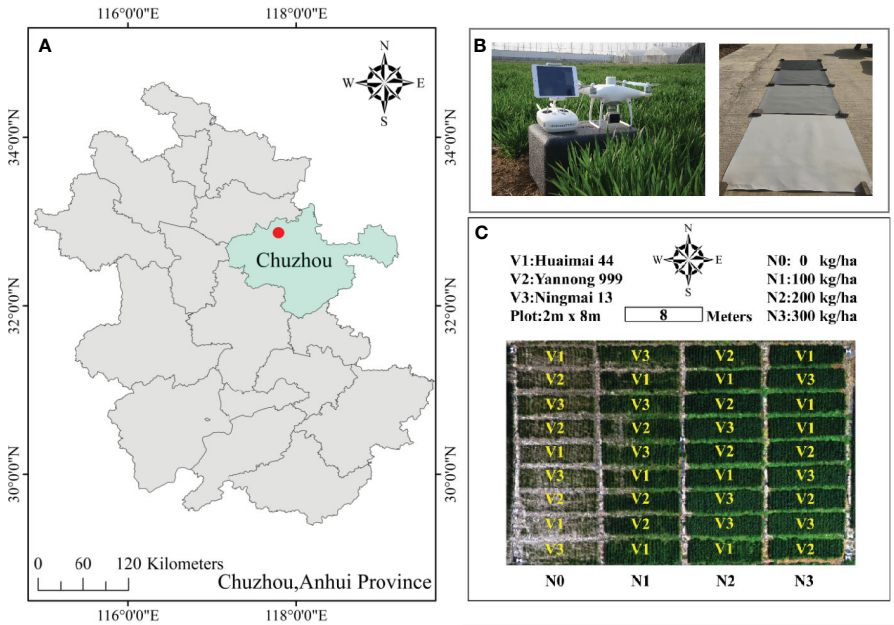


FIGURE 1
Study area **(A)**. The experiment field was in Chuzhou, Anhui province, China. **(B)**, the unmanned aerial vehicle (UAV) for acquiring low-altitude remote sensing data and plates that correct 5%, 10%, 20% and 40% reflectivity of calibrate differences and **(C)** field experimental design.

other field management measures were based on local high-yield cultivation practices. Throughout the entire growth stage of the winter wheat, there were no occurrences of diseases, pests, or weed hazards, and the field environment was good, with no instances of drought or waterlogging.

2.2 UAV-based data collection and pre-processing

The experiment utilized a Phantom 4 Multispectral (P4M) RTK (DJI Technology Co., Shenzhen, China) UAV to collect multispectral (MS) data during the wheat heading stage on April 18, 2021, the flowering stage on April 29, 2021, and the late filling stage on May 24, 2021. The UAV is equipped with five monochrome sensors, each with 2.08 million pixels: Blue, Green, Red, Red edge, and NIR, with center wavelengths of 450 ± 16nm, 560 ± 16nm, 650 ± 16nm, 730 ± 16nm, and 840 ± 26nm, and bandwidths of 20nm, 20nm, 10nm, 10nm, 40nm, respectively. It is also equipped with a real-time kinematic (RTK) system, which allows for vertical positioning accuracy of ±1.5 cm and horizontal positioning accuracy of ±1cm, enabling the P4M to obtain high-precision spectral and texture information (Oktay and Coban, 2017). Flights were conducted between 11:00 and 13:00 on clear, windless days with stable solar radiation intensity. Four radiometric calibration panels with known reflectance were deployed on the ground for radiometric calibration during subsequent data processing (Figure 1B). The DJI GS PRO software (https://www.dji.com/cn/ground-station-pro/) was used to pre-plan the flight route, relying on the UAV's autopilot system to execute the predefined flight plan (Şahin et al., 2022). Each flight lasted 20 minutes, with a flight altitude of 30 m, a flight speed of 2m/s, forward and side overlap of 90% and 85% respectively, and an image resolution of 1600×1300 pixels.

The UAV MS images collected at each growth stage were imported into the PIX4Dmapper software (version 4.4.12, Pix4D SA, Prilly, Switzerland) for image stitching. The initial step involved aligning the images using the feature point matching algorithm. Subsequently, a dense point cloud and texture mesh were generated by utilizing both the UAV image data and location information. To improve data quality, this study carried out radiometric correction on the MS images using the empirical line method (ELM) and the image information of ground radiometric calibration panels with known reflectance (Di Gennaro et al., 2022). A shapefile was generated using ArcGIS 10.2 (Environmental Systems Research Institute, Inc, RedLands, CA, USA), and plot areas were delineated based on the orthoimage of the experimental area.

2.3 SPAD data collection

Field measurements at each stage were conducted prior to the UAV flights on the same day, with chlorophyll values based on SPAD determined for winter wheat across the 36 plots. Winter wheat SPAD was measured using a SPAD-502 chlorophyll meter (Konica Minolta Optics Inc., Osaka, Japan) (Jiang et al., 2022). For each plot, three average-growth wheat plants were selected, and SPAD readings were taken at the 1/6, 3/6, and 5/6 length of the flag leaves. The average of the 9 SPAD values in the sample plot was used as the actual SPAD measurement in the sample plot. A total of 324 times were collected in each growth stage, and the measured SPAD data of 36 sample plots were calculated.

2.4 Extraction of features from UAV imagery

2.4.1 Spectral features extraction

This study used band reflectance and vegetation indices as spectral features (SF). Vegetation indices are obtained by calculations or combinations of characteristic bands, providing robust vegetation information factors and enhancing, to a certain extent, the expressive ability of remote sensing data (Bai et al., 2022). The reflectance of the five original bands from the wheat canopy during the heading, flowering, and late filling stages was obtained from the MS images for vegetation index calculation. This study selected a total of 30 widely-used spectral features for monitoring crop growth and evaluating parameters These features include five original bands (I), seven with only the visible bands (II), five with the RE band but without NIR band (III), eight with the NIR band but without RE band (IV), and five with both the NIR and RE bands (V) (Table 1).

2.4.2 Texture features extraction

As a common means of characterizing image information, texture reflects important information such as surface structure and spatial arrangement in the image without relying on brightness. It generally uses high-energy narrow peaks in the spectrum to detect the periodicity of the image (Bharati et al., 2004; Pu and Cheng, 2015). This study uses the Gray Level Co-occurrence Matrices (GLCM) method to extract the texture information of B, G, R, RE and NIR bands in winter wheat canopy MS images. GLCM requires user-defined parameters such as window size and orientation. The parameters for extracting GLCM textures in this study are as follows: a window size of 3 pixels by 3 pixels, an extraction direction set at 45°, and grayscale quantization levels defaulted to 64. The texture information includes Mean (Me), Variance (Va), Homogeneity (Ho), Contrast (Cn), Dissimilarity (Di), Entropy (En), Second moment (Se) and Correlation (Cr), a total of eight indicators (Table 2). In order to obtain more texture information for feature selection, the maximum value (MAX), minimum value (MIN), mean value (MEAN) and standard deviation (SD) of the GLCM index in each sample area were calculated, resulting in a total of 5(bands) × 8(GLCM indices) × 4(statistical metrics) = 160 texture metrics.

2.5 Features selection strategy and features fusion strategy

UAV multispectral remote sensing data was analyzed in order to identify the most effective approach to maximize improvements in SPAD estimates during later stages of growth (Figure 2). The

TABLE 1 Types and formulas of spectral features extracted from MS images based on UAV.

Component	Variable	Formulation	References
The original bands	Blue	B	–
	Green	G	–
	Red	R	–
	Red edge	RE	–
	Near-infrared	NIR	–
With only the Visible bands	CI	(R-B)/R	(Kumar et al., 2015)
	NGBDI	(G-B)/(G+B)	(Hunt et al., 2005)
	RBI	R/B	(Peña Barragán et al., 2007)
	NPCI	(R-B)/(R+B)	(Peñuelas et al., 1994)
	WI	(G-B)/(R-G)	(Woebbecke et al., 1995)
	ExB	1.4*B-G	(Mao et al., 2003)
	GBI	G/B	(Sellaro et al., 2010)
With the Red edge band but without NIR band	PSRI	(R-G)/RE	(Ren et al., 2017)
	LIC3	B/RE	(Lichtenthaler et al., 1996)
	DCabCxc	R/G*3*RE	(Poblete et al., 2020)
	GM1	RE/G	(Gitelson and Merzlyak, 1996)
	NDREI	(RE-G)/(RE+G)	(Hassan et al., 2018)
With the NIR band but without Red edge band	DVI	NIR-R	(Richardson and Everitt, 1992)
	mSR	(NIR-B)/(RE-B)	(Datt, 1999)
	CVI	NIR*R/(G*G)	(Datt et al., 2003)
	SIPI	(NIR-B)/(NIR-R)	(Hunt et al., 2011)
	GCVI	NIR/G-1	(Gitelson et al., 2003)
	ARI3	NIR*(1/G-1/R)	(Gitelson et al., 2001)
	BDVI	NIR-B	(de Castro et al., 2012)
	NDVI	(NIR-R)/(NIR+R)	(Bannari et al., 1995)
With the NIR band and Red edge band	NDRE	(NIR-RE)/(NIR+RE)	(Le Maire et al., 2008)
	LCI	(NIR-RE)/(NIR+R)	(Hunt et al., 2011)
	MTCI	(NIR-RE)/(RE-R)	(Jurgens, 1997)
	DATT	(NIR-RE)/(NIR-R)	(Datt, 1999)
	mNDblue	(B-RE)/(NIR+B)	(Jay et al., 2017)

(Continued)

TABLE 1 Continued

features selection strategy is used to select SF and TF that are sensitive to SPAD. SPAD prediction model is further constructed to check the difference between implementing and not implementing the features selection strategy, different types of feature subsets are fused in the features fusion strategy to construct SPAD estimation model together.

2.5.1 Features selection

In this study, two features selection methods were selected as features selection strategy: Recursive feature elimination (RFE) and Boruta. The RFE algorithm is a greedy search algorithm that incrementally selects the optimal features based on the model's performance. Initially, RFE requires the selection of a specific

TABLE 2 Formulas of texture features extracted from MS images based on GLCM.

Texture features	Name	Formula	Description
Me	Mean	$Me = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} i * \mathrm{p}(i,j)$	The mean value in the texture
Va	Variance	$Va = \sum_i\sum_j (i-u)^2 \mathrm{p}(i,j)$	the size of the texture change
Ho	Homogeneity	$Ho = \sum_i\sum_j \frac{1}{1+(i-j)^2}\mathrm{p}(i,j)$	The homogeneity of grey level in the texture
Cn	Contrast	$Cn = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \mathrm{p}(i,j) \right\}_{\lvert i-j\rvert = n}$	The clarity in the texture
Di	Dissimilarity	$Di = \sum_{n=1}^{N_g-1} n \left\{ \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \mathrm{p}(i,j) \right\}_{\lvert i-j\rvert = n}$	Same as contrast, The similarity of the pixels in the texture
En	Entropy	$En = -\sum_i\sum_j \mathrm{p}(i,j)\log(\mathrm{p}(i,j))$	The diversity of the pixels in the texture
Se	Second moment	$Se = \sum_i\sum_j \{p(i,j)\}^2$	The uniformity of greyscale in the texture
Cr	Correlation	$Cr = \frac{\sum_i\sum_j (i,j)\mathrm{p}(i,j) - \mu_i\mu_j}{\sigma_i\sigma_j}$	The consistency in the texture

The parameters u_i, u_j, σ_i, and σ_j represent the average and standard deviation of the row and column sums of the GLCM.

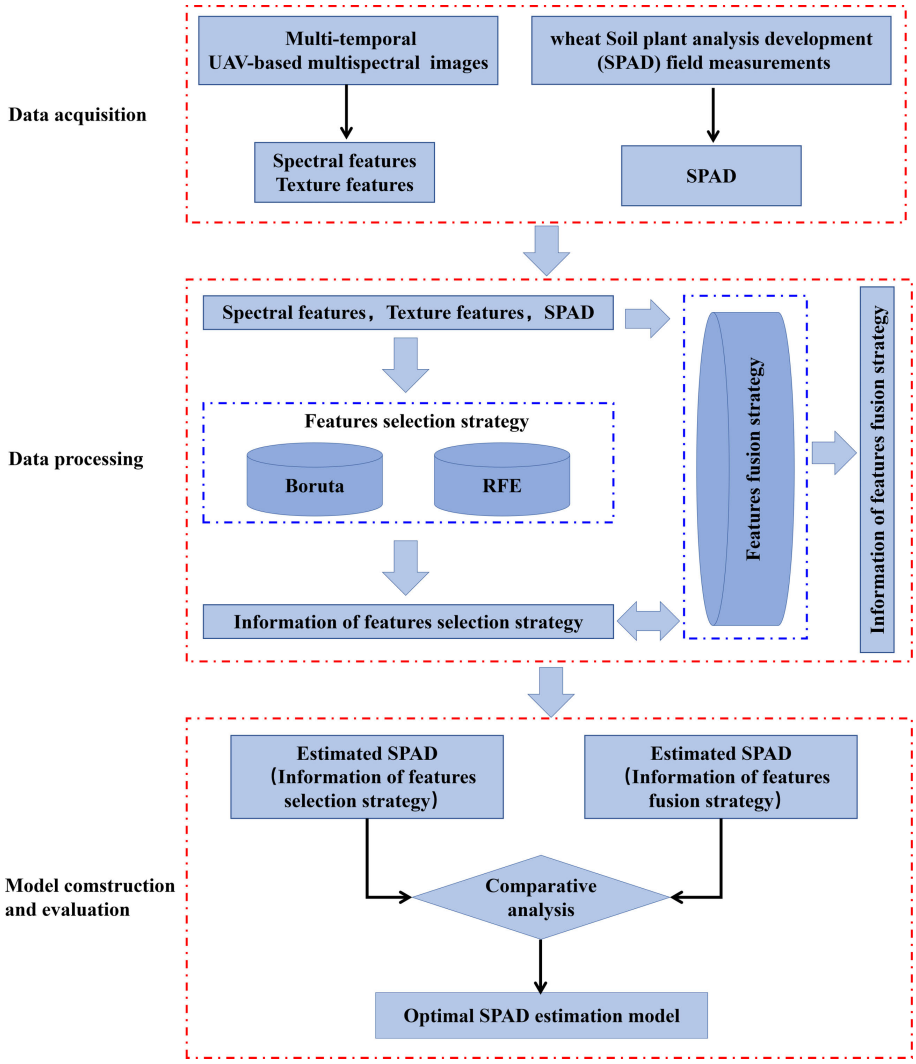


FIGURE 2
Flowchart of features selection strategy and features fusion strategy applied in this experimental study.

machine learning model to serve as the estimator, such as linear regression, support vector machines, and random forests (RF). The RFE algorithm first uses all features to train the model, calculates the importance of each feature and sorts it, and uses each feature subset to train the model, compares the model results obtained from each subset, and then recalculates the model based on specific features. Train the model and repeat this process until the optimal feature combination is selected to maximize model performance (Kuhn, 2008). Ultimately, RFE yields a subset of features that the algorithm deems the most representative and significant. By employing this recursive approach, RFE systematically eliminates irrelevant or redundant features while retaining the most pertinent ones, thereby enhancing the model's generalizability and efficacy. The essence of the RFE algorithm lies in iteratively training the model and selecting features to identify an optimal subset, thereby optimizing the model's performance and interpretability.

The Boruta algorithm is a wrapper based on the RF algorithm (Kursa and Rudnicki, 2010). Therefore, Boruta has the advantages of the random forest algorithm, has low running time cost, and can run results without relying on parameter adjustment. It is an ensemble method where multiple independent decision trees vote for classification, classify all trees based on a given attribute and calculate the importance of all trees, i.e. the Z score, which reflects the fluctuations in accuracy between trees in the forest. When Boruta is running, it will create "shadow" attributes obtained by reshuffling the original attributes, and randomly disrupt the order of feature parameters. When calculating feature importance, the feature parameters are divided into three categories, namely Z score and significant features with a Z score higher than the "Shadow" attribute are called "Confirmed" (important), features with a Z score close to the "Shadow" attribute are called "Tentative" (potentially important), and features with a Z score significantly lower than the "Shadow" attribute are called "Rejected" (not important). In previous studies (Bai et al., 2022; Li Z et al., 2022), "Confirmed+Tentative" was used as the final result of feature selection, and no actual research was conducted to explore which feature parameter set performs better in building a model. Therefore, this study divides the Boruta algorithm screening

results into two categories: "Confirmed" and "Confirmed +Tentative", analyzes the impact of the two feature parameter sets on model construction and compares their accuracy.

In this study, the features screened by the RFE method are prefixed with R-. The feature parameter sets screened by the Boruta method are "Confirmed" and "Confirmed+Tentative", represented by C- and CT- respectively. For example, the spectral features screened by the RFE algorithm are prefixed with R-SF means.

2.5.2 Features fusion

Features fusion is a method of building a model by fusing different types of remote sensing features together (Liu et al., 2022b; Liu et al., 2023). In this study, the features fusion strategy is mainly divided into two parts: 1. First, fuse SF and TF, implement the features selection method to optimize, then build SPAD estimation models, and compare it with the model built by the fused features that did not participate in feature selection. 2. Based on the features selection strategy, the SPAD estimation model is constructed by fusing the selected feature subsets of different categories, to identify a SPAD estimation model that delivers optimal performance during the later stages of wheat growth.

2.6 Model development and accuracy evaluation

Based on R language version 4.1.3 (R Foundation, Vienna, Austria), the support vector machine regression (SVR) algorithm was used to predict winter wheat SPAD. The idea of SVR originated from the support vector machine. It transforms the input low-dimensional sample set into a high-dimensional space and finds a hyperplane that can be closest to all feature sample sets to implement regression, minimize structural risks, and improve sample discreteness (Smola and Schölkopf, 2004). In this study, the SPAD data set of winter wheat in each growth stage was divided into a calibration set and a validation set. The dataset was randomly sampled, 2/3 was used for model training (Calibration), and 1/3 was used to evaluate the performances (Validation) (Table 3). Three statistical indicators are used to test the machine learning model: the coefficient of determination (R^2), root mean square error (RMSE), and the ratio of performance to deviation (RPD) (Bruning et al., 2019). Their calculation formulas are presented as Equations 1–3.

$$R^2 = 1 - \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \quad (2)$$

$$RPD = \frac{SD}{RMSE} \quad (3)$$

In the formula, $\hat{y}_i$ and y_i are the observed values and measured values of SPAD, $\bar{y}$ is the average of SPAD observed values, n is the sample size, and SD is the standard deviation of reference values.

3 Results

3.1 Analysis of SPAD variation

Figure 3 illustrates the dynamic changes in SPAD for wheat under different nitrogen treatments, varieties, and growth stages. As observed in the Figure 3, SPAD exhibits noticeable differences across various nitrogen gradient treatments, generally trending upward with increasing nitrogen application, peaking at the N3 level. The performance of SPAD varies across growth stages, initially showing a slight increase followed by a sharp decrease. Notably, the SPAD values are highest during the Flowering stage and lowest in the Late filling stage. Furthermore, no significant differences in SPAD were found among different wheat varieties, with the trends being relatively consistent across various periods and nitrogen treatments.

3.2 The optical features with SF and TF

In this study, 30 spectral features and 160 texture features were extracted from UAV multispectral images. Not all features are helpful for wheat SPAD estimation, and redundant features may affect model accuracy. Optimizing remote sensing features sensitive to SPAD using features selection strategy (Figure 4). As can be seen from Figures 5A–C, during the heading stage, three identical feature sets were selected, encompassing all five types of spectral features. In the flowering stage, spectral feature types IV and V performed the most effectively, with C-SF and CT-SF selecting all features from

TABLE 3 Descriptive statistics on SPAD for the calibration and validation sets.

Stage	Dataset	Number	Min	Mean	Max	SD
Heading	Calibration	24	31.59	42.76	55.20	7.76
	Validation	12	30.17	44.82	55.31	7.23
Flowering	Calibration	24	34.38	44.62	52.74	6.42
	Validation	12	30.30	45.87	56.12	7.63
Late filling	Calibration	24	5.70	20.71	39.47	12.39
	Validation	12	6.59	22.50	44.29	12.36

Min, Mean, Max, and SD represent the minimum value, mean value, maximum value, and standard deviation of each dataset.

10.3389/fpls.2024.1404238

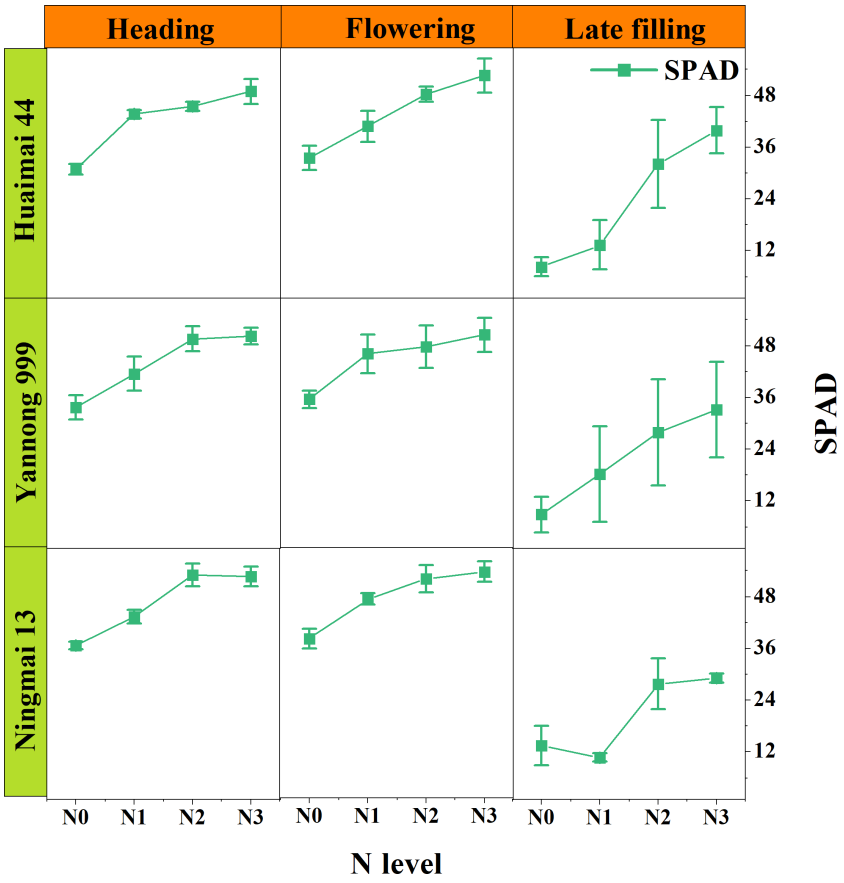


FIGURE 3
SPAD Dynamic Change Line Chart, where the nodes represent the average values of the SPAD.

types IV and V, while R-SF predominantly chose type IV as the optimal variable set. In the late filling stage, types IV and V continued to excel, and the three feature sets showed a high degree of consistency in favoring spectral feature types IV and V. The commonality between spectral feature types IV and V is their composition, which includes participation from the near-infrared band.

The results of texture features band type selection for the three stages are shown in Figures 5D–F. During the heading stage, the R, RE, and NIR bands showed better texture performance, far exceeding the B and G bands. During the flowering stage, the trend of texture feature band types selected by the three methods was consistent, with the Red and NIR bands showing better performance. During the late filling stage, the texture feature band performances were consistent across the three selection methods, with no significant differences among the five texture feature band types. The texture performance of the R and NIR bands was relatively stable, and the number of retained features

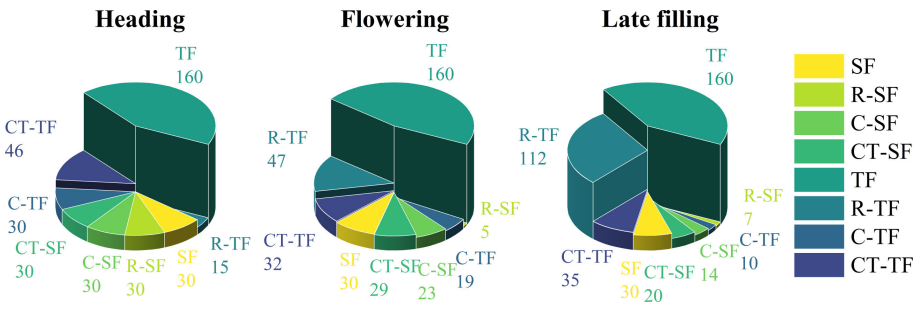


FIGURE 4
The optimal number of variables selected by Boruta and RFE features selection methods for different feature sets during different growth stages.

10.3389/fpls.2024.1404238

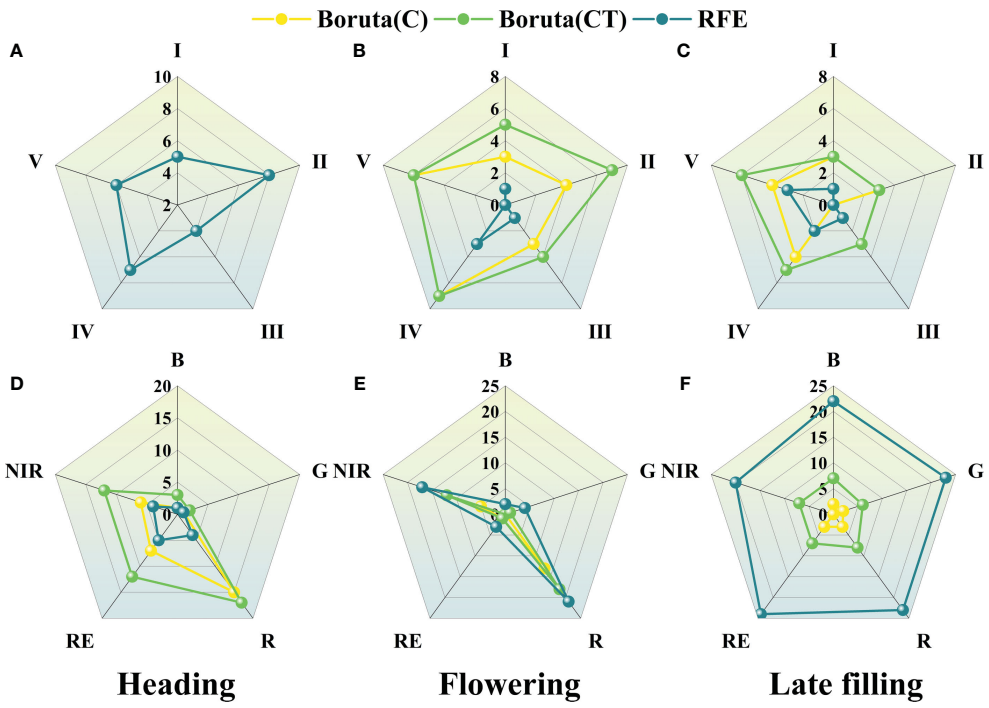


FIGURE 5
The radar chart of features selection for the three growth stages. **(A–C)** spectral features and **(D–F)** texture features.

after selection was significantly higher than that of other bands, except during the late filling stage where there was no significant difference among the five bands.

3.3 Construction and validation of the SPAD prediction model were carried out using different strategy

3.3.1 Estimation of SPAD during the late growth stage of winter wheat using features selection strategy

In this study, SPAD estimation models were developed for the heading stage, flowering stage, and late filling stage based on spectral features and texture features extracted from UAV multispectral images. For each stage, the original feature set was included in the model construction process. A total of 24 (4*2*3) SPAD regression prediction models were built using four feature sets, two types of features (spectral features and texture features), and three growth stages (Figure 6).

From the perspective of model accuracy before and after feature selection, the accuracy of the models constructed using the selected dataset is generally higher compared to that of the initial dataset, but there are also some exceptions. For example, during the flowering stage, the SF dataset (Validation: R^2 = 0.805, RMSE=3.211, RPD=1.998) has higher accuracy than the R-SF, C-SF, and CT-SF datasets for constructing SPAD estimation models. The accuracy among the three datasets composed after features selection varies, and their performance differs in different growth stages and feature types. For example, during the heading stage, the R-TF dataset has lower accuracy (Validation: R^2 = 0.800, RMSE=3.732, RPD=2.080) compared to the C-TF and CT-TF datasets, while during the flowering stage, the R-TF dataset has higher accuracy (Validation: R^2 = 0.799, RMSE=2.941, RPD=2.181) than the C-TF and CT-TF datasets. The accuracy of the C- and CT-datasets selected by the Boruta algorithm is not consistently higher or lower.

3.3.2 Estimation of SPAD during the late growth stage of winter wheat using features fusion strategy

SPAD estimation models were established for three growth stages based on spectral and texture features extracted from UAV multispectral images using a features fusion strategy, followed by feature selection. The original feature set was included in model construction at each stage, and a total of 12 (4*3) SPAD regression prediction models were constructed using four feature sets, one feature type (spectral and texture feature fusion), and three growth stages.

As shown in Figure 7, under this feature strategy, the SPAD estimation model for winter wheat canopy at the heading stage (R-SFTF) constructed using the RFE features selection method combined with the SFTF dataset achieved the highest level of accuracy. The performance indicators of the model were as follows: Validation: R^2 = 0.861, RMSE=3.604, RPD=2.154. The SPAD estimation model for winter wheat at the flowering stage (C-SFTF) constructed using the Boruta features selection method combined with the SFTF dataset had the best accuracy, with specific performance indicators of Validation: R^2 = 0.740, RMSE=3.256, RPD=1.971. The SPAD estimation model for winter wheat at the

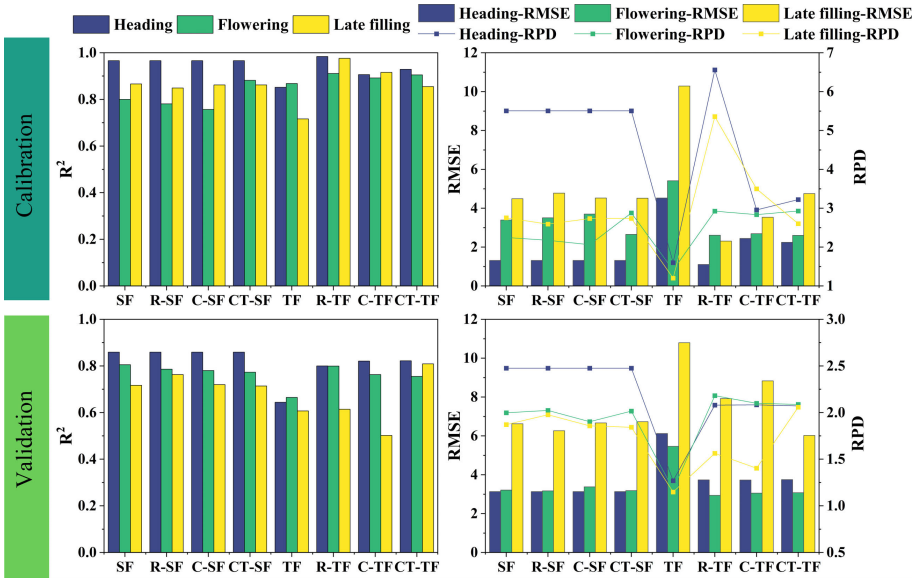


FIGURE 6
The SVR models for estimating SPAD based on different datasets in three growth stages.

late filling stage (R-SFTF) constructed using the RFE features selection method combined with the SFTF dataset had the best performance, with Validation: R^2 = 0.761, RMSE=6.250, RPD=1.983.

On the other hand, the features fusion strategy was executed to establish SPAD estimation models for the heading stage, flowering stage, and late filling stage. For each stage, 27 SPAD prediction models (9*3) were constructed using 9 (3*3) feature sets, 1 feature type (features fusion set), and 3 growth stages. Additionally, 30 regression models were created by adding the unfiltered SFTF feature set as a comparison at each stage (Figure 8).

As shown in Figure 8, under this strategy, during the heading stage, the R-SF, C-SF, and CT-SF feature sets were the same. Therefore, there were three optimal estimation models that fused the R-TF data with these feature sets: R-SF–R-TF, C-SF–R-TF, and CT-SF–R-TF. The performance indicators of these models are as follows: Validation: R^2 = 0.857, RMSE=3.134, RPD=2.477. During the flowering stage, the best accuracy for canopy SPAD estimation in winter wheat was achieved by the fusion model (R-SF–R-TF) constructed using the RFE features selection method combined with the SF and TF datasets. The specific performance indicators are as follows: Validation: R^2 = 0.807,

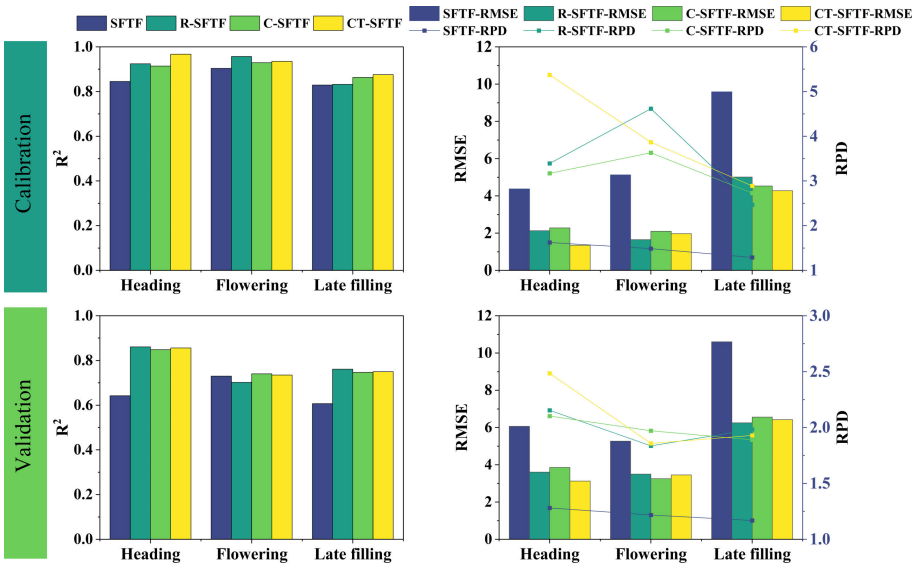


FIGURE 7
SVR models for SPAD estimation based on different fusion datasets at three growth stages.

10.3389/fpls.2024.1404238

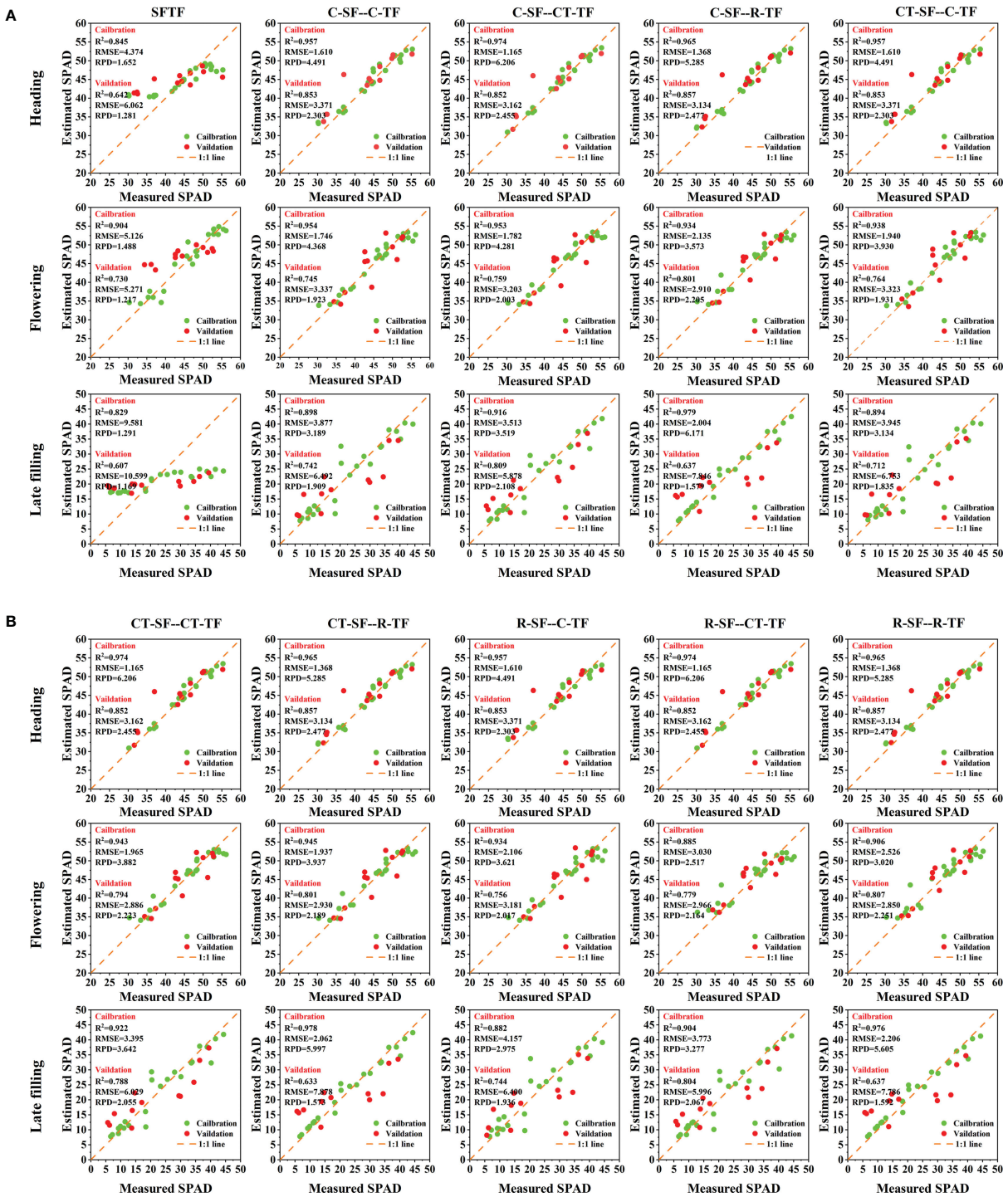


FIGURE 8
The SVR regression models constructed using the fusion dataset of selected spectral and texture features. C-, CT- and R- represent the feature set selected by the Boruta algorithm "Confirmed", Boruta algorithm "Confirmed+Tentative" and RFE algorithm. SF represents the spectral features, TF represents the texture features. A represents the first five fusion treatments, which include SFTF, C-SF–C-TF, C-SF–CT-TF, C-SF–R-TF, and CT-SF–C-TF. B encompasses the latter five fusion treatments, consisting of CT-SF–CT-TF, CT-SF–R-TF, R-SF–C-TF, R-SF–CT-TF, and R-SF–R-TF.

RMSE=2.850, RPD=2.251. For the late filling stage, the best performance was observed in the fusion model (C-SF–CT-TF) constructed by combining the SF and TF datasets using the Boruta features selection method. The performance indicators for this model are as follows: Validation: R^2 = 0.809, RMSE=5.878, RPD=2.108.

To further analyze the accuracy of the SPAD estimation models, From the Figure 9, it can be observed that the optimal monitoring models for the later growth stages of winter wheat were all SPAD estimation models constructed using a dual-strategy. At each stage, the data points are clustered closely around the 1:1 line,

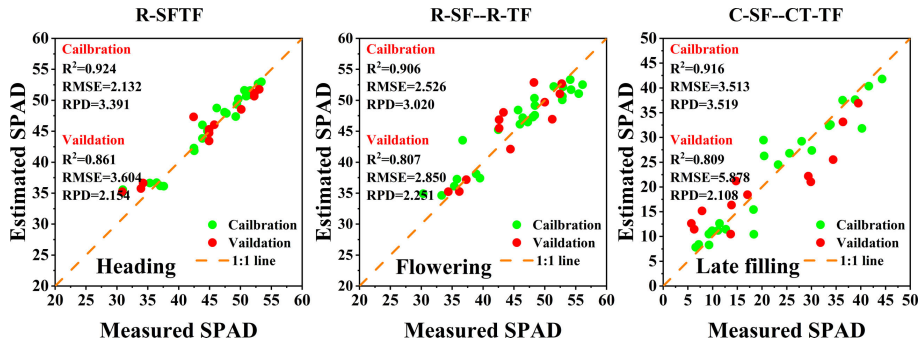


FIGURE 9
Scatter plots of the optimal SPAD estimation models.

demonstrating strong agreement between the model's predicted values and the field-measured values, with minimal errors. These results indicate the models' capability to accurately estimate the SPAD in winter wheat.

4 Discussion

4.1 The main spectral and texture features that affect the SPAD during the later growth stages of winter wheat

Wheat exhibits significant SPAD variation during the heading, flowering, and late filling stages, with a close relationship between SPAD and nitrogen treatment. Specifically, wheat SPAD exhibits a gradual upward trend with increased field nitrogen gradient treatments. This is because nitrogen is one of the essential nutrients for plant growth and development (Wang W et al., 2021) and a crucial component of chloroplasts in leaves. SPAD, as a measure of relative chlorophyll content in leaves, is inevitably closely related to nitrogen. During the heading and flowering stages, wheat plants absorb nitrogen from the soil through their roots. This nitrogen is then translocated to various parts of the plant, including the leaves (El-Hendawy et al., 2017). In contrast to the heading and flowering stages, SPAD in wheat drops sharply during the late filling stage. This may be due to nitrogen translocation within the plant at this time. During the filling stage, nitrogen from the leaves gradually moves into the grains, a process accompanied by a decline in photosynthetic capacity and senescence of the leaves (Bowman et al., 2015). Leaves begin to turn yellow, and the chlorophyll content decreases sharply, leading to lower SPAD values.

Spectral features have been verified to be closely associated with crop agronomic traits and are considered important indicators for estimating crop phenotypic information (Duan et al., 2019). It is well known that the sensitivity of spectral data to vegetation decreases during the later growth stages of crops (Zhu et al., 2023). This can be attributed to the canopy closure phenomenon and the complex spectral mixing and canopy structure effects that affect the sensitivity of spectral data to vegetation (Li R et al., 2022). Additionally, the canopy information generated by the spike emergence further exacerbates this effect (Wang et al., 2022). In this study, it was found that the NIR spectral bands combined with other spectral bands during the later growth stages of crops were the most frequently selected spectral features after optimal treatment (Figures 5, 10). This finding is consistent with the conclusion reached by Yin et al. (2023), that the sensitive spectral features for estimating SPAD in winter wheat are combinations of near-infrared and other bands using three features selection methods. This may be related to the sensitivity of the NIR bands to the growth status of crops. As the growth stage progresses, nitrogen in winter wheat leaves gradually transfers to the spike organs, the leaves wilt, and the appearance of wheat plants changes, leading to a sharp change in reflectance in the near-infrared bands. Zheng et al. (2018) found that the near-infrared spectral information was most important when estimating leaf nitrogen content (LNC) using mean squared error (MSE) to evaluate variable importance. LNC is strongly correlated with SPAD, and the results of this study are similar to their conclusion, indicating the great potential of near-infrared spectral data for estimating SPAD during the later growth stages of winter wheat.

In this study, the texture features of the red and near-infrared bands maintained a high sensitivity to SPAD information during the reproductive growth stage of wheat (Figures 5, 11). This is consistent with the conclusion reached by Bai et al. (2022), that the texture features of the red band can better predict soybean yield compared to the blue and green bands, and with the conclusion reached by Lu et al. (2021), that extracted texture features from the near-infrared band obtained by multispectral sensors are more effective in predicting potassium content in rice plants than other bands. This may be because the texture features of the red and near-infrared bands can better describe the spatial distribution changes of the crop canopy, providing valuable information for estimating crop phenotypes. It also demonstrates the feasibility of using UAV equipped with RGB sensors (including the red band) or multispectral sensors (including the red and near-infrared bands) to monitor crop phenotypic information using texture features. Furthermore, previous studies that used texture features to construct crop phenotyping models mostly considered the relationship between the mean (MEAN) of texture feature indices and crop phenotypes (Fu et al., 2021; Liu et al., 2022a; Luo et al., 2022; Zhou et al., 2022), while rarely exploring other statistical indices. In this study, further analysis of statistical indices of texture

10.3389/fpls.2024.1404238

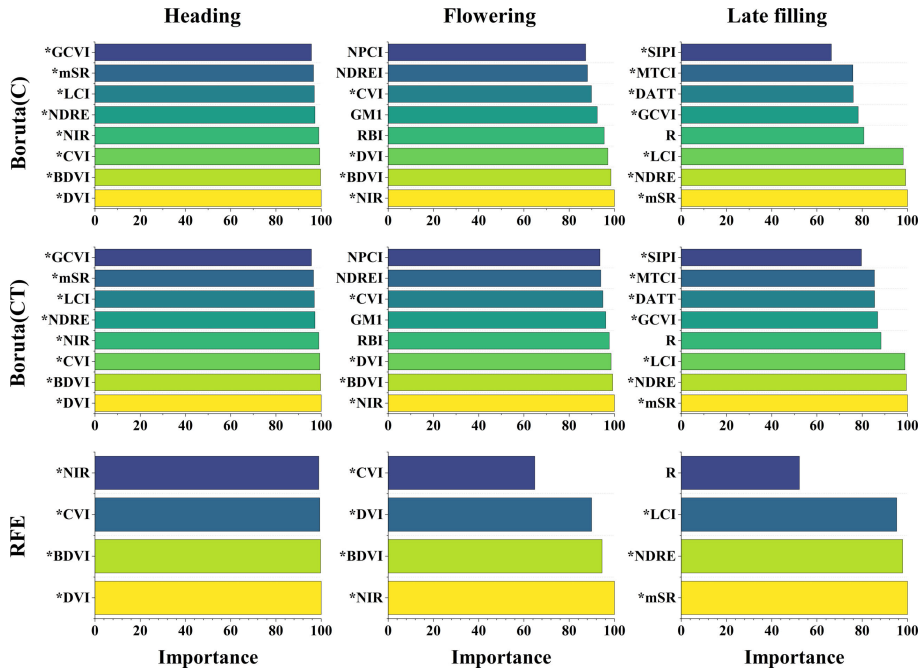


FIGURE 10
The importance of spectral features selected by different features selection methods. The Boruta algorithm displays the top 8 spectral features in terms of importance, while the RFE algorithm retains fewer features and only shows the top 4. The * symbol indicates that the spectral features involves the NIR band.

features (Figures 11, 12) extracted four different statistical indices based on GLCM features for each sample area: MEAN, SD, MAX, and MIN. The commonly used MEAN feature consistently maintained a leading position among the four statistical indices (Figure 12), but its importance gradually decreased as the growth stage progressed (Figure 11). The changing trend of importance for other indices was opposite to MEAN. From the heading stage to the late filling stage, their importance gradually increased and reached a

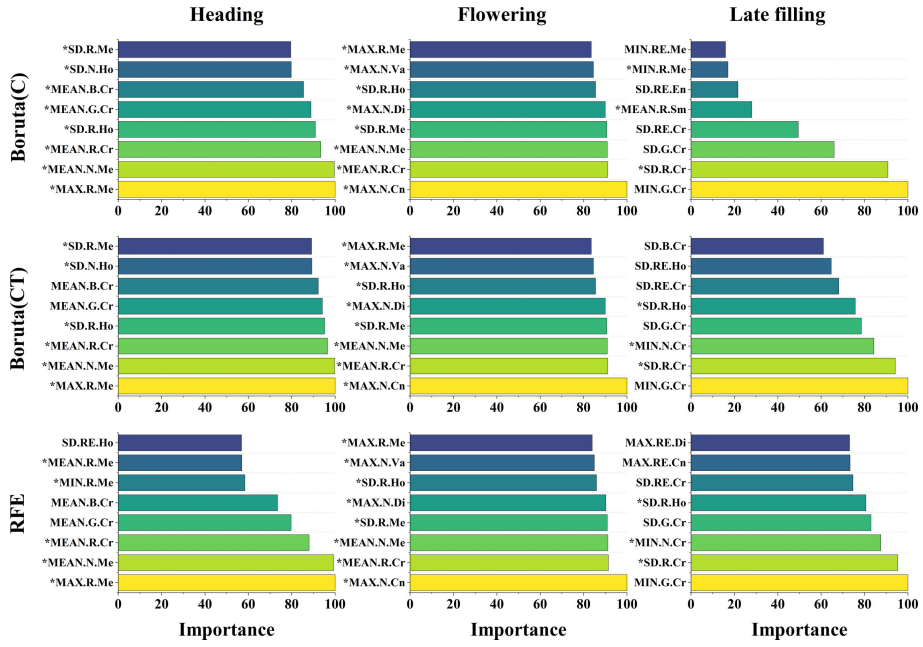


FIGURE 11
The importance of texture features selected by different features selection methods. The Boruta and RFE algorithms display the top 8 texture features in terms of importance. The * symbol indicates that the texture features belongs to the red and NIR bands.

10.3389/fpls.2024.1404238

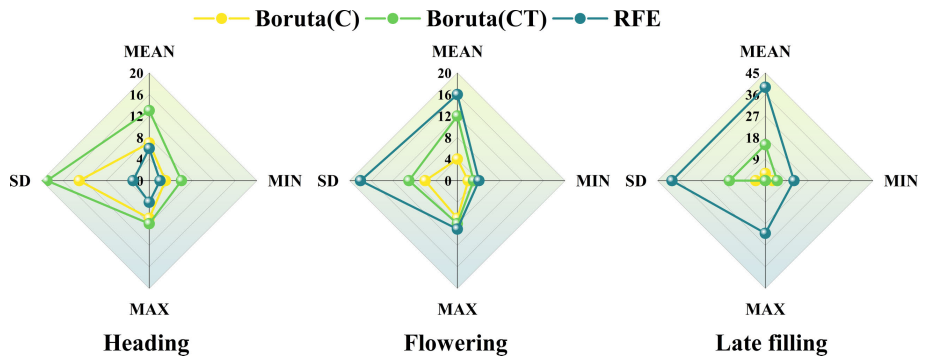


FIGURE 12
The statistical indices of the retained texture features at the three growth stages of wheat.

peak in the late filling stage, with SD showing the best trend. This indicates that other statistical indices also have the potential to estimate SPAD in wheat, especially SD. This is consistent with the conclusion reached by Bai et al. (2022), that using texture features incorporating SD information achieved good accuracy and precision in predicting early soybean yield. The results of this study indicate that the SD information of texture features also has the potential to estimate crop SPAD.

4.2 The impact of features selection strategy on SPAD estimation

Features selection methods play a significant role in machine learning and can directly impact the predictive model (Kursa and Rudnicki, 2010). When estimating crop phenotypes using remote sensing features, selecting appropriate remote sensing features through features selection strategy can improve model performance and eliminate interference (Lin et al., 2021; Xu et al., 2023). In this study, two features selection methods, Boruta and RFE, were used to select optimal spectral and texture features and build models. Additionally, a comparison of model accuracy before and after features selection was conducted to further analyze the mechanism of features selection methods and select suitable methods for crop phenotype studies. From Figure 13, it can be observed that the accuracy of the SPAD estimation model constructed using the subset of variables after features selection is generally better than that of the original variable set. However, there are also some exceptions, as the Boruta method shows a slight decrease in model accuracy (Figure 13, Late filling, TF and C-TF). This may be related to the principle of the Boruta method. Boruta cannot identify and explore features with low individual value but high combined value, leading to the removal of low-value features and reducing or eliminating the high value generated by feature combinations, thus resulting in decreased model accuracy. Furthermore, Boruta focuses on selecting all features related to SPAD (Kursa and Rudnicki, 2010), overlooking the relationship between features, which may lead to multicollinearity among features and subsequently reduce model performance.

Therefore, features selection techniques should not only consider the relationship between features and SPAD but also take into account the relationship between features. RFE does this well by continuously combining subsets of variables to build models in order to find the optimal variable set with the minimum RMSE. For example, Wang J et al. (2021) used two features selection methods, RF and r, to construct RF and SVR models for estimating winter wheat SPAD, and compared them with the RFE features selection method. They found that the RFE features selection method combined with RF and SVR models had higher predictive accuracy. Lin et al. (2021) compared three features selection methods, Boruta, Sequential Forward Selection (SFS), and Permutation Importance-Recursive Feature Elimination (PI-RFE), and found that the PI-RFE method had the best dimensionality reduction effect on multidimensional feature sets, greatly improving the accuracy of FVC inversion. In this study, whether it was at the heading stage or flowering stage, the RFE

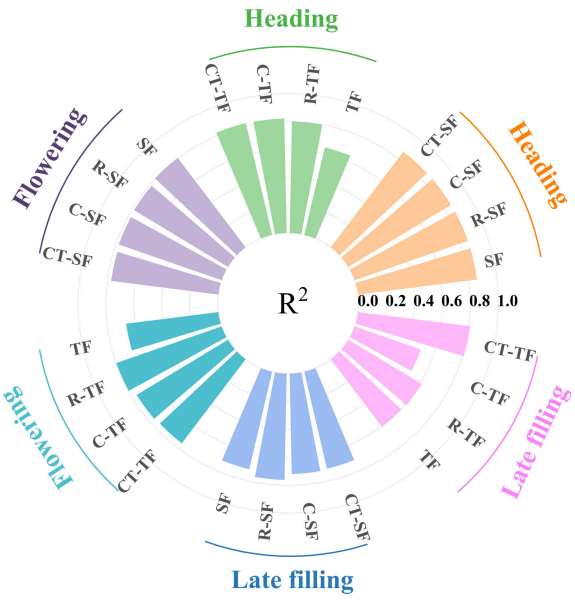


FIGURE 13
The comparison of R^2 values for the validation accuracy of winter wheat SPAD estimation models constructed using features selection methods to select suitable remote sensing features at three growth stages.

features selection method combined with the SVR algorithm provided the best SPAD prediction model accuracy. In the late filling stage, the RFE-SVR combination still showed high SPAD estimation ability (Figures 7, 8). Overall, compared with the Boruta method, the RFE combined with the SVR algorithm can achieve higher model accuracy and stronger model performance. RFE takes into account the interrelationships between features and between features and SPAD, thereby improving the performance of prediction model.

4.3 The impact of features fusion strategy on SPAD estimation

Features fusion strategy involve merging different types of features into a more optimal feature set, combining features with different features to construct models, and have the potential to mitigate the impact of spectral saturation effects, enhance model performance and increase the interpretability of prediction models (Zhou et al., 2022; Sun et al., 2023). This explains why the optimal SPAD estimation models in the late growth stage of winter wheat are constructed using features fusion strategy. This study comprehensively reports the essence of the impact of features fusion strategy on model performance through two features fusion approaches. When only implementing features fusion strategy (Table 4), it was found that not all feature fusions lead to improved accuracy. Blindly combining features together to build models may have a counterproductive effect, causing data redundancy and reducing model performance (Son et al., 2015; Li et al., 2020). This conclusion is in line with the discoveries of Zhou L et al. (2023), who observed a decrease in model accuracy when agronomic practice information (API) was fused with spectral and texture features to construct a rice yield estimation model. The essence of the impact of features fusion strategy on model accuracy lies in the contribution of multi-characteristic features to the model and the interactions between features. This also explains why there are differences in predictive model results when different variables are fused (Figure 8). Similar to the findings of Liu J et al. (2022), who estimated rice nitrogen use efficiency (NUE) using different numbers of spectral features and concluded that increasing the number of variables does not necessarily improve model accuracy, this study also discovered that there is no essential correlation between model accuracy and the quantity of features (Table 4; Figures 4, 6–8, 12, 13). Adding variables does not necessarily make the model more robust, and features selection strategy are powerful means to address these issues.

TABLE 4 Constructing SPAD estimation model using the dataset of SF, TF and SFTF.

Stage	Dataset (original)	Number	Validation		
			R^2	RMSE	RPD
Heading	SF	30	0.859	3.135	2.476
	TF	160	0.644	6.119	1.269
	SFTF	190	0.642	6.062	1.281
Flowering	SF	30	0.805	3.211	1.998
	TF	160	0.665	5.463	1.175
	SFTF	190	0.730	5.271	1.217
Late filling	SF	30	0.717	6.624	1.871
	TF	160	0.607	10.794	1.148
	SFTF	190	0.607	10.599	1.169

SF represents initial spectral features, TF represents initial texture features, and SFTF represents the fused feature set obtained by combining the initial spectral features and initial texture features.

After combining features fusion strategy with features selection strategy, the stability, accuracy, and robustness of the SPAD estimation models for winter wheat are at the forefront (Figure 9). As the growth stage progresses, the improvement in model accuracy gradually increases. The highest increase in model accuracy is observed in the late filling stage, where compared to models using only initial spectral or texture features, R^2_{Val} increased by 0.092 to 0.202, $RMSE_{Val}$ decreased by 0.076 to 4.916, and RPD_{Val} increased by 0.237 to 0.960. Sun et al. (2023) developed maize LAI estimation model using SVM, RF, BPNN, and PLSR algorithms by selecting five spectral features and three texture indices through correlation analysis. The results showed that the SVM algorithm with multi-variable fusion achieved the highest accuracy. Compared to SVM models using only spectral or texture features as input variables, R^2_{Val} increased by 0.023 to 0.192, $RMSE_{Val}$ decreased by 0.015 to 0.036, and RPD_{Val} increased by 0.074 to 0.746. Ma et al. (2022) used four features selection methods, including correlation coefficient, MIC, RF, and RFE, to select three spectral features and three texture features to construct a cotton yield estimation model. The results showed that the RFE_ELM model based on the fusion of spectral and texture features achieved the highest accuracy, with an increase in R^2_{Val} of 0.073 to 0.187 compared to the non-fusion model. The findings of this study surpass those of prior research, indicating that the combination of features fusion strategy and features selection strategy is an effective method to mitigate the impact of spectral saturation effects and improve the accuracy of crop phenotype estimation, fully tapping into the potential of multi-characteristic features in estimating crop phenotypes.

4.4 Limitations and future research perspectives

This study conducted field experiments on winter wheat using three locally stable and high-yielding varieties (Huaimai 44, Yannong 999, and Ningmai 13) and four nitrogen fertilizer application levels (N0-N3: 0, 100, 200, 300 kg/ha). In the future, we will systematically supplement the dataset and conduct field experiments on more varieties and nitrogen fertilizer treatments. At the same time, because of the swift advancement in sensor technology, there are still limitations in building SPAD prediction models relying solely on single sensor data. In the future, we plan to incorporate RGB and hyperspectral data into the SPAD prediction model. The machine learning algorithms employed in this study are

relatively single, and in the future, we will introduce multiple algorithms for comparison and select the most suitable algorithm for SPAD estimation. Furthermore, the use of multi-source remote sensing information and background effect removal can also improve the accuracy of crop phenotype estimation based on UAV. In the future, we will combine them to further improve the accuracy of SPAD estimation for winter wheat. Although the combination of features fusion strategy and features selection strategy performs well in SPAD estimation during the later growth stage of wheat, the results of this study only come from small-scale experiments. In the future, we will conduct large-scale integrated winter wheat SPAD remote sensing monitoring research using satellite and UAV platforms.

5 Conclusion

This study demonstrates that spectral and texture features have good potential for monitoring the SPAD status of winter wheat in the reproductive stage. Spectral features in the near-infrared band can fully capture the spectral differences of wheat SPAD in the reproductive growth stage, while texture features in the red and near-infrared bands are more sensitive to wheat SPAD. Among the SPAD estimation models under different strategy, the SVR model combining features selection and features fusion strategy had the highest accuracy, and compared with the SVR model that only used initial spectral or texture features as input, the stability of the model was improved. As the growth stage progresses, the enhancement of model accuracy by this method becomes more significant, with the greatest improvement observed during the late filling stage. R^2_{Val} increased from 0.092 to 0.202, $RMSE_{Val}$ decreased from 0.076 to 4.916, and RPD_{Val} increased from 0.237 to 0.960. This study shows that UAV remote sensing technology combined with features selection and fusion strategy has good prospects for monitoring winter wheat SPAD in the reproductive growth stage, and can accomplish precise monitoring of winter wheat growth under different varieties and nitrogen treatments, providing scientific guidance and theoretical support for fine management of field crops nutrition.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

XS: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Writing – original draft, Writing – review & editing, Software. YN: Data curation, Formal analysis, Writing – review & editing. HS: Writing – review & editing. AH: Data curation, Methodology, Software, Writing – review & editing. HY: Data curation, Formal analysis, Writing – review & editing. YZ: Conceptualization, Data curation, Methodology, Writing – review & editing. JLi: Data curation, Methodology, Software, Writing – review & editing. WW: Data curation, Methodology, Writing – review & editing. HW: Supervision, Writing – review & editing. QM: Writing – review & editing, Formal analysis, Methodology. JLiu: Methodology, Writing – review & editing, Funding acquisition, Software, Writing – original draft. XL: Methodology, Software, Writing – original draft, Writing – review & editing, Funding acquisition, Supervision, Validation. YA: Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by Scientific research projects in higher education institutions of Anhui Province (no. 2023AH051855; 2022AH051623); Anhui Province Crop Intelligent Planting and Processing Technology Engineering Research Center Open Research Project (no. ZHKF03); Natural Science Foundation of Hebei Province (no. C2023408010); Scientific research projects in higher education institutions of Hebei Province (no. QN2024158).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Asseng, S., Foster, I., and Turner, N. C. (2011). The impact of temperature variability on wheat yields. *Glob. Change Biol.* 17, 997–1012. doi: 10.1111/j.1365-2486.2010.02262.x

Bai, D., Li, D., Zhao, C., Wang, Z., Shao, M., Guo, B., et al. (2022). Estimation of soybean yield parameters under lodging conditions using rgb information from unmanned aerial vehicles. *Front. Plant Sci.* 13, 1012293. doi: 10.3389/fpls.2022.1012293

Bannari, A., Morin, D., Bonn, F., and Huete, A. (1995). A review of vegetation indices. *Remote Sens. Rev.* 13, 95–120. doi: 10.1080/02757259509532298

Bharati, M. H., Liu, J. J., and MacGregor, J. F. (2004). Image texture analysis: methods and comparisons. *Chemometrics Intell. Lab. Syst.* 72, 57–71. doi: 10.1016/j.chemolab.2004.02.005

Bowman, B. C., Chen, J., Zhang, J., Wheeler, J., Wang, Y., Zhao, W., et al. (2015). Evaluating grain yield in spring wheat with canopy spectral reflectance. *Crop Sci.* 55, 1881–1890. doi: 10.2135/cropsci2014.08.0533

Bruning, B., Liu, H., Brien, C., Berger, B., Lewis, M., and Garnett, T. (2019). The development of hyperspectral distribution maps to predict the content and distribution of nitrogen and water in wheat (Triticum aestivum). *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01380

Chen, L., Xing, M., He, B., Wang, J., Shang, J., Huang, X., et al. (2021). Estimating soil moisture over winter wheat fields during growing season using machine-learning methods. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 14, 3706–3718. doi: 10.1109/JSTARS.2021.3067890

Datt, B. (1999). A new reflectance index for remote sensing of chlorophyll content in higher plants: tests using Eucalyptus leaves. *J. Plant Physiol.* 154, 30–36. doi: 10.1016/S0176-1617(99)80314-9

Datt, B., McVicar, T. R., Van Niel, T. G., Jupp, D. L., and Pearlman, J. S. (2003). Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. *IEEE Trans. Geosci. Remote Sens.* 41, 1246–1259. doi: 10.1109/TGRS.2003.813206

de Castro, A. I., Jurado-Expósito, M., Peña-Barragán, J. M., and López-Granados, F. (2012). Airborne multi-spectral imagery for mapping cruciferous weeds in cereal and legume crops. *Precis. Agric.* 13, 302–321. doi: 10.1007/s11119-011-9247-0

Di Gennaro, S. F., Toscano, P., Gatti, M., Poni, S., Berton, A., and Matese, A. (2022). Spectral comparison of UAV-Based hyper and multispectral cameras for precision viticulture. *Remote Sens.* 14, 449. doi: 10.3390/rs14030449

Duan, B., Liu, Y., Gong, Y., Peng, Y., Wu, X., Zhu, R., et al. (2019). Remote estimation of rice LAI based on Fourier spectrum texture from UAV image. *Plant Methods* 15, 1–12. doi: 10.1186/s13007-019-0507-8

El-Hendawy, S. E., Hassan, W. M., Al-Suhaibani, N. A., and Schmidhalter, U. (2017). Spectral assessment of drought tolerance indices and grain yield in advanced spring wheat lines grown under full and limited water irrigation. *Agric. Water Manage.* 182, 1–12. doi: 10.1016/j.agwat.2016.12.003

Fan, Y., Feng, H., Yue, J., Liu, Y., Jin, X., Xu, X., et al. (2023). Comparison of different dimensional spectral indices for estimating nitrogen content of potato plants over multiple growth periods. *Remote Sens.* 15, 602. doi: 10.3390/rs15030602

Fan, K., Li, F., Chen, X., Li, Z., and Mulla, D. J. (2022). Nitrogen balance index prediction of winter wheat by canopy hyperspectral transformation and machine learning. *Remote Sens.* 14, 3504. doi: 10.3390/rs14143504

Feng, H., Tao, H., Li, Z., Yang, G., and Zhao, C. (2022). Comparison of UAV RGB imagery and hyperspectral remote-sensing data for monitoring winter wheat growth. *Remote Sens.* 14, 3811. doi: 10.3390/rs14153811

Fu, Y., Yang, G., Song, X., Li, Z., Xu, X., Feng, H., et al. (2021). Improved estimation of winter wheat aboveground biomass using multiscale textures extracted from UAV-based digital images and hyperspectral feature analysis. *Remote Sens.* 13, 581. doi: 10.3390/rs13040581

Gitelson, A. A., and Merzlyak, M. N. (1996). Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. *J. Plant Physiol.* 148, 494–500. doi: 10.1016/S0176-1617(96)80284-7

Gitelson, A. A., Merzlyak, M. N., and Chivkunova, O. B. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochem. Photobiol.* 74, 38–45. doi: 10.1562/0031-8655(2001)074<0038:OPANEO>2.0.CO;2

Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., and Leavitt, B. (2003). Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys. Res. Lett.* 30, 1248. doi: 10.1029/2002GL016450

Gitelson, A. A., Viña, A., Ciganda, V., Rundquist, D. C., and Arkebauer, T. J. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* 32, L08403. doi: 10.1029/2005GL022688

Guo, W., Carroll, M. E., Singh, A., Swetnam, T. L., Merchant, N., Sarkar, S., et al. (2021). UAS-based plant phenotyping for research and breeding applications. *Plant Phenomics* 2021. doi: 10.34133/2021/9840192

Han, Y., Tang, R., Liao, Z., Zhai, B., and Fan, J. (2022). A novel hybrid GOA-XGB model for estimating wheat aboveground biomass using UAV-based multispectral vegetation indices. *Remote Sens.* 14, 3506. doi: 10.3390/rs14143506

Hassan, M. A., Yang, M., Rasheed, A., Jin, X., Xia, X., Xiao, Y., et al. (2018). Time-series multispectral indices from unmanned aerial vehicle imagery reveal senescence rate in bread wheat. *Remote Sens.* 10, 809. doi: 10.3390/rs10060809

Hunt, E. R., Cavigelli, M., Daughtry, C. S., Mcmurtrey, J. E., and Walthall, C. L. (2005). Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. *Precis. Agric.* 6, 359–378. doi: 10.1007/s11119-005-2324-5

Hunt, E. R. Jr., Daughtry, C., Eitel, J. U., and Long, D. S. (2011). Remote sensing leaf chlorophyll content using a visible band index. *Agron. J.* 103, 1090–1099. doi: 10.2134/agronj2010.0395

Impollonia, G., Croci, M., Blandinières, H., Marcone, A., and Amaducci, S. (2022). Comparison of PROSAIL model inversion methods for estimating leaf chlorophyll content and LAI using UAV imagery for hemp phenotyping. *Remote Sens.* 14, 5801. doi: 10.3390/rs14225801

Jay, S., Baret, F., Dutartre, D., Malatesta, G., Héno, S., Comar, A., et al. (2019). Exploiting the centimeter resolution of UAV multispectral imagery to improve remote-sensing estimates of canopy structure and biochemistry in sugar beet crops. *Remote Sens. Environ.* 231, 110898. doi: 10.1016/j.rse.2018.09.011

Jay, S., Gorretta, N., Morel, J., Maupas, F., Bendoula, R., Rabatel, G., et al. (2017). Estimating leaf chlorophyll content in sugar beet canopies using millimeter-to centimeter-scale reflectance imagery. *Remote Sens. Environ.* 198, 173–186. doi: 10.1016/j.rse.2017.06.008

Jiang, J., Johansen, K., Stanschewski, C. S., Wellman, G., Mousa, M. A., Fiene, G. M., et al. (2022). Phenotyping a diversity panel of quinoa using UAV-retrieved leaf area index, SPAD-based chlorophyll and a random forest approach. *Precis. Agric.* 23, 961–983. doi: 10.1007/s11119-021-09870-3

Jimenez-Berni, J. A., Deery, D. M., Rozas-Larraondo, P., Condon, A. T. G., Rebetzke, G. J., James, R. A., et al. (2018). High throughput determination of plant height, ground cover, and above-ground biomass in wheat with LiDAR. *Front. Plant Sci.* 9, 237. doi: 10.3389/fpls.2018.00237

Jin, X., Yang, G., Xu, X., Yang, H., Feng, H., Li, Z., et al. (2015). Combined multi-temporal optical and radar parameters for estimating LAI and biomass in winter wheat using HJ and RADARSAR-2 data. *Remote Sens.* 7, 13251–13272. doi: 10.3390/rs71013251

Jurgens, C. (1997). The modified normalized difference vegetation index (mNDVI) a new index to determine frost damages in agriculture based on Landsat TM data. *Int. J. Remote Sens.* 18, 3583–3594. doi: 10.1080/014311697216810

Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Software* 28, 1–26. doi: 10.18637/jss.v028.i05

Kumar, S., Gautam, G., and Saha, S. K. (2015). Hyperspectral remote sensing data derived spectral indices in characterizing salt-affected soils: a case study of Indo-Gangetic plains of India. *Environ. Earth Sci.* 73, 3299–3308. doi: 10.1007/s12665-014-3613-y

Kursa, M. B., and Rudnicki, W. R. (2010). Feature selection with the Boruta package. *J. Stat. Software* 36, 1–13. doi: 10.18637/jss.v036.i11

Le Maire, G., François, C., Soudani, K., Berveiller, D., Pontailler, J., Bréda, N., et al. (2008). Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sens. Environ.* 112, 3846–3864. doi: 10.1016/j.rse.2008.06.005

Li, Z., Chen, Z., Cheng, Q., Duan, F., Sui, R., Huang, X., et al. (2022). UAV-based hyperspectral and ensemble machine learning for predicting yield in winter wheat. *Agronomy* 12, 202. doi: 10.3390/agronomy12010202

Li, R., Wang, D., Zhu, B., Liu, T., Sun, C., and Zhang, Z. (2022). Estimation of nitrogen content in wheat using indices derived from RGB and thermal infrared imaging. *Field Crop Res.* 289, 108735. doi: 10.1016/j.fcr.2022.108735

Li, B., Xu, X., Zhang, L., Han, J., Bian, C., Li, G., et al. (2020). Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *Isprs-J. Photogramm. Remote Sens.* 162, 161–172. doi: 10.1016/j.isprsjprs.2020.02.013

Lichtenthaler, H. K., Lang, M., Sowinska, M., Heisel, F., and Miehe, J. A. (1996). Detection of vegetation stress via a new high resolution fluorescence imaging system. *J. Plant Physiol.* 148, 599–612. doi: 10.1016/S0176-1617(96)80081-2

Lin, X., Chen, J., Lou, P., Yi, S., Qin, Y., You, H., et al. (2021). Improving the estimation of alpine grassland fractional vegetation cover using optimized algorithms and multi-dimensional features. *Plant Methods* 17, 1–18. doi: 10.1186/s13007-021-00796-5

Liu, Y., Feng, H., Yue, J., Fan, Y., Bian, M., Ma, Y., et al. (2023). Estimating potato above-ground biomass by using integrated unmanned aerial system-based optical, structural, and textural canopy measurements. *Comput. Electron. Agric.* 213, 108229. doi: 10.1016/j.compag.2023.108229

Liu, Y., Feng, H., Yue, J., Jin, X., Li, Z., and Yang, G. (2022a). Estimation of potato above-ground biomass based on unmanned aerial vehicle red-green-blue images with different texture features and crop height. *Front. Plant Sci.* 13, 938216. doi: 10.3389/fpls.2022.938216

Liu, Y., Feng, H., Yue, J., Li, Z., Yang, G., Song, X., et al. (2022b). Remote-sensing estimation of potato above-ground biomass based on spectral and spatial features extracted from high-definition digital camera images. *Comput. Electron. Agric.* 198, 107089. doi: 10.1016/j.compag.2022.107089

Liu, Y., Su, L., Wang, Q., Zhang, J., Shan, Y., and Deng, M. (2020). Comprehensive and quantitative analysis of growth characteristics of winter wheat in China based on growing degree days. *Adv. Agron.* 159, 237–273. doi: 10.1016/bs.agron.2019.07.007

Liu, J., Zhu, Y., Tao, X., Chen, X., and Li, X. (2022). Rapid prediction of winter wheat yield and nitrogen use efficiency using consumer-grade unmanned aerial vehicles multispectral imagery. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1032170

Lu, J., Eitel, J. U., Engels, M., Zhu, J., Ma, Y., Liao, F., et al. (2021). Improving Unmanned Aerial Vehicle (UAV) remote sensing of rice plant potassium accumulation by fusing spectral and textural information. *Int. J. Appl. Earth Obs. Geoinf.* 104, 102592. doi: 10.1016/j.jag.2021.102592

Luo, S., Jiang, X., He, Y., Li, J., Jiao, W., Zhang, S., et al. (2022). Multi-dimensional variables and feature parameter selection for aboveground biomass estimation of potato based on UAV multispectral imagery. *Front. Plant Sci.* 13, 948249. doi: 10.3389/fpls.2022.948249

Ma, Y., Ma, L., Zhang, Q., Huang, C., Yi, X., Chen, X., et al. (2022). Cotton yield estimation based on vegetation indices and texture features derived from RGB image. *Front. Plant Sci.* 13, 925986. doi: 10.3389/fpls.2022.925986

Madec, S., Baret, F., De Solan, B., Thomas, S., Dutartre, D., Jezequel, S., et al. (2017). High-throughput phenotyping of plant height: comparing unmanned aerial vehicles and ground LiDAR estimates. *Front. Plant Sci.* 8, 2002. doi: 10.3389/fpls.2017.02002

Mao, W., Wang, Y., and Wang, Y. (2003). "Real-time detection of between-row weeds using machine vision," in *2003 ASAE Annual Meeting:American Society of Agricultural and Biological Engineers*, Vol. 1. doi: 10.13031/2013.15381

Mutanga, O., Masenyama, A., and Sibanda, M. (2023). Spectral saturation in the remote sensing of high-density vegetation traits: A systematic review of progress, challenges, and prospects. *Isprs-J. Photogramm. Remote Sens.* 198, 297–309. doi: 10.1016/j.isprsjprs.2023.03.010

Mutanga, O., and Skidmore, A. K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *Int. J. Remote Sens.* 25, 3999–4014. doi: 10.1080/01431160310001654923

Netto, A. T., Campostrini, E., de Oliveira, J. G., and Bressan-Smith, R. E. (2005). Photosynthetic pigments, nitrogen, chlorophyll a fluorescence and SPAD-502 readings in coffee leaves. *Sci. Hortic.* 104, 199–209. doi: 10.1016/j.scienta.2004.08.013

Oktay, T., and Coban, S. (2017). Simultaneous longitudinal and lateral flight control systems design for both passive and active morphing TUAVs. *Elektron. Elektrotech* 23, 15–20. doi: 10.5755/j01.eie.23.5.19238

Peña Barragán, J. M., López Granados, F., Jurado Expósito, M., and García Torres, L. (2007). Mapping Ridolfia segetum patches in sunflower crop using remote sensing. *Weed Res.* 47, 164–172. doi: 10.1111/j.1365-3180.2007.00553.x

Peñuelas, J., Gamon, J. A., Fredeen, A. L., Merino, J., and Field, C. B. (1994). Reflectance indices associated with physiological changes in nitrogen-and water-limited sunflower leaves. *Remote Sens. Environ.* 48, 135–146. doi: 10.1016/0034-4257(94)90136-8

Poblete, T., Camino, C., Beck, P., Hornero, A., Kattenborn, T., Saponari, M., et al. (2020). Detection of Xylella fastidiosa infection symptoms with airborne multispectral and thermal imagery: Assessing bandset reduction performance from hyperspectral analysis. *Isprs-J. Photogramm. Remote Sens.* 162, 27–40. doi: 10.1016/j.isprsjprs.2020.02.010

Pu, R., and Cheng, J. (2015). Mapping forest leaf area index using reflectance and textural information derived from WorldView-2 imagery in a mixed natural forest area in Florida, US. *Int. J. Appl. Earth Obs. Geoinf.* 42, 11–23. doi: 10.1016/j.jag.2015.05.004

Qiao, L., Gao, D., Zhao, R., Tang, W., An, L., Li, M., et al. (2022). Improving estimation of LAI dynamic by fusion of morphological and vegetation indices based on UAV imagery. *Comput. Electron. Agric.* 192, 106603. doi: 10.1016/j.compag.2021.106603

Rasel, S. M., Chang, H., Ralph, T. J., Saintilan, N., and Diti, I. J. (2021). Application of feature selection methods and machine learning algorithms for saltmarsh biomass estimation using Worldview-2 imagery. *Geocarto Int.* 36, 1075–1099. doi: 10.1080/10106049.2019.1624988

Ren, S., Chen, X., and An, S. (2017). Assessing plant senescence reflectance index-retrieved vegetation phenology and its spatiotemporal response to climate change in the Inner Mongolian Grassland. *Int. J. Biometeorol.* 61, 601–612. doi: 10.1007/s00484-016-1236-6

Richardson, A. J., and Everitt, J. H. (1992). Using spectral vegetation indices to estimate rangeland productivity. *Geocarto Int.* 7, 63–69. doi: 10.1080/10106049209354353

Rigon, J., Capuani, S., Fernandes, D. M., and Guimarães, T. M. (2016). A novel method for the estimation of soybean chlorophyll content using a smartphone and image analysis. *Photosynthetica* 54, 559–566. doi: 10.1007/s11099-016-0214-x

Şahin, H., Kose, O., and Oktay, T. (2022). Simultaneous autonomous system and powerplant design for morphing quadrotors. *Aircr. Eng. Aerosp. Technol.* 94, 1228–1241. doi: 10.1108/AEAT-06-2021-0180

Sampson, P. H., Zarco-Tejada, P. J., Mohammed, G. H., Miller, J. R., and Noland, T. L. (2003). Hyperspectral remote sensing of forest condition: Estimating chlorophyll content in tolerant hardwoods. *For. Sci.* 49, 381–391. doi: 10.1093/forestscience/49.3.381

Sellaro, R., Crepy, M., Trupkin, S. A., Karayekov, E., Buchovsky, A. S., Rossi, C., et al. (2010). Cryptochrome as a sensor of the blue/green ratio of natural radiation in Arabidopsis. *Plant Physiol.* 154, 401–409. doi: 10.1104/pp.110.160820

Shao, G., Han, W., Zhang, H., Wang, Y., Zhang, L., Niu, Y., et al. (2022). Estimation of transpiration coefficient and aboveground biomass in maize using time-series UAV multispectral imagery. *Crop J.* 10, 1376–1385. doi: 10.1016/j.cj.2022.08.001

Shu, M., Bai, K., Meng, L., Yang, X., Li, B., and Ma, Y. (2023). Assessing maize lodging severity using multitemporal UAV-based digital images. *Eur. J. Agron.* 144, 126754. doi: 10.1016/j.eja.2023.126754

Shu, M., Zuo, J., Shen, M., Yin, P., Wang, M., Yang, X., et al. (2021). Improving the estimation accuracy of SPAD values for maize leaves by removing UAV hyperspectral image backgrounds. *Int. J. Remote Sens.* 42, 5862–5881. doi: 10.1080/01431161.2021.1931539

Smola, A. J., and Schölkopf, B. (2004). A tutorial on support vector regression. *Stat. Comput.* 14, 199–222. doi: 10.1023/B:STCO.0000035301.49549.88

Son, H., Kim, C., Kim, C., and Kang, Y. (2015). Prediction of government-owned building energy consumption based on an RReliefF and support vector machine model. *J. Civ. Eng. Manage.* 21, 748–760. doi: 10.3846/13923730.2014.893908

Sun, X., Yang, Z., Su, P., Wei, K., Wang, Z., Yang, C., et al. (2023). Non-destructive monitoring of maize LAI by fusing UAV spectral and textural features. *Front. Plant Sci.* 14, 1158837. doi: 10.3389/fpls.2023.1158837

Tatsumi, K., Igarashi, N., and Mengxue, X. (2021). Prediction of plant-level tomato biomass and yield using machine learning with unmanned aerial vehicle imagery. *Plant Methods* 17, 1–17. doi: 10.1186/s13007-021-00761-2

Uddling, J., Gelang-Alfredsson, J., Piikki, K., and Pleijel, H. (2007). Evaluating the relationship between leaf chlorophyll concentration and SPAD-502 chlorophyll meter readings. *Photosynth. Res.* 91, 37–46. doi: 10.1007/s11120-006-9077-5

Vélez, S., Vacas, R., Martín, H., Ruano-Rosa, D., and Álvarez, S. (2022). A novel technique using planar area and ground shadows calculated from UAV RGB imagery to estimate pistachio tree (Pistacia vera L.) canopy volume. *Remote Sens.* 14, 6006. doi: 10.3390/rs14236006

Wang, W., Wu, Y., Zhang, Q., Zheng, H., Yao, X., Zhu, Y., et al. (2021). AAVI: A novel approach to estimating leaf nitrogen concentration in rice from unmanned aerial vehicle multispectral imagery at early and middle growth stages. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 14, 6716–6728. doi: 10.1109/JSTARS.2021.3086580

Wang, W., Zheng, H., Wu, Y., Yao, X., Zhu, Y., Cao, W., et al. (2022). An assessment of background removal approaches for improved estimation of rice leaf nitrogen concentration with unmanned aerial vehicle multispectral imagery at various observation times. *Field Crop Res.* 283, 108543. doi: 10.1016/j.fcr.2022.108543

Wang, J., Zhou, Q., Shang, J., Liu, C., Zhuang, T., Ding, J., et al. (2021). UAV-and machine learning-based retrieval of wheat SPAD values at the overwintering stage for variety screening. *Remote Sens.* 13, 5166. doi: 10.3390/rs13245166

Woebbecke, D. M., Meyer, G. E., Von Bargen, K., and Mortensen, D. A. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Trans. Asae* 38, 259–269. doi: 10.13031/2013.27838

Wu, Q., Zhang, Y., Zhao, Z., Xie, M., and Hou, D. (2023). Estimation of relative chlorophyll content in spring wheat based on multi-temporal UAV remote sensing. *Agronomy* 13, 211. doi: 10.3390/agronomy13010211

Xu, S., Xu, X., Blacker, C., Gaulton, R., Zhu, Q., Yang, M., et al. (2023). Estimation of leaf nitrogen content in rice using vegetation indices and feature variable optimization with information fusion of multiple-sensor images from UAV. *Remote Sens.* 15, 854. doi: 10.3390/rs15030854

Yang, H., Hu, Y., Zheng, Z., Qiao, Y., Zhang, K., Guo, T., et al. (2022). Estimation of potato chlorophyll content from UAV multispectral images with stacking ensemble algorithm. *Agronomy* 12, 2318. doi: 10.3390/agronomy12102318

Yang, B., Wang, M., Sha, Z., Wang, B., Chen, J., Yao, X., et al. (2019). Evaluation of aboveground nitrogen content of winter wheat using digital imagery of unmanned aerial vehicles. *Sensors* 19, 4416. doi: 10.3390/s19204416

Yang, X., Yang, R., Ye, Y., Yuan, Z., Wang, D., and Hua, K. (2021). Winter wheat SPAD estimation from UAV hyperspectral data using cluster-regression methods. *Int. J. Appl. Earth Obs. Geoinf.* 105, 102618. doi: 10.1016/j.jag.2021.102618

Yin, C., Lv, X., Zhang, L., Ma, L., Wang, H., Zhang, L., et al. (2022). Hyperspectral UAV images at different altitudes for monitoring the leaf nitrogen content in cotton crops. *Remote Sens.* 14, 2576. doi: 10.3390/rs14112576

Yin, Q., Zhang, Y., Li, W., Wang, J., Wang, W., Ahmad, I., et al. (2023). Estimation of winter wheat SPAD values based on UAV multispectral remote sensing. *Remote Sens.* 15, 3595. doi: 10.3390/rs15143595

Yue, J., Yang, G., Tian, Q., Feng, H., Xu, K., and Zhou, C. (2019). Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices. *Isprs-J. Photogramm. Remote Sens.* 150, 226–244. doi: 10.1016/j.isprsjprs.2019.02.022

Yue, J., Yang, H., Yang, G., Fu, Y., Wang, H., and Zhou, C. (2023). Estimating vertically growing crop above-ground biomass based on UAV remote sensing. *Comput. Electron. Agric.* 205, 107627. doi: 10.1016/j.compag.2023.107627

Zhang, S., Zhao, G., Lang, K., Su, B., Chen, X., Xi, X., et al. (2019). Integrated satellite, unmanned aerial vehicle (UAV) and ground inversion of the SPAD of winter wheat in the reviving stage. *Sensors* 19, 1485. doi: 10.3390/s19071485

Zheng, H., Li, W., Jiang, J., Liu, Y., Cheng, T., Tian, Y., et al. (2018). A comparative assessment of different modeling algorithms for estimating leaf nitrogen content in winter wheat using multispectral images from an unmanned aerial vehicle. *Remote Sens.* 10, 2026. doi: 10.3390/rs10122026

Zheng, H., Ma, J., Zhou, M., Li, D., Yao, X., Cao, W., et al. (2020). Enhancing the nitrogen signals of rice canopies across critical growth stages through the integration of textural and spectral information from unmanned aerial vehicle (UAV) multispectral imagery. *Remote Sens.* 12, 957. doi: 10.3390/rs12060957

Zhou, C., Gong, Y., Fang, S., Yang, K., Peng, Y., Wu, X., et al. (2022). Combining spectral and wavelet texture features for unmanned aerial vehicles remote estimation of rice leaf area index. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.957870

Zhou, L., Meng, R., Yu, X., Liao, Y., Huang, Z., Lü, Z., et al. (2023). Improved yield prediction of ratoon rice using unmanned aerial vehicle-based multi-temporal feature method. *Rice Sci.* 30, 247–256. doi: 10.1016/j.rsci.2023.03.008

Zhou, X., Yang, M., Chen, X., Ma, L., Yin, C., Qin, S., et al. (2023). Estimation of cotton nitrogen content based on multi-angle hyperspectral data and machine learning models. *Remote Sens.* 15, 955. doi: 10.3390/rs15040955

Zhu, Y., Liu, J., Tao, X., Su, X., Li, W., Zha, H., et al. (2023). A three-dimensional conceptual model for estimating the above-ground biomass of winter wheat using digital and multispectral unmanned aerial vehicle images at various growth stages. *Remote Sens.* 15, 3332. doi: 10.3390/rs15133332

Zhu, W., Sun, Z., Yang, T., Li, J., Peng, J., Zhu, K., et al. (2020). Estimating leaf chlorophyll content of crops via optimal unmanned aerial vehicle hyperspectral data at multi-scales. *Comput. Electron. Agric.* 178, 105786. doi: 10.1016/j.compag.2020.105786

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 05 June 2024
DOI 10.3389/fpls.2024.1411178

Check for updates

OPEN ACCESS

EDITED BY
Liujun Xiao,
Nanjing Agricultural University, China

REVIEWED BY
Yongbing Long,
South China Agricultural University, China
Yunchao Tang,
Dongguan University of Technology, China

*CORRESPONDENCE
Jiong Mu
jmu@sicau.edu.cn

RECEIVED 02 April 2024
ACCEPTED 17 May 2024
PUBLISHED 05 June 2024

CITATION
Zhang L, Luo P, Ding S, Li T, Qin K and Mu J (2024) The grading detection model for fingered citron slices (citrus medica 'fingered') based on YOLOv8-FCS.
Front. Plant Sci. 15:1411178.
doi: 10.3389/fpls.2024.1411178

COPYRIGHT
© 2024 Zhang, Luo, Ding, Li, Qin and Mu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The grading detection model for fingered citron slices (citrus medica 'fingered') based on YOLOv8-FCS

Lingtao Zhang[1,2], Pu Luo[1,2], Shaoyun Ding[1,2], Tingxuan Li[1,2], Kebei Qin[1,2] and Jiong Mu[1,2]*

[1]College of Information Engineering, Sichuan Agricultural University, Ya'an, China, [2]Ya'an Digital Agriculture Engineering Technology Research Center, Sichuan Agricultural University, Ya'an, China

Introduction: Fingered citron slices possess significant nutritional value and economic advantages as herbal products that are experiencing increasing demand. The grading of fingered citron slices plays a crucial role in the marketing strategy to maximize profits. However, due to the limited adoption of standardization practices and the decentralized structure of producers and distributors, the grading process of fingered citron slices requires substantial manpower and lead to a reduction in profitability. In order to provide authoritative, rapid and accurate grading standards for the market of fingered citron slices, this paper proposes a grading detection model for fingered citron slices based on improved YOLOv8n.

Methods: Firstly, we obtained the raw materials of fingered citron slices from a dealer of Sichuan fingered citron origin in Shimian County, Ya'an City, Sichuan Province, China. Subsequently, high-resolution fingered citron slices images were taken using an experimental bench, and the dataset for grading detection of fingered citron slices was formed after manual screening and labelling. Based on this dataset, we chose YOLOv8n as the base model, and then replaced the YOLOv8n backbone structure with the Fasternet main module to improve the computational efficiency in the feature extraction process. Then we redesigned the PAN-FPN structure used in the original model with BiFPN structure to make full use of the high-resolution features to extend the sensory field of the model while balancing the computation amount and model volume, and finally we get the improved target detection algorithm YOLOv8-FCS.

Results: The findings from the experiments indicated that this approach surpassed the conventional RT-DETR, Faster R-CNN, SSD300 and YOLOv8n models in most evaluation indicators. The experimental results show that the grading accuracy of the YOLOv8-FCS model reaches 98.1%, and the model size is only 6.4 M, and the FPS is 130.3.

Discussion: The results suggest that our model offers both rapid and precise grading for fingered citron slices, holding significant practical value for promoting the advancement of automated grading systems tailored to fingered citron slices.

KEYWORDS

finger citron, YOLOv8, object detection, smart agriculture, grading

1 Introduction

Fingered citron (citrusmedical. Var. Sarco-dactylisSwingle), a plant of the genus Citron in the family Rutaceae, is named because its fruit petals are shaped like fingers, resembling the hand of the Buddha, and is also known as Bergamot, bergamot citron, miro, and longevity oranges, which is one of the major traditional Chinese medicinal herbs in China (Wang et al., 2020). *The Pharmacopoeia of the People's Republic of China* has included fingered citron since the 1963 edition, and it has been included until now (Zhang et al., 2018). Fingered citron has a long history of medicinal use. The fruit is used as medicine, which has the effect of regulating the flow of Qi and stopping vomiting, harmonizing the stomach and spleen, eliminating food and resolving phlegm (Luo et al., 2020). Modern medicine is even more research has proved that fingered citron has antidepressant, antibacterial, anti-inflammatory, anticancer, antitumor, antiaging, blood pressure lowering and other effects (Peng et al., 2009; Karp and Hu, 2018; Li et al., 2018). Fingered citron slices are harvested in its autumn when the fruits have not yet turned yellow or yellow, and are often primed by slicing and then sun-drying or low-temperature drying to become fingered citron slices (Klein, 2014; Wu, 2015; Chen et al., 2020).

In today's market, fingered citron slices, as the most widely distributed fingered citron agricultural by-products, are preferred to those with large slices, green skin and white flesh, and strong aroma (Xu et al., 2021a). They are classified into three grades, namely first-grade, second-grade and third-grade. At the same time, due to the low prevalence of standardization and the fragmentation of farmers and dealers, unusable bad slices and counterfeit slices are often mixed in (Xu et al., 2021b), resulting in the loss of interests of many parties. Moreover, most of the agricultural products such as fingered citron slices are judged by laborers using manual visual estimation to determine the grade, which is inconsistent and labor-intensive (Meena et al., 2011; Ismail and Malik, 2022).The continuous loss of agricultural labor in recent years and the unprecedented increase in labor costs (Patel and Patil, 2024), which became more prominent after the COVID-19 pandemic (Nawaz et al., 2021), are detrimental to the market for fingered citron slices and the expansion of fingered citron cultivation scale. In summary, fingered citron slices grading technology, as a key factor to improve the quality of fingered citron slices and to liberate the orchard labor force, is of great significance to increase the added value of the product, to improve the competitiveness of the market, and to alleviate the shortage of orchard labor. A high-precision, high-speed and damage-free grading method is needed to grade fingered citron slices effectively and objectively.

With the continuous development of computer science and technology, computer vision technology in deep learning has penetrated into people's production and life. More and more research has applied computer vision technology to agricultural production and has a wide range of practical applications (Chen et al., 2024a). In research focusing on the grading of agricultural goods using traditional machine learning, Castro and colleagues (Castro et al., 2019) assessed the efficacy of four distinct machine learning strategies paired with three different color spaces for the categorization of cape gooseberry fruit based on their stage of ripeness. Their findings revealed that the utilization of the Lab* color space in conjunction with a Support Vector Machine (SVM) classifier yielded the highest levels of precision and f-measure. In another study, Moallem and associates (Moallem et al., 2017) introduced an algorithm that employs computer vision to evaluate apples. This method involved extracting both textural and geometric attributes from defective regions of the apples. The fruits were then categorized into first-grade, second-grade, and irregular categories using a combination of SVM, Multi-Layer Perceptron (MLP), and K-Nearest Neighbors (KNN) classifiers. Gui et al (Gui et al., 2014). suggested a method for classifying apple shapes based on wavelet rectangles, which successfully sorted apples into three categories: normal, slightly misshapen, and severely misshapen, with respective classification accuracies of 86.2%, 85.8%, and 90.8%. Among these machine learning-based grading techniques, preprocessing of images is a common necessity, and the reliance on singular features for classification can lead to challenges such as suboptimal real-time performance and diminished robustness.

In a research project centered on fruit grading using deep learning techniques, Chakraborty and colleagues (Chakraborty et al., 2023) crafted a specialized, lightweight Convolutional Neural Network (CNN) known as "SortNet". This model is designed for straightforward implementation on edge devices to facilitate real-time citrus fruit sorting through visual analysis. The model's performance, as demonstrated on a test dataset, is detailed in the accompanying table, highlighting an impressive accuracy rate of 97.6%. Chen et al (Chen et al., 2024b. introduced a multi-task Deep Convolutional Neural Network (DCNN) detection model, MTD-YOLOv7, which is an enhancement of the YOLOv7 framework. This model is tailored

for ripeness detection in cherry tomato clusters and has achieved a composite score of 86.6% in multi-task learning, with an average inference time of merely 4.9 milliseconds. Momeny and team (Momeny et al., 2020) enhanced the CNN's generalization capabilities by integrating Max-pooling and Average-pooling techniques for cherry classification. The CNN model delivered classification accuracy of 99.4%, 98.7%, and 99.1% for various input image dimensions. Fan et al (Fan et al., 2020). employed a Convolutional Neural Network (CNN) to recognize apple quality. After training the CNN, they achieved a 96.5% accuracy on the test set. They developed a CNN-based classification software and integrated a computer vision module into a four-threaded fruit sorting machine, capable of sorting at a pace of five fruits per second, with an overall classification accuracy of 92%. However, the model's size is considerable, leading to relatively low computational efficiency. Raikar et al (Raikar et al., 2020). explored the quality classification of okra, categorizing it into four size-based types: small, medium, large, and extra-large. They utilized three deep learning models—AlexNet, GoogLeNet, and ResNet50—with the ResNet model achieving a remarkable accuracy exceeding 99%. Luna et al (De Luna et al., 2019). proposed a deep learning-based approach for identifying defective areas in individual tomatoes. This method utilized the OpenCV library and Python programming. They collected 1200 tomato images of varying qualities using an image capture box. These images were used to train VGG16, InceptionV3, and ResNet50 deep learning models. Upon comparing the experimental outcomes, they determined that VGG16 is the most effective deep learning model for defect detection. Asriny et al (Asriny et al., 2020). suggested a deep learning-based CNN model for categorizing orange images into five distinct classes. They experimented with ReLU and Tanh activation functions, finding that the ReLU activation function surpassed Tanh in the hidden layer. The classification results using the ReLU activation function for data training showed an accuracy of 98.6%, while the validation data yielded 92.8%, and the test data achieved 96%. Fu and associates (Fu et al., 2022) constructed a linear regression model to assess and measure fruit freshness by analyzing the darkness and color changes in the fruit's skin. They evaluated a range of fruits, including apples, bananas, dragon fruits, kiwis, oranges, and pears, for freshness grading, and the highest average accuracy attained was 96.34%. The results indicated that deep learning algorithms are highly effective in addressing this issue. Gururaj and colleagues (Gururaj et al., 2023) developed an innovative CNN architecture for the in-depth grading of three mango varieties. They achieved an accuracy of 93.23% for variety identification and an impressive 95.11% for quality grading.

However, the above deep learning model grading methods still have problems such as insufficient model optimization, poor real-time performance, and are less applied to herbal medicines such as fingered citron slices. Drawing from this foundation, the present study focuses on fingered citron slices as the subject of investigation and delves into the development of an efficient grading detection system specifically tailored for these segments. The paper proposes a fingered citron slices grading detection algorithm based on an improved YOLOv8n, using FasterNet instead of the original YOLO8 backbone, which consumes fewer memory accesses and tends to have higher FLOPS, in order to achieve an improved computational efficiency of the model. The architecture of the initial model is re-envisioned by incorporating the BiFPN (Bi-directional Feature Pyramid Network) module, enhancing the model's capacity for feature integration while simultaneously eliminating superfluous connections to achieve a more streamlined design. Experiment data indicates that this refined approach successfully elevates the detection accuracy of the model, all without incurring additional costs in the training process. Finally, we substantiated the method's viability and dependability through a series of ablation studies and rigorous statistical analyses. The outcomes substantiate that the YOLOv8-FCS algorithm, as introduced in this work, is adept at executing swift and precise grading of fingered citron slices. It can provide a relatively authoritative grading standard and the basis of counterfeit detection for the fingered citron slices market, and help to promote the development of automation in the fingered citron industry.

2 Materials and methods

2.1 Dataset construction

Fingered citron grows in most of the warm citrus producing areas of China (Karp and Hu, 2018). The fingered citron slices used in this dataset were obtained from a Sichuan fingered citron origin dealer in Shimian County, Ya'an City, Sichuan Province, China. The image acquisition devices used in the experiment were Canon camera (60D, 18 megapixel DIGIC4, 18–200 IS lens), iphone15, and Redmi K50 (Xiaomi Technology Co. Ltd., Beijing, China) in order to simulate a variety of filming devices in the real application environment. The above equipment is installed on a self-made shooting platform to shoot at a fixed angle, and partially uses LED light strips as supplementary light sources to capture images of fingered citron slices under diffuse illumination. The fixed position for shooting is 40 centimeters away from the sample. The shooting platform is shown in Figure 1.

According to the 2020 edition of *Chinese Pharmacopoeia*, *Identification of Traditional Chinese Medicines*, *Introduction to the Identification of Practical Chinese Medicinal Traits* and *Instructions for Clinical Use of the Pharmacopoeia of the People's Republic of China*, fingered citron slices are preferable to those with large slices, green skin and white flesh, and strong aroma. It can be seen that the original medicinal materials are preferred to those with heavy quality in terms of commercial specifications. In the traditional grade evaluation, the appearance of the skin-type tablets is required to be thick skin, fine silk and excellent quality. In conclusion, fingered citron slices are preferable to the heavy ones, and the external form is mainly concerned with its quality and thickness. On the other hand, fingered citron slices is preferable to thick skin and thin silk, and its external form is mainly concerned with quality and width. Therefore, the fingered citron slices collected are classified into 4 grades with reference to relevant industry standards and experts' experience.

First-grade products are often packaged as selected pieces and enter the market, with the highest selling price. Second-grade

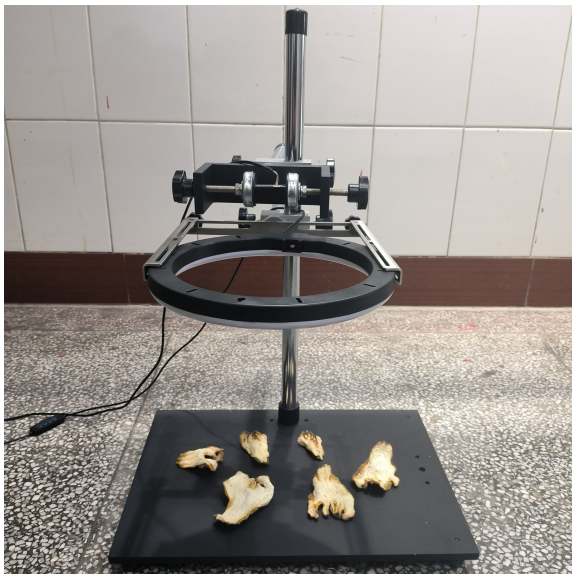


FIGURE 1
Shooting platform.

and third-grade products are bergamot slices with some defects or long storage life, usually entering the market as a single piece, with significant price fluctuations. Bad products include moldy, blackened bergamot slices, or excessively baked scraps, which are usually not sold. Therefore, under the guidance of experts, we classify the bergamot slices based on their color and morphological characteristics. The classification criteria are as follows:

- The color of the first-grade products are slightly yellow, with a regular shape and almost no damage or defects, as shown in Figure 2A.
- The color of the second-grade products are slightly yellow, with a relatively regular shape and slight damage or defects, as shown in Figure 2B.
- The third-grade products have a brown color, irregular shape, and obvious damage or defects, as shown in Figure 2C.
- The color of the bad products are black brown, with an irregular shape and scattered damages or defects, as shown in Figure 2D.

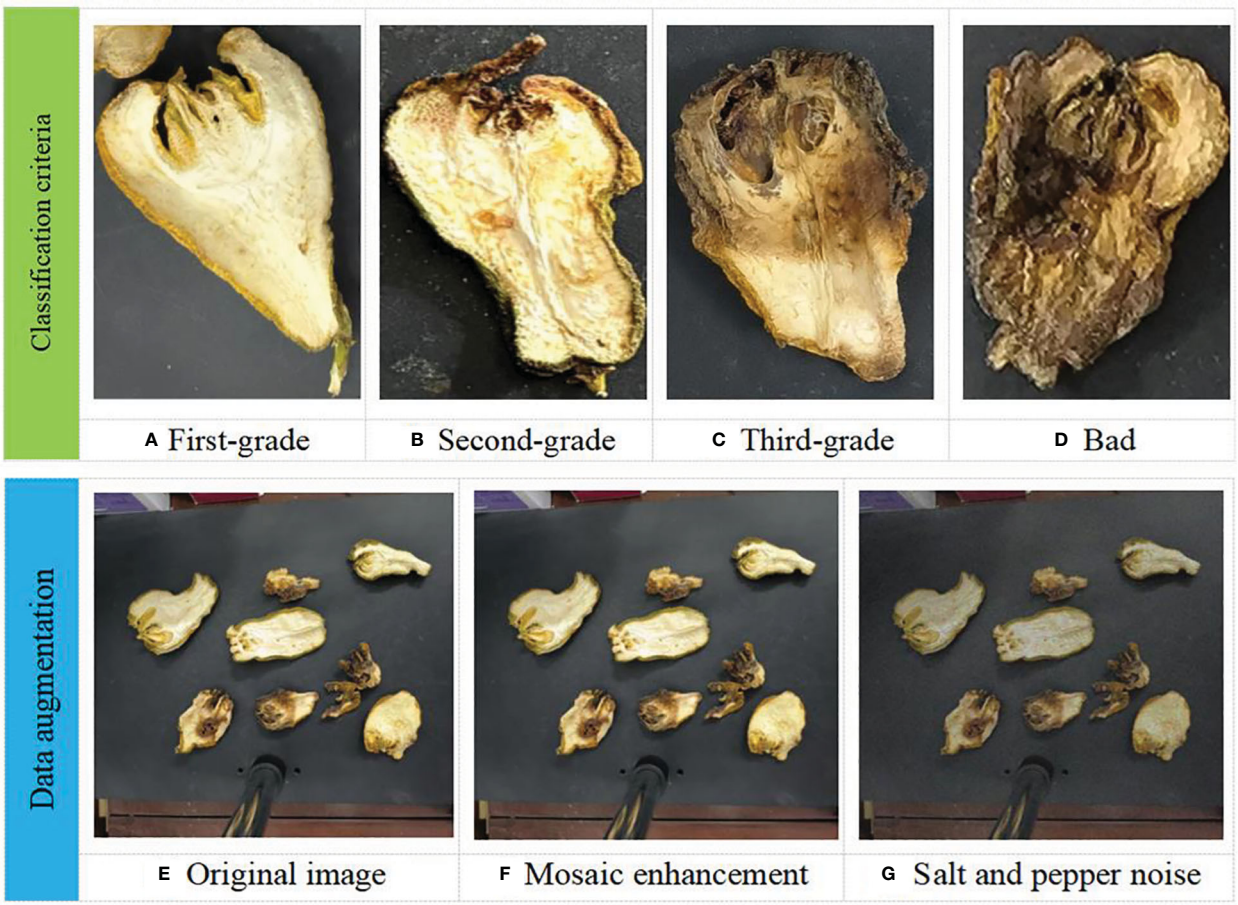


FIGURE 2
Four levels of fingered citron slices **(A–D)** and images before and after data augmentation **(E–G)**.

After we manually filtered out some duplicate and similar data to simulate real application scenarios, we used Gaussian blur and salt and pepper noise to enhance our dataset to improve the generalization and robustness of the model. Examples of four levels of fingered citron and images before and after data augmentation are shown in Figures 2E–G.

Finally we selected 609 images as our initial dataset. We then labeled the dataset according to the above criteria using the open source script LabelImg on GitHub under the guidance of experts, and divided the dataset into training, validation, and test sets according to the ratio of 6:2:2. The labeling process produced a total of 4755 labeled frames, including 1911 first-grade labeled frames, 1620 second-grade labeled frames, 663 third-grade labeled frames, and 561 bad labeled frames. The division of the dataset and the distribution of the number of labeled frames are shown in Table 1.

TABLE 1 Partition of dataset and distribution of labeled boxes.

Data	Proportion	Number of images	Number of labels
Training set	60%	365	2832
Validation set	20%	122	975
Testing set	20%	122	948
sum	100%	609	4755

2.2 Design of fingered citron slices grading method based on improved YOLOv8n

2.2.1 FasterNet

Past means of lightweighting are mainly based on Deep Separable Convolution (DWConv) (Howard et al., 2017). DWConv is a commonly used operation in convolutional neural networks that reduces the number of parameters (FLOPs) and computational complexity by dividing the convolution into two steps: deep convolution and point-by-point convolution. Deep convolution performs the convolution operation independently for each input channel, capturing features on different channels. While point-by-point convolution performs inter-channel combination and fusion of feature maps through a 1×1 convolution kernel. DWConv has the advantages of reduced number of parameters, reduced computational complexity and improved feature expression, making it an efficient convolution operation. However, the nature of its channel-by-channel convolutional operation results in the need for frequent memory accesses in practical operation. This makes the floating-point operation speed (FLOPS) of DWConv become low, which reduces the actual running speed of this operation. Chen et al (Chen et al., 2023). further proposed FasterNet in view of the new PConv and the off-the-shelf PWConv as the main operators.

We found that GhostNet has a certain degree of redundancy in the convolutional channels (Han et al., 2020). FasterNet uses standard convolution on this basis, but processes only part of the channels, and the features of the other part of the channels are directly mapped as constant. In this way, the high FLOPS advantage of standard convolution is utilized, while the FLOPs of normal convolution are reduced due to processing only part of the channels.

The main structure of FasterNet that The FasterNet Block is designed with three parts. The feature input first passes through PConv, which processes only part of the features, followed by 1×1 Conv to expand the number of channels to twice. After going through BN and ReLU, another 1×1 Conv reduces the number of channels and finally a residual link is done.

The overall FasterNet architecture is shown in the Figure 3. It has 4 hierarchical stages, Each level is initiated by either an embedding layer, which consists of a standard 4×4 convolution with a stride of 4, or a merging layer, featuring a standard 2×2 convolution with a stride of 2, both of which are utilized to achieve spatial downsampling and an increase in the channel count. Within each hierarchical level, a multitude of FasterNet blocks is integrated.

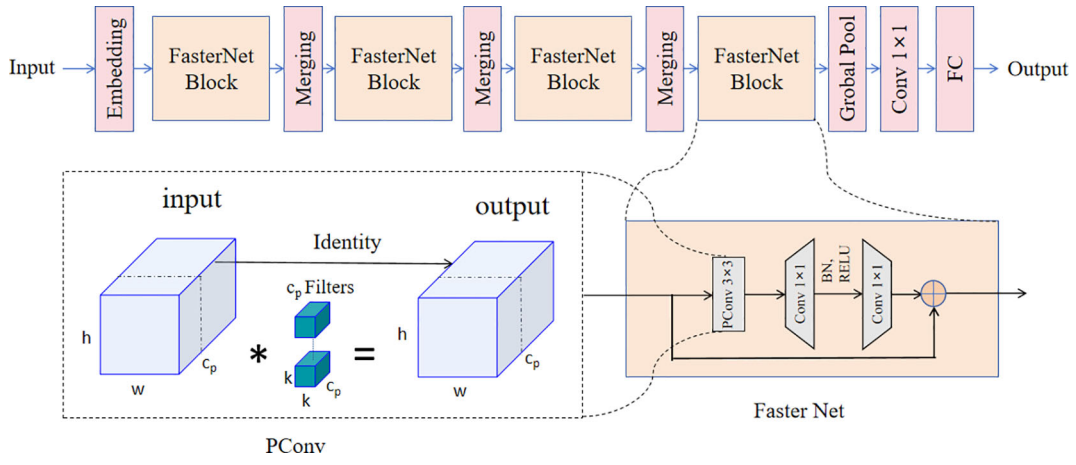


FIGURE 3
The structure of FasterNet.

Given that the blocks in the penultimate and final stages demand less memory and are inclined to offer greater floating-point operations per second (FLOPS), there is an increased deployment of these blocks, with a corresponding augmentation in computational resources dedicated to these stages. A typical FasterNet block commences with a PConv layer, which is then succeeded by 2 PWConv layers or 1×1 convolutional layers, they form an inverted residual configuration. In this setup, the intermediate layer boasts an expanded channel capacity, and a shortcut connection is implemented to redeploy the features from the input. FasterNet constitutes a novel breed of neural networks, renowned for their rapid execution. It excels at a variety of visual tasks, benefited by its streamlined architecture that is both user-friendly and compatible with a broad spectrum of hardware.

2.2.2 BiFPN

After replacing the backbone structure with the FasterNet structure, although it can significantly improve the model mAP, the inference time and model size have increased. Therefore, we expect to find a more efficient way of feature fusion as a way to achieve lightweighting and improve accuracy.

During training, the surface defect targets of fingered citron slices have different shapes and sizes, resulting in features with different resolution sizes. Inspired by PAnet (Liu et al., 2018), the neck of YOLOv8 is designed as a PAN-FPN structure. The regular linear superposition of these features lead to an uneven weighting of features in the fusion output associated with surface defect targets in different fingered citron slices. This imbalance lead to the dominance of large-scale features in the post-fusion output, overwhelming smaller features to influence the grading judgment of fingered citron slices. The structure of PAN-FPN is shown in Figure 4.

To address this challenge and bolster the Neck network's capacity to manage intricate features, it is imperative to select a proficient feature fusion network that can effectively process the features extracted by the feature extraction network across various layers.

The architecture of the BiFPN was initially conceived by Google as part of the EfficientDet (Tan et al., 2020) object detection algorithm. We have adopted BiFPN to serve as the neck network in our model, effectively superseding the previous PAN-FPN framework utilized in YOLOv8. Figure 5 illustrates the configuration of the BiFPN structure.

The BiFPN framework offers an enhanced capacity for feature fusion and more efficient inter-scale connectivity when compared with the PAN-FPN structure. BiFPN streamlines the network by pruning nodes that possess a single input edge, as these have a minimal impact on the comprehensive feature network. Moreover, BiFPN refines the feature integration process by establishing additional connections between the initial input and the output nodes, and by iterating the bi-directional pathway across the same layer multiple times. This approach facilitates robust feature fusion while maintaining a relatively low increment in computational expenditure. It should be noted that the fusion module used in this experiment is a weighted fusion model. In other words, the feature maps are added directly in the spatial and channel dimensions.

The core idea of BiFPN structure is to utilize the information flow in both bottom-up and top-down directions to construct the feature pyramid. Meanwhile, a repetition-weighted fusion method is used at each pyramid level. By utilizing the information flow from both directions, the BiFPN structure can fuse features at different levels to better fit objects of various sizes. By repeating the weighted fusion process, the BiFPN structure enhances the accuracy and generalization of the model, which leads to improved target detection performance.

2.2.3 YOLOv8-FCS deep learning network structure

YOLO (You Only Look Once) is a typical single-stage target detection algorithm (Redmon et al., 2016; Redmon and Farhadi, 2017, 2018; Bochkovskiy et al., 2020; Wang et al., 2023). YOLOv8 is

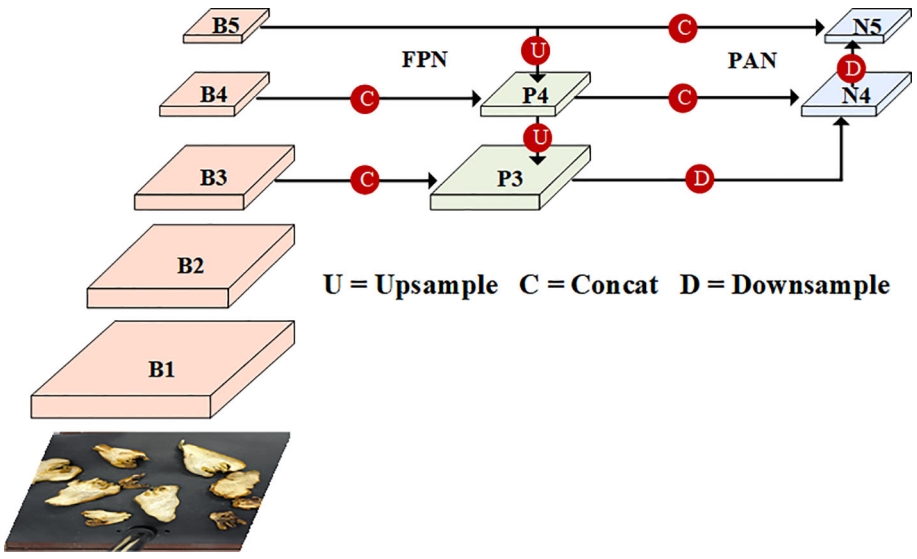


FIGURE 4
The structure of PAN-FPN.

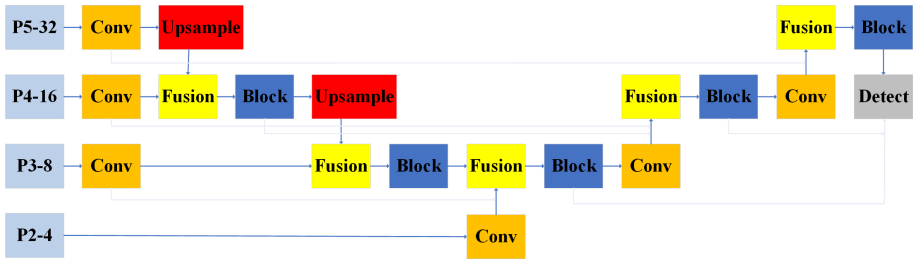


FIGURE 5
The structure of BiFPN.

the structure of YOLO family introduced in 2023, which performs target detection by means of a unique two-path prediction and tightly connected convolutional network. The algorithm employs a lightweight network structure while maintaining a high performance and is therefore efficient. In YOLOv8, the target detection task is decomposed into two independent subtasks, namely classification and localization. Each subtask has its own network path, which enables the algorithm to better handle targets of different sizes. In addition, YOLOv8 adopts the ideas of cascading and pyramiding, which enables the algorithm to deal with targets of different sizes. The feature extraction capability of YOLOv8 has been significantly improved compared to the previous model while maintaining a lightweight network structure. The main structure of YOLOv8 consists of Backbone, Neck, and Head, as shown in the Figure 6.

The challenge of evaluating and detecting fingered citron slices is constrained by the availability of computational resources, necessitating

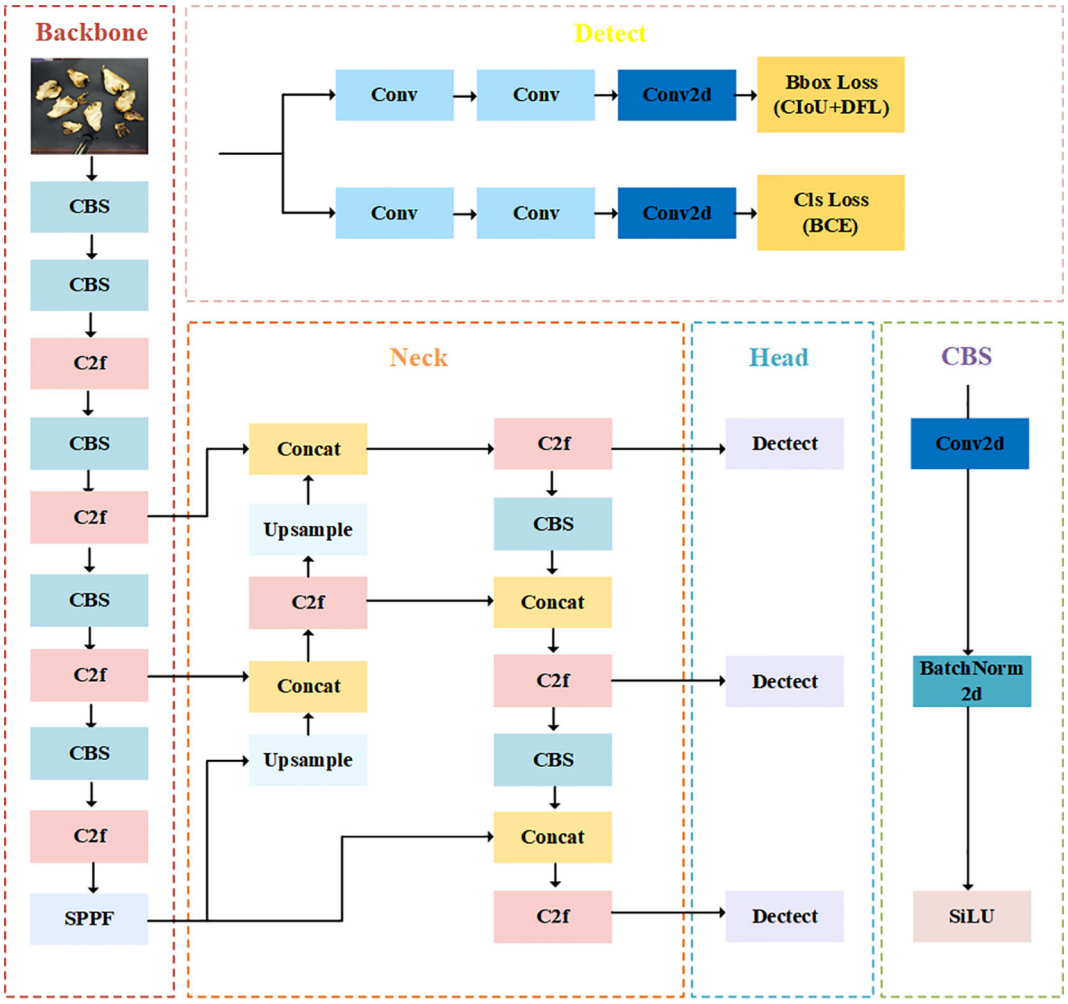


FIGURE 6
The main structure of YOLOv8.

models that are structurally straightforward, exhibit minimal delay, and offer substantial data processing capabilities. Traditional lightweight networks, such as MobileNet (Howard et al., 2017), ShuffleNet (Zhang et al., 2018), and GhostNet (Han et al., 2020), leverage deep convolution or swarm convolution to distill spatial features from visual data. While the primary objective of these lightweight models is to diminish the count of floating-point operations (FLOP), there is a dearth of research that has explored models with a low rate of floating-point operations per second (FLOPS). It's important to note that merely reducing the model's parameters does not completely equate to a proportional boost in computational velocity. Consequently, some studies have attempted to engineer nimble and rapid neural network components through the use of deep convolution or group convolution. However, these attempts do not always accelerate model performance and can sometimes even increase latency. In light of these considerations, this paper proposes the substitution of the Darknet-53 backbone network in YOLOv8n with the FasterNet-T0 network. This change is intended to augment the model's parameter count and computational capacity, thereby enhancing its computational efficiency.

In the YOLOv8 framework, the PAN-FPN structure is utilized, which represents an enhanced iteration of the classic FPN. This PAN-FPN configuration is designed to mitigate the issue of partial loss of localization details by integrating a bottom-up PAN architecture atop the FPN, thereby replenishing the missing localization data. While this approach does bolster the semantic and spatial information to an extent, the PAN-FPN structure is still susceptible to refinements. To more effectively tackle the limitations of PAN-FPN, particularly its challenges with handling large-scale feature maps and the loss of some pristine information in feature maps post-sampling, this study proposes a restructured feature fusion mechanism for YOLOv8n that draws inspiration from the BiFPN.

The modified network structure is shown in Figure 7, which consists of the Backbone module with the improved FasterNet structure, the Neck module reconfigured according to the BiFPN structure, and the head module of the original model, which named YOLOv8-FCS.

YOLOv8-FCS is able to extract features better and obtain better recognition accuracy in comparison with the base YOLOv8n model, as verified in experiments in Section 3. The improved model effectively improves the overall detection performance of the model yet becomes complicated to a small extent. From the subsequent experiments, it can be seen that the model achieves better results in the task of grading and detecting fingered citron slices.

2.3 Experimental platform training parameters

In this study, the experiments were run on Windows 10. The framework image source used is PyTorch 1.8.1, with Python 3.8 as the training environment and the Compute Unified Device Architecture (CUDA) 11.1 as the computing architecture. The GPU for the hardware part is RTX 2080×1 with 8 GB of its video memory. the CPU is a 4-core, 8-thread 12th Gen Intel(R) Core(TM) i3–12100F 3.30 GHz with 16 GB of RAM.

The settings of each hyperparameter in this experiment are shown in Table 2. In order to balance the different shooting devices, the input shape value of the input model was standardized to 640×640, and all of them were trained for 200 epochs, and the mosaic enhancement was turned off at the last 10 epochs. The optimizer was SGD, the batch size was set to 8. The maximum learning rate was 1e-2, the minimum learning rate was 1e-4. The momentum is set to 0.937, and the IoU threshold for the mean Average Precision (mAP) is set to 0.7.

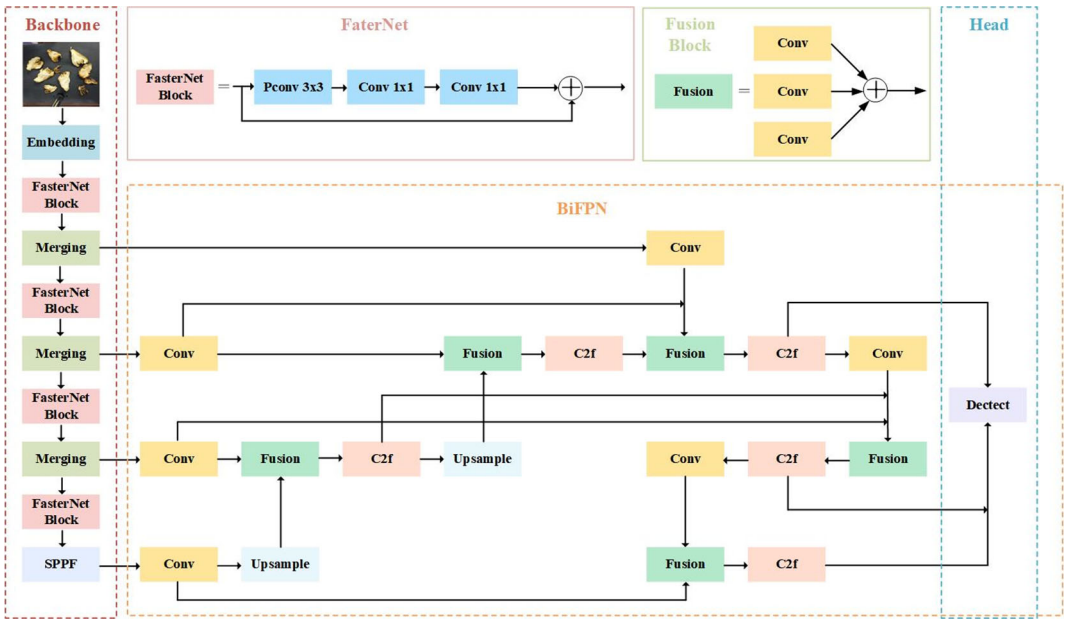


FIGURE 7
The main structure of YOLOv8-FCS.

TABLE 2 Hyperparameter settings.

Parameter	Setting
Input Shape	(640, 640, 3)
Train Epoch	200
Close Mosaic	10
Batch Size	8
Workers	8
Optimizer	SGD
Maximum Learning Rate	1e-2
Minimum Learning Rate	1e-4
Momentum	0.937
IoU	0.7

3 Results

3.1 Evaluation indicators

To more accurately assess the precision and assurance levels of the model's classifications, this study has chosen to employ the loss function curve (Loss), mean Average Precision (mAP), recall rate, the Giga Floating Point Operations Per second (GFLOPs), and the Frames Per Second (FPS) as the key metrics for evaluating the algorithm's performance. The calculation of related evaluation indexes is shown in Equations 1–5.

$$mAP = \frac{\sum_{1}^{N} AP}{N} = \frac{\sum_{1}^{N} \int_{0}^{1} P(R)dR}{N} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$Mean\ AveragePrecesion = \frac{\sum AveragePrecesion}{n(Class)} \quad (4)$$

$$F-score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

In our experiment, the True Positives (TP) denotes the count of instances that were accurately identified as positive instances, the True Negatives (TN) signifies the count of instances that were correctly classified as negative instances, the False Positives (FP) represents the instances that were incorrectly categorized as positive when they were actually negative, and the False Negatives (FN) corresponds to the instances that were mistakenly labeled as negative when they were positive. The Average Precision (AP) is calculated as the area beneath the Precision-Recall curve (P-R curve), which is a measure of the precision across various recall levels and represents the average precision. The mean Average Precision (mAP) is the mean value of the AP across different object categories. The n represents the total number of sample categories that were subjected to testing. The F-score is a statistical measure that balances the impact of precision and recall, providing a single score that encompasses both metrics for a more comprehensive assessment of the model's performance.

3.2 Impact of data enhancement

To explore the impact of data augmentation on the outcomes of the experiments, this study employs both a non-augmented dataset and an augmented dataset to perform tests on the YOLOv8n and YOLOv8-FCS models respectively. Both datasets are segmented in a 6:2:2 ratio for training, validation, and testing, and the training is carried out for 200 epochs. The findings from these experiments are presented in Figure 8.

Observations indicate that, when compared with the results obtained from the non-augmented dataset, both the YOLOv8n and YOLOv8-FCS models exhibit an improvement in their correlation metrics on the augmented dataset, with an enhancement range of 2% to 5%. It is worth mentioning that in around 190 epochs of this study, the model experienced a significant decrease in loss value. This is to prevent the training images generated by Mosaic data augmentation from deviating from the true distribution of natural images and introducing a large number of inaccurate annotations. Therefore, Mosaic data augmentation was turned off in the last 10 rounds of training, and the model began to learn real image data. This transformation significantly reduces label errors, leading to a rapid decrease in training loss values. These experimental outcomes confirm that the data augmentation techniques applied in this study play a significant role in bolstering the accuracy of the model's detection capabilities.

3.3 Comparison with YOLOv8 before improvements

Figure 9 show the comparison of the main algorithmic performance evaluation metrics of the pre-improved YOLOv8n model and the improved YOLOv8-FCS. During the training process, the loss function of YOLOv8-FCS shows a faster convergence trend, as well as a lower training loss. In addition, the improved YOLOv8-FCS model gets better results in precision, recall, and mAP metrics. This indicates that the improvement of this experiment enables the model to learn the feature information related to the grading of fingered citron slices faster, so that the convergence speed of the model increases and the training loss decreases, which leads to better training effect and accuracy.

Figure 10 illustrates the confusion matrix for both the original and the enhanced models across four categories of fingered citron slices. In this matrix, each row corresponds to the true class of the samples, each column corresponds to the class predicted by the model, and the diagonal elements indicate the count of samples correctly classified for each class. The data reveals that there is a high rate of accurate predictions for the first and bad grades of fingered citron slices, whereas the second and third-grades have a

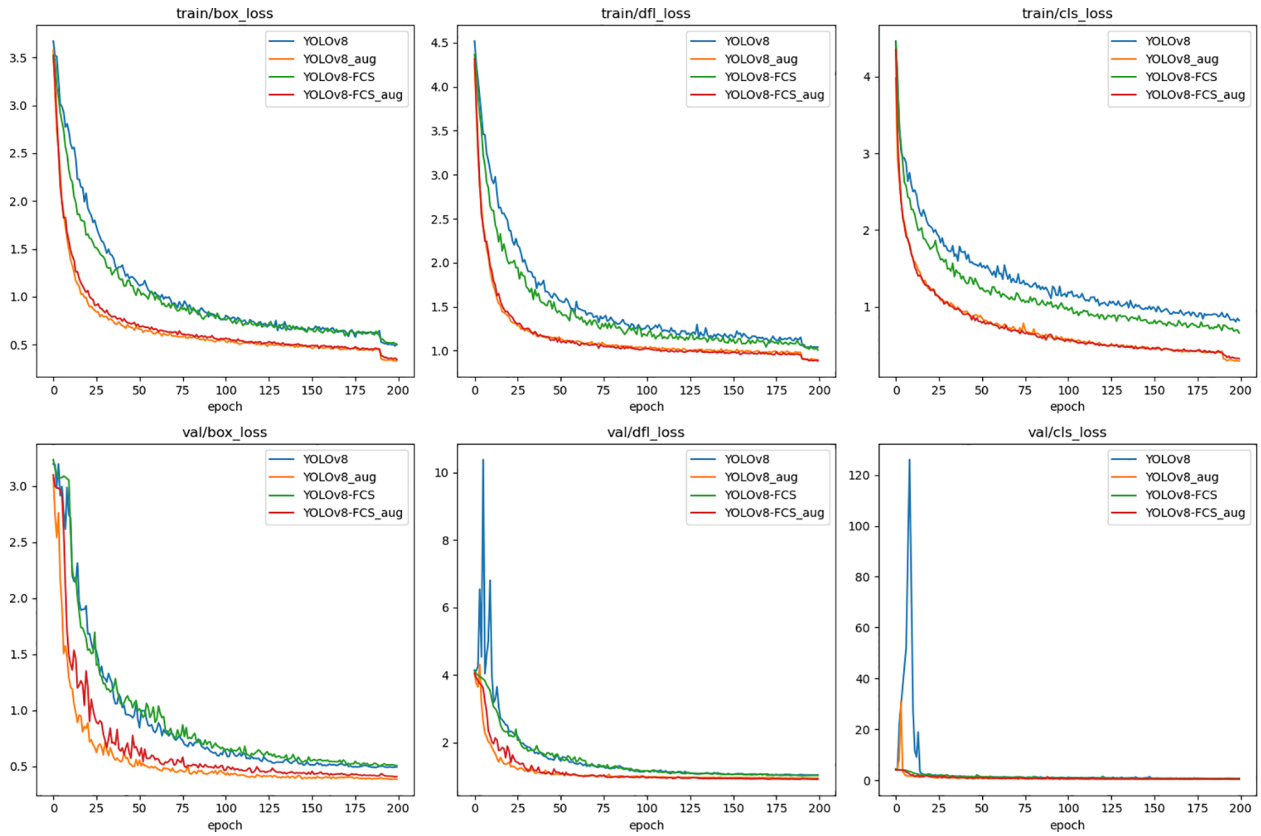


FIGURE 8
Loss chart before and after data augmentation.

comparatively lower accuracy in predictions. The confusion matrix highlights that there is a tendency for the model to confuse the first and second class goods. Upon examination of the images that led to incorrect classifications, it was observed that a significant number of these instances involved targets with occlusion or adhesion. Such conditions can result in certain features being obscured, causing defects and characteristics to be incorrectly identified due to the masking effect.

By comparing the confusion matrices of the models before and after the improvement, it can be found that the number of first-grade products misjudged as second-grade products is greatly reduced, which directly improves the judgment accuracy of first-grade products. At the same time, the number of second-grade and third-grade samples judged correctly are both improved to some extent. This is because we have redesigned the neck file of the model using the BiFPN structure to improve feature fusion capability. This suggests that the improvements in this experiment have a positive effect on the accurate identification of each class of fingered citron slices. However, from Figure 10B, it can be seen that the improved model has a slight decrease in the ability to distinguish between bad and third-grade products. This is because the weakened expression ability of the model in the feature extraction process for some subtle features of bergamot slices, resulting in some fuzzy recognition at the boundaries of these two categories. Overall, the prediction accuracy of the four categories of fingered citron slices is relatively high, indicating that the model is suitable for the classification detection task of fingered citron slices from a fixed perspective.

In order to further observe the effect of the improvement in this experiment, this study also tries to use HiRes-CAM (Esmaeili et al., 2021) to visualize the sample regions that are more concerned by the YOLOv8 and the improved YOLOv8-FCS model respectively, as shown in Figure 11. Different highlighted regions indicate the degree of influence of pixel points at different locations on the recognition results, where red highlighting indicates that the pixel values in the region have the greatest influence on the recognition results and get the highest convolutional attention. By comparison, we find that YOLOv8-FCS pays more attention to the target itself and focuses the attention constriction on the target. Furthermore, for the adhesion case, YOLOv8-FCS is able to distinguish the target better, instead of confusing the adhesion part. In addition, for problems such as misdetecting the background as the target, which existed in the YOLOv8n model, YOLOv8-FCS is able to avoid them to a large extent. In summary, it is further demonstrated that the YOLOv8-FCS model proposed in this paper can improve the learning ability of important features for grading samples of fingered citron slices.

In summary, the YOLOv8-FCS model proposed in this paper is superior to the YOLOv8n model in terms of comprehensive metrics, and is more suitable for the task of grading fingered citron slices, which is the focus of this experiment.

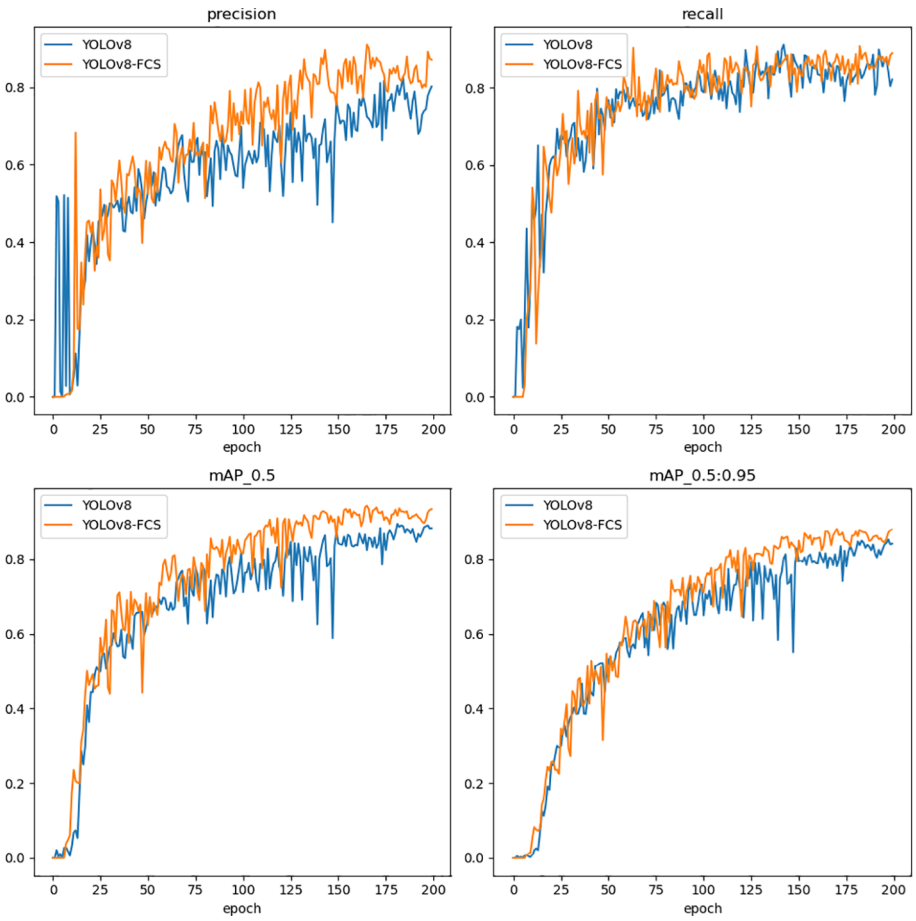


FIGURE 9
Comparison of the Evaluation indicator chart of YOLOv8 and YOLOv8-FCS.

3.4 Ablation experiments

In order to validate the feasibility and effectiveness of the enhanced model, the improved methodology is divided into two key components, namely, replacing the FasterNet module in the backbone network of YOLOv8n and redesigning the PAN-FPN structure of YOLOv8n using the BiFPN structure. Four different ablation experiments were conducted in order to fully assess the

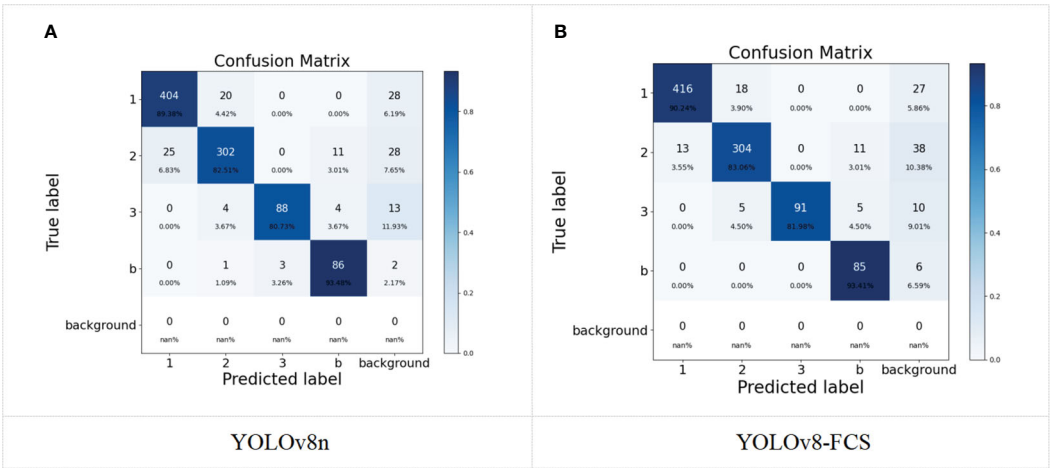


FIGURE 10
Confusion Matrix for YOLOv8n **(A)** and YOLOv8-FCS **(B)**.

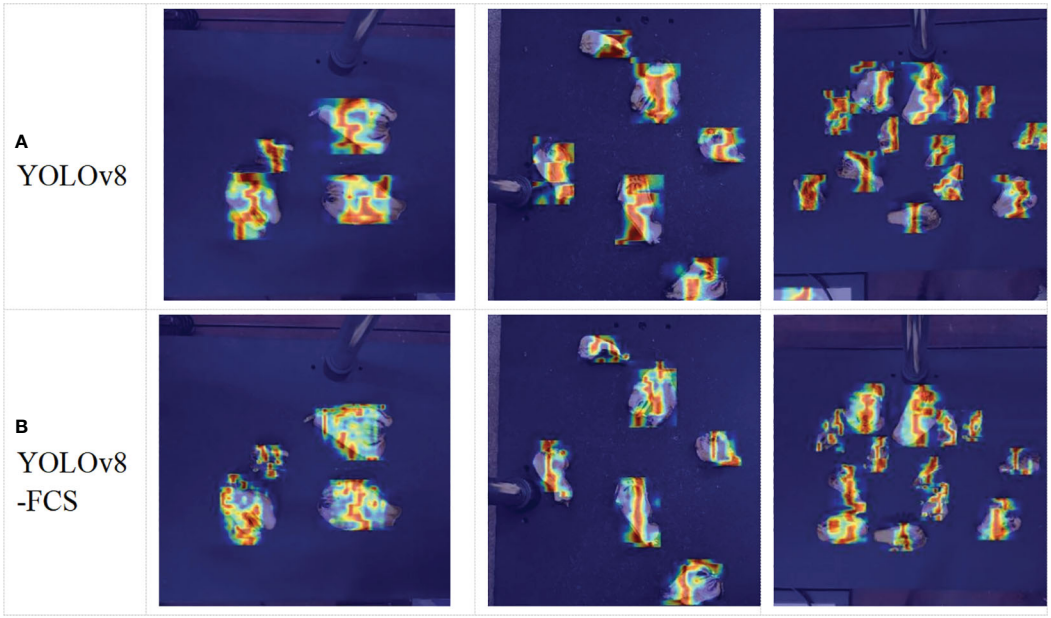


FIGURE 11
Heat maps of YOLOv8n **(A)** and YOLOv8-FCS **(B)**.

impact of changes and combinations of individual modules on the performance of the algorithm. The results of these experiments are combined and summarized in Table 3.

The purpose of this experiment is to evaluate the effect of different components on the performance of the YOLOv8n model by observing the changes in precision, recall, computation volume (GFLOPs), model size, Frames Per Second (FPS) transmitted on the screen, and mean Average Precision (mAP). The original YOLOv8n model achieves 93.3% precision and 90.9% recall, with a mAP of 96.7% and FPS of 194.3, while maintaining a low computational effort (8.1 GFLOPs) and model volume (5.95 MB). This shows that YOLOv8n has been designed to be efficient and accurate.

The design goal of FasterNet is to improve the computational efficiency of the model. After introducing FasterNet as the backbone network (YOLOv8-FasterNet), the computational volume of the model increased to 10.7 GFLOPs, the size increased to 8.20 MB, the precision and recall were improved, the mAP increased to 97.8%, and the recall reached the highest value of 0.961 for this ablation experiment. Although FasterNet optimizes the computational efficiency, this optimization comes at the cost of increasing the size and complexity of the model. This is due to the fact that FasterNet introduces more parameters or more complex network structures, which increases the model's representational power. The increased model complexity allows FasterNet to capture more features and details, which is directly reflected in improved precision and recall. Increased precision means that the model correctly identifies an increased percentage of targets, while increased recall means that the model is able to identify more real targets and reduces missed detections. The mAP is increased to show that the model performs well under a wide range of conditions. The recall rate reaches the highest value in the experiment, which further confirms the effectiveness of FasterNet in the target detection task.

BiFPN as a multi-scale feature fusion technique, is able to reduce the size and computational effort of the model while maintaining performance. When BiFPN is added as the neck structure (YOLOv8-BiFPN), the model volume reaches the lowest value of 4.15 MB for this ablation experiment, and the amount of computation is reduced to 7.5 GFLOP. This is due to BiFPN's ability to efficiently integrate features at different scales, which reduces the need for a large number of parameters. Despite the reduction in model size and computation, the addition of BiFPN still results in a slightly better performance than the original model. This suggests that BiFPN does a good job of integrating multi-scale information

TABLE 3 Results of ablation experiment.

Model	Precision	Recall	mAP	FPS	GFLOPs	ModelSize
YOLOv8n	0.939	0.909	0.967	**194.3**	8.1	5.95MB
YOLOv8-Fasternet	0.916	**0.961**	0.978	138.2	10.7	8.20MB
YOLOv8-BiFPN	0.931	0.943	0.973	169.1	**7.5**	**4.15MB**
YOLOv8-FCS	**0.961**	0.949	**0.981**	130.3	10.1	6.40MB

The bold values represent the most outstanding performance among the evaluation indicators.

and is able to improve the model's ability to detect targets of different sizes. It is mentioned in the paper that the improvements in BiFPN are useful for a specific task that fingered citron slice grading detection. This is because the task requires accurate identification and classification of slices of different sizes and shapes, and BiFPN is able to provide just this kind of multi-scale feature fusion.

When the two improved components are used together (YOLOv8-FCS), the volume of the model is 6.40 MB and the computational volume increases to 10.1 GFLOPs, which is higher compared to the original YOLOv8n model but lower than the model that replaces the FasterNet structure alone. This suggests that the combination of the two improves computational efficiency while keeping the computational cost under control. Meanwhile, the improved precision and recall reached 96.1% and 94.9%, respectively. mAP was 98.1%, which was the highest among all models. This result shows that the combination of YOLOv8n, FasterNet and BiFPN maximizes the performance of the model and achieves a balance between high efficiency and high precision. Although the performance of the two improved models is lower than any of the improved as well as the base model in terms of FPS, it still reaches a respectable value of 130.3, which is fully adequate for deployment.

To summarize, the experimental results show that the detection performance of the YOLOv8n model can be significantly improved by reasonably increasing the number of parameters and computational volume of the model. In particular, the FasterNet module plays a key role in improving precision and recall, and BiFPN performs well in controlling the model volume and computation. Meanwhile, FasterNet and BiFPN optimize the feature extraction and multi-scale feature fusion capabilities of the model to different degrees. The combined use of these improved components provides an effective way to achieve more efficient and accurate target detection.

3.5 Comparison with other models

In order to fully evaluate the performance of the YOLOv8-FCS model, we compare it with several other popular target detection models, including RT-DETR (Zhao et al., 2023), YOLOv7-tiny (Wang et al., 2023), YOLOv5n, YOLOv5s, SSD (Liu et al., 2016) and FasterR-CNN (Ren et al., 2016). The experimental results of these models on the same dataset are shown in Table 4.

As can be seen from Table 4, the mAP value of the YOLOv8-FCS model proposed in this paper reaches 98.1% with the same experimental setup. In contrast, RT-DETR-resnet18 (92.4%), the faster FasterR-CNN (86.3%), SSD (90.7%), YOLOv5n (94.7%), YOLOv5s (97.9%), and YOLOv7 (97.7%) are not as good as the YOLOv8-FCS model in terms of mAP. Comparison experiments show that the real-time detection rate (FPS) of both Faster R-CNN (18.4) and SSD (26.6) do not meet the real-time requirements. While YOLOv5n (190.2), YOLOv5s (124.3) and YOLOv7-tiny (143.8) perform well in this regard. The YOLOv8-FCS model (130.3) proposed in this paper does not have the highest real-time detection rate, but it fully meets the practical deployment requirements. YOLOv7-tiny and YOLOv5s are 1–2 of the size of this experimental model, but their detection accuracy, recall and mAP values are not gap with the YOLOv8-FCS model. In contrast, the size of the improved algorithm model proposed in this paper is 6.4 MB, which is much smaller than Faster R-CNN, RT-DETR-resnet18 and SSD models, and only slightly larger than the YOLOv5n model, but it performs much better in the evaluation metrics.

In summary, the YOLOv8-CFS model proposed in this paper not only meets the real-time detection requirements, but also improves the detection accuracy, minimizes the model size, and exchanges a very small arithmetic cost for a large performance improvement, which has higher generality and practical value. This result not only demonstrates the superiority of YOLOv8-CFS in target detection tasks, but also shows that the detection accuracy of the model can be significantly improved without significantly affecting the computational efficiency through our well-designed network structure improvement.

4 Discussion

4.1 Feasibility analysis

In order to promote the intelligent development of fingered citron slices industry, this study applied the YOLOv8n model to the task of grading and detecting fingered citron slices under the fixed perspective, and improved the YOLOv8n for the limitations, and

TABLE 4 Results of comparison test.

Model	Precision	Recall	mAP	FPS	ModelSize
SSD	0.836	0.899	0.907	26.6	92.1MB
FasterR-CNN	0.754	0.937	0.863	18.4	108MB
RT-DETR-resnet18	0.936	0.913	0.924	46.7	38.5MB
YOLOv5n	0.945	0.893	0.947	**190.2**	**5.02MB**
YOLOv5s	0.938	**0.957**	0.979	124.3	17.60MB
YOLOv7-tiny	0.955	0.948	0.977	143.8	12.30MB
YOLOv8-FCS	**0.961**	0.949	**0.981**	130.3	6.40MB

The bold values represent the most outstanding performance among the evaluation indicators.

finally obtained the YOLOv8-FCS fingered citron slices grading and detecting model. The results show that the YOLOv8-FCS model can accomplish the fingered citron slices grading detection task better.

The size of the YOLOv8-FCS network, as introduced in this study, is 6.4MB, marking an increment of merely 0.4MB over the standard YOLOv8 model. The model requires a mere 0.046 seconds to process a single image within the experimental setup, which includes a GPU (RTX 2080 with 8GB of video memory). The evaluation metrics of the refined model significantly surpass those of other prevalent models typically used for object detection tasks. The empirical findings indicate that our model aligns well with hardware deployment requirements, and its processing velocity satisfies practical demands. Consequently, the model demonstrated in this research holds a distinct accuracy advantage over existing models. It is capable of effectively executing the grading detection task for fingered citron slices from fixed viewpoints and holds promise for future integration into assembly lines or sorting machinery designed for fingered citron slices.

4.2 Contribution to the intellectualization of the fingered citron slices industry

At present, the majority of the fingered citron slices sector relies on conventional manual methods. These methods are characterized by low efficiency, substantial labor expenses, and a significant degree of subjectivity. Consequently, there is a pressing need to explore intelligent processing solutions within the fingered citron slices industry to address these shortcomings. Manual sorting of fingered citron slices is a labor-intensive task that requires a lot of manpower, and the use of mechanical grading helps to reduce the labor demand and can improve grading accuracy and avoid the influence of subjectivity. In this study, a deep learning algorithm is utilized to achieve the task of grading fingered citron slices detection under fixed viewpoint. We propose the YOLOv8-FCS model to complete the task of grading fingered citron slices, which has performed well in testing and can process 130 images per second (FPS=130.3). This processing speed not only meets the requirements of industrial applications, but also has the potential for practical deployment on fingered citron slices processing flowlines or automatic sorting machinery. The efficiency of the YOLOv8 FCS model means that it can significantly improve the automation level and production efficiency of the classification of fingered citron slices while maintaining high accuracy. In addition, the design flexibility and scalability of this model also provide possibilities for its application in different working environments, thereby providing strong technical support for the automatic grading of fingered citron slices and the automation of subsequent packaging, sales, and other processes.

Non-intrusive examination techniques have become extensively integrated into agricultural practices and are often merged with a variety of innovative technologies (Kondo, 2010), signaling a trend towards the automation of post-harvest activities as the next frontier in agricultural advancement. The utilization of machine vision technology enables swift and dependable evaluation of agricultural commodities like fruits and vegetables, with non-destructive approaches that can minimize financial losses while simultaneously boosting productivity and economic gain. In the context of this research, the multi-input model has been applied to the task of grading and detecting fingered citron slices, representing an instance of processed produce in sliced form. As such, the model exhibits considerable promise for extending its utility to the grading of other sliced herbs and potentially for the quality assessment of fruits and vegetables more broadly.

4.3 Limitations and future work

Compared with the original YOLOv8n model, the improved YOLOv8-FCS model in this paper improves the accuracy of fingered citron slices grading detection. Nevertheless, a small number of false detection exist, which may be attributed to the following factors. The first reason is related to the features of the images. The difference features between the grades of fingered citron slices are more variable, which increases the detection challenge. The second reason is the complexity of the datasets, where the captured images not only have scattered targets of fingered citron slices, but also partially adhered and overlapped. These factors tend to mask the differential features of fingered citron slices, leading to false detection. The third problem is the limitation of the datasets. The datasets of this experiment were all taken from the same batch of fingered citron slices, which may not be representative enough. The above problems indicate that the model has some room for improvement in dealing with the problem of datasets with complex features. In the future, we will evaluate the performance of the datasets under more equipment shots and try to add other batches and provenances of fingered citron slices to enhance the representativeness of the datasets.

5 Conclusion

In this study, we apply deep learning techniques to the grading detection of fingered citron slices, and improve the YOLOv8n model to get the YOLOv8-FCS model which is more effective in grading detection of fingered citron slices. Firstly, we enhanced and labeled the image data of fingered citron slices collected under the fixed viewing angle to form the fingered citron slices datasets. Then we replace the YOLOv8n backbone structure with the FasterNet main module to improve the computational efficiency in the feature extraction process. The PAN-FPN structure used in the original model is redesigned with the BiFPN structure in order to balance the computational volume and model size while making full use of the high-resolution features to expand the sensory field of the model. In the end, the detection accuracy is improved significantly at a small cost of arithmetic power. The main conclusions are as follows:

(1) The mAP value of the YOLOv8-FCS model proposed in this study reaches 98.1%, which is improved by 1.4% compared with the YOLOv8n model. It is found that the YOLOv8-FCS model is able to achieve higher detection accuracy with a very small model size and computation enhancement, which alleviates the YOLOv8n model's

misdetection to a certain extent. Therefore, we believe that the improved method we used is effective and the YOLOv8-FCS model can effectively provide technical support for grading fingered citron slices.

(2) Following a comparative evaluation, it has been observed that the YOLOv8-FCS model outperforms current mainstream object detection networks in terms of accuracy for the task of grading detection of fingered citron slices. The model size of 6.4M is acceptable. The arithmetic capacity is 10.1GLOPs, which is very low, and the FPS is 130, which can meet the practical use requirements. The experimental results show that YOLOv8-FCS can accurately accomplish the task of fingered citron slices grading detection and can run on lightweight devices.

(3) The method proposed in this study can well accomplish the grading detection of fingered citron slices under the fixed viewpoint, which effectively reduces the problems of low efficiency, high cost, and great influence by subjectivity that exist in the traditional manual detection. Boasting both high precision and real-time capabilities, this model is well-suited to fulfill the grading requirements of agricultural producers and small to medium-sized enterprises. It holds significant practical value for application within the industry dedicated to the grading of fingered citron slices, contributing to the advancement of a standardized and efficient fingered citron slices market.

(4) Given that the primary focus of this research is on the grading detection of fingered citron slices, a type of herbaceous slice, future work may explore the potential of adapting the methodologies presented in this paper to the evaluation of other sliced agricultural products. In addition, we will also consider the use of more informative sensors such as multispectral cameras for image acquisition in order to grade fingered citron slices more accurately, authoritatively, and efficiently by combining the scales of active ingredient content.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

LZ: Writing – original draft, Writing – review & editing, Conceptualization, Formal analysis, Investigation, Methodology, Software, Supervision, Visualization. PL: Data curation, Methodology, Supervision, Writing – original draft. SD: Investigation, Validation, Writing – review & editing. TL: Validation, Writing – review & editing. KQ: Methodology, Supervision, Writing – original draft. JM: Writing – original draft, Funding acquisition, Methodology, Project administration, Resources, Supervision.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study has been funded by the Science and Technology Department of Sichuan Province (Contract Number: 2022NSFSC0172 and 22ZDYF0095).

Acknowledgments

Thanks to Yi Li, Kebei Qin, Jingsha Shi and Pu Luo for his help in data acquisition and Professor Fan Liu from Sichuan Agricultural University for her valuable comments.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Asriny, D. M., Rani, S., and Hidayatullah, A. F. (2020). Orange fruit images classification using convolutional neural networks. *IOP Conf. series: Materials Sci. Eng.* 803, 12020. doi: 10.1088/1757-899X/803/1/012020

Bochkovskiy, A., Wang, C., and Liao, H. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv [Preprint]*. doi: 10.48550/arXiv.2004.10934

Castro, W., Oblitas, J., De-La-Torre, M., Cotrina, C., Bazán, K., and Avila-George, H. (2019). Classification of cape gooseberry fruit according to its level of ripeness using machine learning techniques and different color spaces. *IEEE Access* 7, 27389–27400. doi: 10.1109/ACCESS.2019.2898223

Chakraborty, S. K., Subeesh, A., Dubey, K., Jat, D., Chandel, N. S., Potdar, R., et al. (2023). Development of an optimally designed real-time automatic citrus fruit grading–sorting machine leveraging computer vision-based adaptive deep learning model. *Eng. Appl. Artif. Intell.* 120, 105826. doi: 10.1016/j.engappai.2023.105826

Chen, M., Chen, Z., Luo, L., Tang, Y., Cheng, J., Wei, H., et al. (2024a). Dynamic visual servo control methods for continuous operation of a fruit harvesting robot

working throughout an orchard. *Comput. Electron. Agric.* 219, 108774. doi: 10.1016/j.compag.2024.108774

Chen, J., Kao, S. H., He, H., Zhuo, W., Wen, S., Lee, C., et al. (2023). "Run, don't walk: chasing higher FLOPS for faster neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Vancouver, BC, Canada), 12021–12031. doi: 10.1109/CVPR52729.2023.01157

Chen, X., Chen, H., Xiao, J., Liu, J., Tang, N., and Zhou, A. (2020). Variations of volatile flavour compounds in finger citron (Citrus medica L. var. sarcodactylis) pickling process revealed by E-nose, HS-SPME-GC-MS and HS-GC-IMS. *Food Res. Int.* 138, 109717. doi: 10.1016/j.foodres.2020.109717

Chen, W., Liu, M., Zhao, C., Li, X., and Wang, Y. (2024b). MTD-YOLO: Multi-task deep convolutional neural network for cherry tomato fruit bunch maturity detection. *Comput. Electron. Agric.* 216, 108533. doi: 10.1016/j.compag.2023.108533

De Luna, R. G., Dadios, E. P., Bandala, A. A., and Vicerra, R. R. P. (2019). "Tomato fruit image dataset for deep transfer learning-based defect detection," in *Proceedings of the 2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)*. (Bangkok, Thailand), 356–361. doi: 10.1109/CIS-RAM47153.2019.9095778

Esmaeili, M., Vettukattil, R., Banitalebi, H., Krogh, N. R., and Geitung, J. T. (2021). Explainable artificial intelligence for human-machine interaction in brain tumor localization. *J. Pers. Med.* 11 (11), 1213. doi: 10.3390/jpm11111213

Fan, S., Li, J., Zhang, Y., Tian, X., Wang, Q., He, X., et al. (2020). On line detection of defective apples using computer vision system combined with deep learning methods. *J. Food Eng.* 286, 110102. doi: 10.1016/j.jfoodeng.2020.110102

Fu, Y., Nguyen, M., and Yan, W. Q. (2022). Grading methods for fruit freshness based on deep learning. *SN Comput. Sci.* 3, 264. doi: 10.1007/s42979-022-01152-7

Gui, J., Zhang, Q., Hao, L., and Bao, X. (2014). Apple shape classification method based on wavelet moment. *Sensors transducers* 178, 182.

Gururaj, N., Vinod, V., and Vijayakumar, K. (2023). Deep grading of mangoes using convolutional neural network and computer vision. *Multimedia Tools Appl.* 82, 39525–39550. doi: 10.1007/s11042-021-11616-2

Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., and Chang, X. (2020). "GhostNet: more features from cheap operations," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. (Seattle, USA), 1580–1589. doi: 10.48550/arXiv.1911.11907

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). "MobileNets: efficient convolutional neural networks for mobile vision applications," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. (Honolulu, Hawaii, USA). doi: 10.48550/arXiv.1704.04861

Ismail, N., and Malik, O. A. (2022). Real-time visual inspection system for grading fruits using computer vision and deep learning techniques. *Inf. Process. Agric.* 9, 24–37. doi: 10.1016/j.inpa.2021.01.005

Karp, D., and Hu, X. (2018). The citron (Citrus medica L.) in China. *Hortic. Rev.* 45, 143–196. doi: 10.1002/9781119431077.ch5

Klein, J. D. (2014). "Citron cultivation, production and uses in the Mediterranean region," in *Medicinal aromatic Plants Middle-East, medicinal and aromatic plants of the world*, vol 2. (Springer, Dordrecht), 199–214. doi: 10.1007/978–94-017– 9276-9_10

Kondo, N. (2010). Automation on fruit and vegetable grading system and food traceability. *Trends Food Sci. Technol.* 21, 145–152. doi: 10.1016/j.tifs.2009.09.002

Li, Z., Cai, M., Liu, Y., and Sun, P. (2018). Development of finger citron (Citrus medica L. var. sarcodactylis) essential oil loaded nanoemulsion and its antimicrobial activity. *Food Control* 94, 317–323. doi: 10.1016/j.foodcont.2018.07.009

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., et al. (2016). "Ssd: Single shot multibox detector," in *Proceedings of the computer vision–ECCV 2016: 14th European conference*. (Amsterdam, Netherlands). doi: 10.1007/978–3-319–46448-0_2

Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). "Path aggregation network for instance Segmentation," in *IEEE conference on computer vision and pattern recognition (CVPR)*. (Hong Kong, China), 8759–8768. doi: 10.48550/arXiv.1803.01534

Luo, X., Wang, J., Chen, H., Zhou, A., Song, M., Zhong, Q., et al. (2020). Identification of flavoanoids from finger citron and evaluation on their antioxidative and antiaging activities. *Front. Nutr.* 7. doi: 10.3389/fnut.2020.584900

Meena, A. K., Kandale, A., Rao, M. M., Panda, P., and Govind, R. (2011). A review on citron-pharmacognosy, phytochemistry and medicinal uses. *Int. Res. J. Pharm.* 2, 14.

Moallem, P., Serajoddin, A., and Pourghassem, H. (2017). Computer vision-based apple grading for golden delicious apples based on surface features. *Inf. Process. Agric.* 4, 33–40. doi: 10.1016/j.inpa.2016.10.003

Momeny, M., Jahanbakhshi, A., Jafarnezhad, K., and Zhang, Y. (2020). Accurate classification of cherry fruit using deep CNN based on hybrid pooling approach. *Postharvest Biol. Technol.* 166, 111204. doi: 10.1016/j.postharvbio.2020.111204

Nawaz, S. A., Li, J., Bhatti, U. A., Bazai, S. U., Zafar, A., Bhatti, M. A., et al. (2021). A hybrid approach to forecast the COVID-19 epidemic trend. *PloS One* 16, e0256971. doi: 10.1371/journal.pone.0256971

Patel, H. B., and Patil, N. J. (2024). An intelligent grading system for automated identification and classification of banana fruit diseases using deep neural network. *Int. J. Computing Digital Syst.* 15, 761–773.

Peng, C., Ker, Y., Weng, C., Peng, C., Huang, C., Lin, L., et al. (2009). Insulin secretagogue bioactivity of finger citron fruit (Citrus medica L. var. Sarcodactylis Hort, Rutaceae). *J. Agric. Food Chem.* 57, 8812–8819. doi: 10.1021/jf902143x

Raikar, M. M., S M, M., Kuchanur, C., Girraddi, S., and Benagi, P. (2020). Classification and Grading of Okra-ladies finger using Deep Learning. *Proc. Comput. Sci.* 171, 2380–2389. doi: 10.1016/j.procs.2020.04.258

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788.

Redmon, J., and Farhadi, A. (2017). "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (Honolulu, Hawaii, USA). 7263–7271. doi: 10.48550/arXiv.1612.08242

Redmon, J., and Farhadi, A. (2018). "YOLOv3: an incremental improvement". in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Salt Lake City, Utah, USA). doi: 10.48550/arXiv.1804.02767

Ren, S., He, K., Girshick, R., and Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Tan, M., Pang, R., and Le, Q. V. (2020). "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. (Seattle, USA). 10781–10790. doi: 10.48550/arXiv.1911.09070

Wang, C., Bochkovskiy, A., and Liao, H. M. (2023). "YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. (Vancouver, BC, Canada), 7464–7475. doi: 10.48550/arXiv.2207.02696

Wang, F., You, H., Guo, Y., Wei, Y., Xia, P., Yang, Z., et al. (2020). Essential oils from three kinds of fingered citrons and their antibacterial activities. *Ind. Crops Products* 147, 112172. doi: 10.1016/j.indcrop.2020.112172

Wu, Z. (2015). Effect of different drying methods on chemical composition and bioactivity of finger citron polysaccharides. *Int. J. Biol. Macromolecules* 76, 218–223. doi: 10.1016/j.ijbiomac.2015.02.043

Xu, W., Islam, M. N., Cao, X., Tian, J., and Zhu, G. (2021a). Effect of relative humidity on drying characteristics of microwave assisted hot air drying and qualities of dried finger citron slices. *Lwt* 137, 110413. doi: 10.1016/j.lwt.2020.110413

Xu, W., Pei, Y., Tian, J., Cao, X., Li, G., Jiang, Y., et al. (2021b). Effects of different drying methods on sensory qualities and aroma compounds of finger citron (Citrus medica L. var. sarcodactylis Swingle) slices. *J. Food measurement characterization* 15, 4465–4474. doi: 10.1007/s11694-021-01008-3

Zhang, S. D., Yang, H. Y., Zeng, J., and Li, M. (2018). Research progress on Citri sarcodactylis fructus. *China J. Traditional Chin. Med. Pharm.* 33, 3510–3514.

Zhang, X., Zhou, X., Lin, M., and Sun, J. (2018). "ShuffleNet: an extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. (Salt Lake City, Utah, USA), 6848–6856. doi: 10.48550/arXiv.1707.01083

Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., et al. (2023). Detrs beat yolos on real-time object detection. *arxiv [Preprint]* arxiv:2304.08069. doi: 10.48550/arXiv.2304.08069

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 19 June 2024
DOI 10.3389/fpls.2024.1411510

Check for updates

OPEN ACCESS

EDITED BY
Ting Sun,
China Jiliang University, China

REVIEWED BY
Hongwei Sun,
Hangzhou Dianzi University, China
Yanjie Li,
Chinese Academy of Forestry, China

*CORRESPONDENCE
Zhaomei Qiu
qzm381@126.com
Xin Jin
jx.771@163.com

RECEIVED 03 April 2024
ACCEPTED 31 May 2024
PUBLISHED 19 June 2024

CITATION
Qing S, Qiu Z, Wang W, Wang F, Jin X, Ji J, Zhao L and Shi Y (2024) Improved YOLO-FastestV2 wheat spike detection model based on a multi-stage attention mechanism with a LightFPN detection head.
Front. Plant Sci. 15:1411510.
doi: 10.3389/fpls.2024.1411510

COPYRIGHT
© 2024 Qing, Qiu, Wang, Wang, Jin, Ji, Zhao and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Improved YOLO-FastestV2 wheat spike detection model based on a multi-stage attention mechanism with a LightFPN detection head

Shunhao Qing[1], Zhaomei Qiu[1]*, Weili Wang[1], Fei Wang[1], Xin Jin[1,2]*, Jiangtao Ji[1], Long Zhao[1] and Yi Shi[1]

[1]College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang, Henan, China, [2]Science and Technology Innovation Center for Completed Set Equipment, Longmen Laboratory, Luoyang, China

The number of wheat spikes has an important influence on wheat yield, and the rapid and accurate detection of wheat spike numbers is of great significance for wheat yield estimation and food security. Computer vision and machine learning have been widely studied as potential alternatives to human detection. However, models with high accuracy are computationally intensive and time consuming, and lightweight models tend to have lower precision. To address these concerns, YOLO-FastestV2 was selected as the base model for the comprehensive study and analysis of wheat sheaf detection. In this study, we constructed a wheat target detection dataset comprising 11,451 images and 496,974 bounding boxes. The dataset for this study was constructed based on the Global Wheat Detection Dataset and the Wheat Sheaf Detection Dataset, which was published by PP Flying Paddle. We selected three attention mechanisms, Large Separable Kernel Attention (LSKA), Efficient Channel Attention (ECA), and Efficient Multi-Scale Attention (EMA), to enhance the feature extraction capability of the backbone network and improve the accuracy of the underlying model. First, the attention mechanism was added after the base and output phases of the backbone network. Second, the attention mechanism that further improved the model accuracy after the base and output phases was selected to construct the model with a two-phase added attention mechanism. On the other hand, we constructed SimLightFPN to improve the model accuracy by introducing SimConv to improve the LightFPN module. The results of the study showed that the YOLO-FastestV2-SimLightFPN-ECA-EMA hybrid model, which incorporates the ECA attention mechanism in the base stage and introduces the EMA attention mechanism and the combination of SimLightFPN modules in the output stage, has the best overall performance. The accuracy of the model was P=83.91%, R=78.35%, AP= 81.52%, and F1 = 81.03%, and it ranked first in the

GPI (0.84) in the overall evaluation. The research examines the deployment of wheat ear detection and counting models on devices with constrained resources, delivering novel solutions for the evolution of agricultural automation and precision agriculture.

KEYWORDS

YOLO-FastestV2, wheat spike, efficient channel attention, Efficient Multi-Scale Attention, SimLightFPN

1 Introduction

The yield and quality of wheat, one of the most important food crops in the world, are directly related to the sustainable development of agriculture as well as the guarantee of global food security (Misra et al., 2021). The number of wheat spikes is a key determinant of the number of wheat grains per unit area and the yield. Therefore, the number of wheat spikes is an important indicator of wheat yield (Slafer et al., 2014; Ferrante et al., 2017; Jin et al., 2017). The Measurement of the number of wheat spikes is important for fine-tuning the management of agricultural production, allocation of agricultural resources, and prediction of wheat yield (Pask, 2012; Zhou et al., 2021).

Currently, the methods used for the detection and counting of wheat spikes are categorized into manual counting, traditional machine learning detection and counting, and deep learning-based detection and counting methods. The manual counting method is cumbersome, labor-intensive, and subjective (Xiong et al., 2019). In traditional machine learning approaches, the detection and counting of wheat spikes usually relies on the selection of features, such as the shape, texture, and color of the ears, which are subsequently used to construct classification models for automated recognition and counting of wheat spikes (Zhao et al., 2014; Grillo et al., 2017; Ganeva et al., 2022). However, the method of manually extracting wheat features has the disadvantages of complex design, weak migration, and cumbersome manual design, which are less effective for applications in scenes with dense wheat spikes (Kamilaris and Prenafeta-Boldú, 2018; Zhang et al., 2020). Deep learning techniques have demonstrated excellent capabilities in the field of computer vision, particularly in the tasks of detecting and counting intensive objects, such as wheat spikes (Ye et al., 2023). These methods can automatically learn and extract key features from massive amounts of image data, thus effectively solving the complexity of manual feature extraction and the mobility problem in different scenarios using traditional methods (Sermanet et al., 2013). Therefore, deep learning provides a solution for the wheat spike counting task that is both efficient and accurate.

Many researchers have used deep-learning-based detection counting for detection and counting in agricultural-intensive scenarios (Aich et al., 2018; Wosner et al., 2021; Alkhudaydi and De La Iglesia, 2022). Common algorithms used for target detection in agriculture fall into two main categories: algorithms for single- and two-stage detection. One-stage algorithms, exemplified by Single Shot MultiBox Detector (SSD) (Liu et al., 2016), EfficientDet (Tan et al., 2020), and the YOLO family of algorithms (Redmon et al., 2016), directly classify and localize targets. Conversely, two-stage algorithms, such as SPP-Net (He et al., 2015), Mask R-CNN (He et al., 2017), and Faster R-CNN (Ren et al., 2015), first propose regions using RPN before refining classification and localization. Li et al. (2021) used the R-CNN method to detect, count, and analyze wheat spikes. The findings indicated that although the method demonstrated high recognition accuracy, it exhibited a slow detection speed and was unsuitable for deployment in a real-time detection device. Zhao et al. (2021) improved YOLOv5 by cleaning and enhancing data, adding a microscale detection layer, and adjusting the confidence loss function. They achieved an average accuracy of 94.1% in detecting wheat spikes in UAV images, surpassing the standard YOLOv5 by 10.8%. Li et al. (2022) used the Faster R-CNN model for image-based detection and metrology of the number of wheat spikes per unit area, demonstrating its practical application. Zang et al. (2022) introduced an efficient channel attention module (ECA) into the C3 module of the backbone structure of the YOLOv5s network model and simultaneously inserted a global attention mechanism module (GAM) between the neck and head structures. The results showed that the improved YOLOv5s method improved its applicability in complex field environments with better results. Wang et al. (2021) enhanced the wheat spike counting target detection model using Convolutional Block Attention Module (CBAM) in EfficientDet-D0, and the results of the study showed that the improved EfficientDet-D0 model achieves a counting accuracy of 94%, which is approximately 2% higher than that of the original model, and has a greater improvement in the occlusion problem. Wen et al. (2022) developed SpikeRetinaNet, a model consisting of a bi-directional feature pyramid network and a feature pyramid network, for detecting and counting the number of wheat spikes in field datasets for global wheat spike detection (GWHD). In general, among the various detection algorithms, single-stage algorithms have the advantages of higher speed and easier deployment in mobile segments, whereas two-stage algorithms, although they have higher accuracy and achieve better results, require large floating-point operations, and cannot achieve real-

time recognition in edge devices with limited computational and storage resources (He et al., 2017; Ge et al., 2021). However, in practice, even single-stage algorithms often need to run on mobile devices that are more capable to ensure detection speed, which may not be sufficiently fast to meet real-time requirements on average mobile devices.

To achieve effective deployment and application of the model to mobile devices, researchers have developed a series of lightweight network architectures that can efficiently perform target detection tasks on mobile devices with limited resources (Zhang et al., 2021; Wang et al., 2022; Ma et al., 2023). Xu et al. (2020) introduced a hybrid YOLO V3-Lite lightweight model by incorporating residual blocks into YOLO-Lite. Using the Jetson AGX Xavier device, 43 FPS can be achieved for recognizing a 224 × 224 video image. Despite this, the number of parameters was still large at 20.5 MB. Tang et al. (2020) introduced a lightweight grape disease classification model using ShuffleNet as the backbone of the model and optimized it with the channel attention mechanism. The model achieved a recognition accuracy of 99.14% with only 1.1 M parameters. Yun et al. (2022) constructed a depthwise separable residual module by combining depthwise separable convolution and non-bottleneck residual modules. They replaced the VGG backbone network in the SSD network with this module and a depthwise separable convolutional structure to reduce the parameter count of the object detection model and enhance detection speed. Dong et al. (2024) improved the YOLOv5 algorithm by introducing Involution modules, Multi-Level Spatial Pyramid Pooling, Efficient Channel Attention (ECA) mechanism, and Content-Aware ReAssembly of Features upsampling method. As a result, the model's parameter count decreased from 7.03 million to 6.09 million without sacrificing performance. Gao et al. (2021) designed a lightweight network backbone based on a novel spatial attention mechanism and migration learning to categorize garbage images and achieved good results in the garbage categorization task on the Huawei cloud platform with an accuracy rate of 96.17% and several floating-point operations of approximately 450M FLOPs. Dog-qiuqiu (2021) introduced the YOLO-FastestV2 model for real-time detection on mobile devices. The model aimed to substitute the YOLOV5 backbone with ShuffleNetV2 while mitigating the Feature Pyramid Network (LightFPN) structure. The parameter size of the model was only 237.55 kB. Good real-time detection can be achieved, even on embedded devices with limited computational resources. YOLO-FastestV2 has the advantages of having a small number of network parameters and low equipment requirements. Therefore, we selected YOLO-FastestV2 as the primary network for wheat spike detection.

When faced with the challenges of high density, severe occlusion, and overlap in wheat spike image detection, traditional detection methods are often difficult to accurately recognize, resulting in significant errors and missed detection problems. To solve these problems, an improved YOLO-FastestV2 target detection method was proposed in this study. In this study, three attention mechanism modules, Large Separable Kernel Attention (LSKA), Efficient Channel Attention (ECA), and Efficient Multi-Scale Attention (EMA), are introduced to enhance the feature extraction capability of the backbone network, effectively filter out irrelevant information, and improve the accuracy of wheat spike detection. For the LightFPN module of the detection head in YOLO-FastestV2, SimConv was used to replace the traditional Conv layer to further improve the accuracy of the model. Subsequently, a hybrid-enhanced wheat spike detection model was constructed by combining the improvements in the backbone network with the detection head. In addition, this article provides a comprehensive evaluation of the accuracy, size, and computational complexity of the model, with the aim of finding a model that strikes a balance between performance and resource usage.

2 Materials and methods

2.1 Dataset construction

The wheat spike detection dataset utilized in this study is sourced from a variety of datasets, including the Global Wheat Detection 2020 dataset available at Kaggle, the Global Wheat Detection 2021 dataset found at Global Wheat, and a custom dataset provided by PP Flying Paddle, accessible at Baidu AI Studio. The Global Wheat Detection dataset came from 11 countries/regions and covered 44 unique measurement sessions (David et al., 2021). A measurement session is a set of images acquired using a specific sensor at the same location within a coherent timestamp (typically a few hours). This study reorganized three datasets, deleted blank and duplicate images, and retained 11,451 images. We reorganized all wheat spike categories and named all categories "wheat-spike". We retained only one unique category label, wheat spikes, to enhance the generalization performance of the training model, capture the common specific points of each type of wheat spike, and accommodate the features of the different categories of wheat spikes. In this study, the dataset was divided into a training set, validation set, and test set in the ratio of 8:1:1. The number of images in the training set was 9160 and the number of detection frames was 397411. The number of images in the validation set was 1145 and the number of detection frames was 50137. The number of images in the test set was 1146 and the number of detection frames was 49426. We selected several pictures of wheat spikes of different maturities and under different light conditions, as shown in Figure 1.

2.2 YOLO-FastestV2 network architecture

YOLO-FastestV2 consists of three parts: input layer, backbone network, and detection header. The main structure of the system is shown in Figure 2. The input layer consisted of a 24 × 24 convolutional module and a pooling layer. The backbone network uses the ShuffleNetV2 (Chen et al., 2022) network as the model for backbone feature extraction, which reduces the memory access cost, improves speed, and reduces model weight. The detection header part is used by the LightFPN module and Yolov5 anchor-matching mechanism to predict the actual detection frame. The structure of the LightFPN module is shown in Figure 3. Using 1 × 1 convolution and Depth Separable Convolution (DWConv) techniques, the LightFPN effectively reduces the computation and number of parameters of the

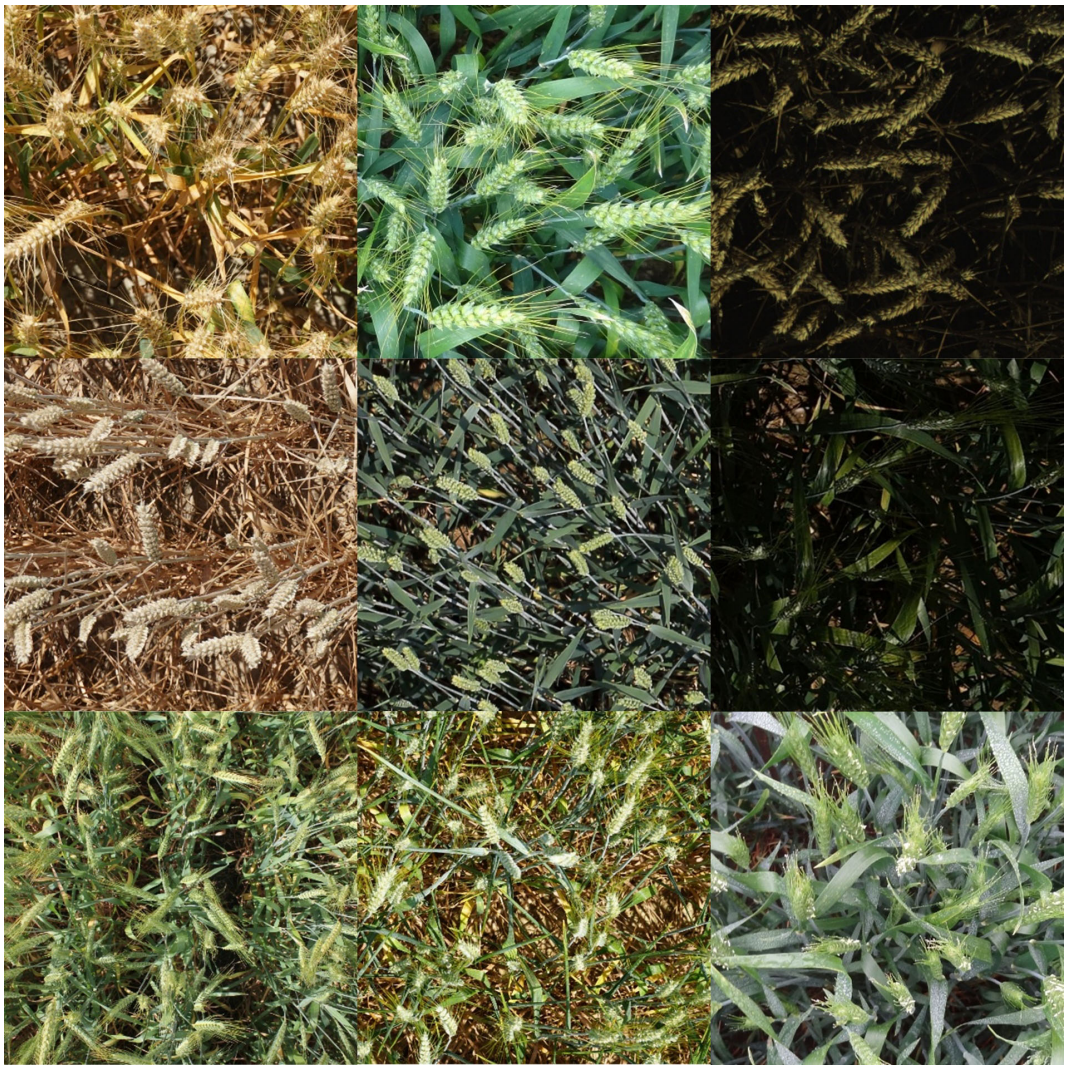


FIGURE 1
Example of the wheat spike dataset.

model, making it more efficient to run on resource-constrained devices. LightFPN integrates the feature maps of different layers through up-sampling and down-sampling operations to capture the information of the image at different scales. LightFPN uses up-sampling and down-sampling techniques to effectively fuse feature maps of different resolutions to capture image information at different scales. Furthermore, using DWConv modules and hopping connections, LightFPN significantly improves the representation of the features. The feature map of Yolo-Fastest V2 can be decoupled into three different feature maps. The foreground context classification and detection classes use the same network branches and share the parameters to enable lightweight target detection.

2.3 Attention mechanism module

The attention mechanism plays a key role in the field of deep learning by weighting the input feature maps, reinforcing important features, and suppressing unimportant ones, allowing the network to focus better on the target region and improve the representation of features (Qiu et al., 2020). In this study, we used three different attentional mechanisms (ECA, LSKA, and EMA) to improve the YOLO-FastestV2 model to improve the detection accuracy of the model.

2.3.1 ECA attention module

The ECA attention mechanism is an attention model designed to enhance the capacity of neural networks to model image features (Wang et al., 2020). At first, ECA abandons the dimensionality reduction operation in the SE module and maintains the direct correlation of the channels by maintaining the integrity of the information through 1×1 convolution. Secondly, ECA employs one-dimensional convolution to enhance local interactions between channels, which not only reduces information loss due to dimensionality reduction, but also allows the network to more fully utilize spatio-temporal information. In addition, the convolutional kernel size of ECA can be adaptively adjusted according to the number of channels and the depth of the

10.3389/fpls.2024.1411510

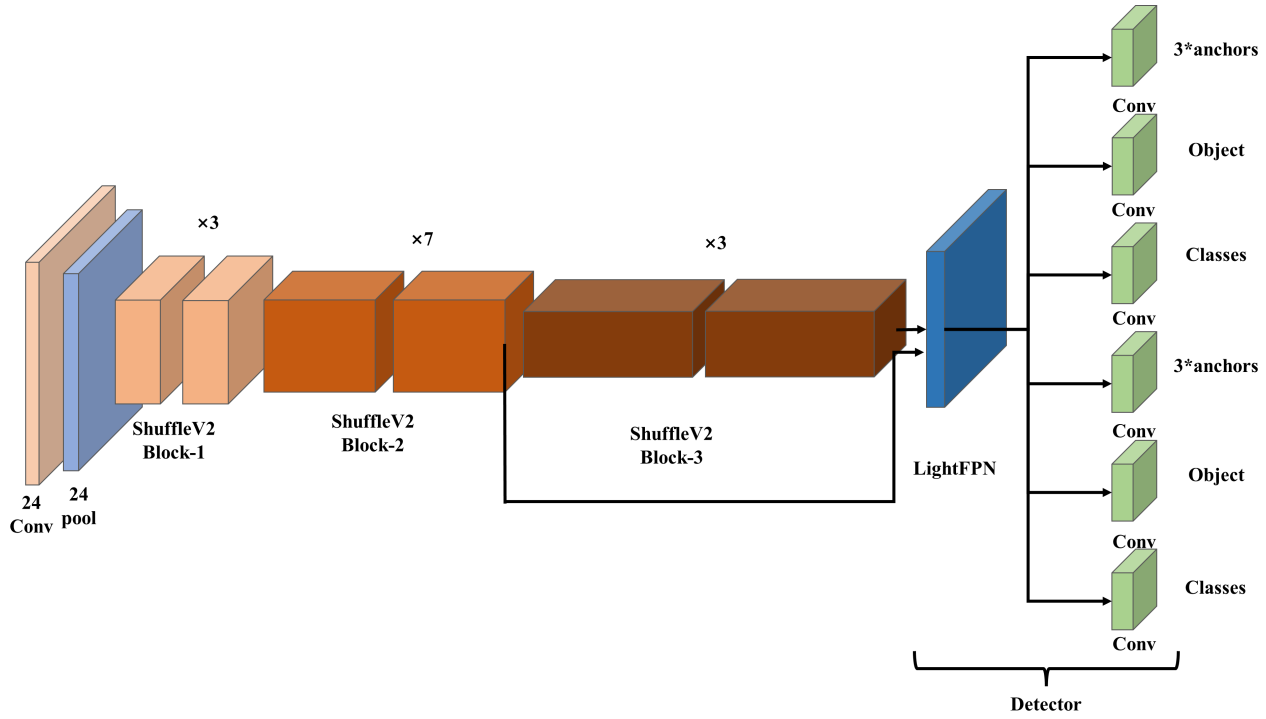


FIGURE 2
Diagram of YOLO-FastestV2 network structure.

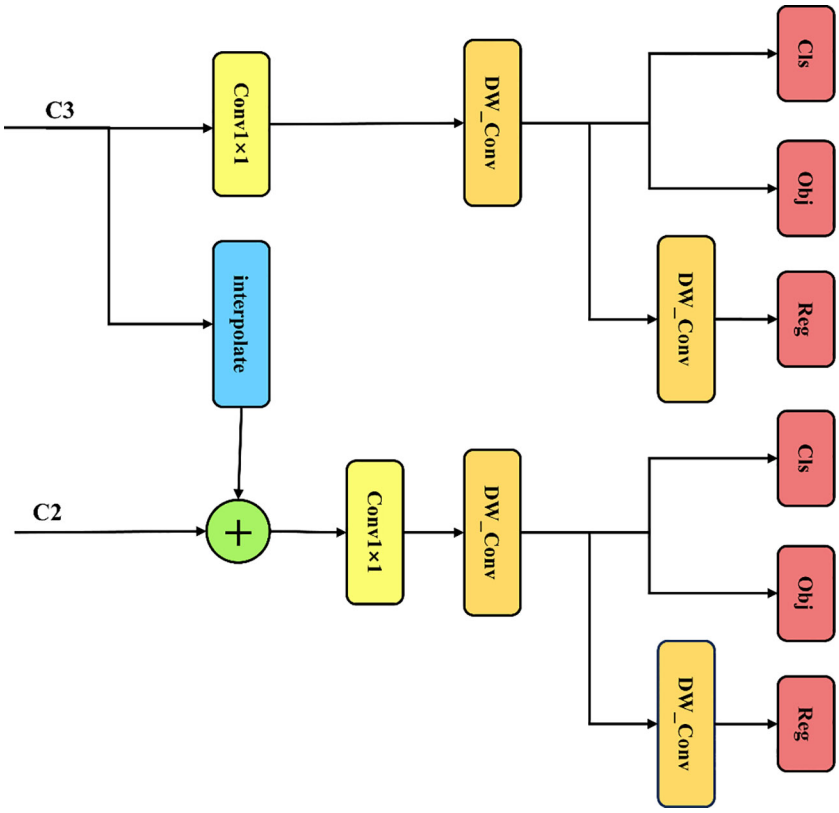


FIGURE 3
LightFPN module structure.

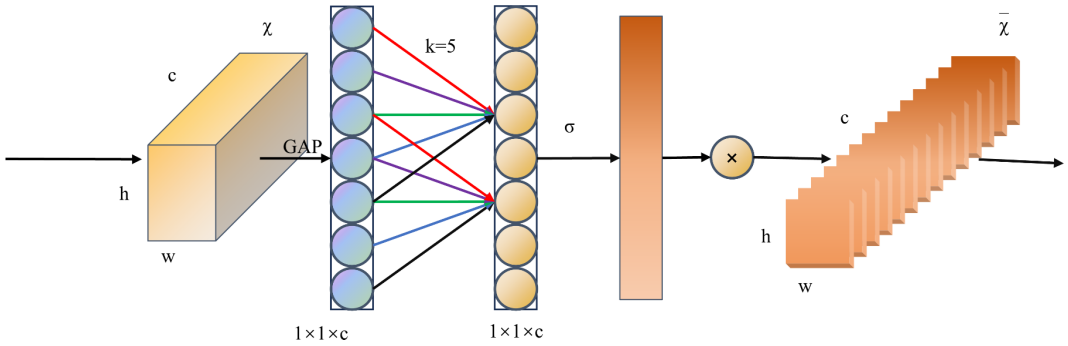


FIGURE 4
Structure of the ECA Attention Mechanism Module.

network, providing flexibility to capture different ranges of channel dependencies. These features work together to enhance the performance of the model. The main structure of the ECA attention mechanism is illustrated in Figure 4. ECA first performs global average pooling (GAP) on the acquired feature maps, followed by application of a one-dimensional convolution operation with kernel size k. In this process, the weights of each channel are computed using a sigmoid activation function, the expression of which is shown in Equation 1.

$$\omega = \sigma(C1D_k(y)) \tag{1}$$

where $1D_k$ denotes a one-dimensional convolution with a convolution kernel size k. Kernel size k determines the range of the channel interaction and is related to the channel dimension, which denotes the visible channel function. In general, it is believed that the larger the channel size, the stronger are the long-term interactions, whereas the smaller the channel size, the stronger are the short-term interactions. In ECA, an adaptive approach is used to determine the size of the nucleus. Finally, the weights are multiplied by the corresponding elements of the original input feature map to obtain the final output feature map.

2.3.2 LSKA attention mechanisms

The principle of the LSKA is an attention mechanism proposed in response to the problems faced by traditional Large Kernel Attention (LKA) when dealing with large convolutional kernels, such as high computational and memory requirements (Lau et al., 2024). The LSKA attention mechanism decomposes the two-dimensional convolutional kernels of the deep convolutional layer into cascading horizontal and vertical one-dimensional kernels, significantly reducing the number of parameters and computational complexity of the model. The LSKA maintains similar performance to the original LKA, effectively capturing key features of the image such as edges, texture and shape. In addition, the design of LSKA allows the use of large convolutional kernels directly in the attention module without the need for additional blocks, enhancing the model's ability to model long distance dependencies of the input image. While reducing computational and memory requirements, LSKA improves the performance of vision tasks and enhances the robustness of the model to image perturbations. The LSKA attention structure is shown in Figure 5. The core of LSKA involves decomposing a conventional 2D convolutional kernel into two 1D convolutional kernels. First, it decomposes a large 2D kernel, horizontally and vertically, into two 1D kernels. This decomposition drastically reduces the number of parameters and computational complexity. Despite the decomposition and concatenation strategies, the performance of LSKA is similar to that of the original LKA. This means that the LSKA is effective in capturing important information when dealing with the key features of an image, such as edges, textures, and shapes.

2.3.3 EMA attention mechanism

EMA is an efficient multi-scale attention module based on cross-spatial learning designed to enhance the performance of deep learning models in computer vision tasks (Ouyang et al., 2023). The EMA attention mechanism captures multi-scale features through parallel 1x1 and 3x3 convolutional branches, where the 1x1 branch encodes channel attention using one-dimensional global average pooling, while the 3x3 branch is used to capture more detailed multi-scale spatial features. To enhance the feature representation through cross-spatial information aggregation

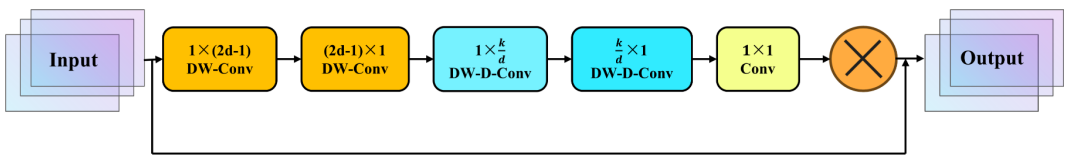


FIGURE 5
Structure of the LSKA attention mechanism.

while avoiding information loss through dimensionality reduction operations, EMA maintains direct correlation between channels. In addition, EMA reduces computational overhead by reshaping some of the channels into bulk dimensions, while enhancing the capacity of the model to perceive the global context of the image by encoding global spatial information. The main structure is illustrated in Figure 6. The EMA module divides the input feature map into subfunctions for each channel dimension. The EMA module uses three parallel routes to extract attention weight descriptors: two in the 1 × 1 branch and one in the 3 × 3 branch. The 1 × 1 branch uses one-dimensional global average pooling to encode channel attention in both spatial directions. The 3 × 3 branch uses a single 3 × 3 kernel to capture the multi-scale feature representations. The EMA module also passes the original input elements through a 3 × 3 branch to expand the feature space. The EMA module uses cross-spatial information aggregation to enrich the feature aggregation. The EMA module encodes global spatial information using two-dimensional global average pooling in 1 × 1 and 3 × 3 branches. The EMA module encodes global spatial information by applying two-dimensional global average pooling in 1 × 1 and 3 × 3 branches, generates spatial attention maps using matrix dot product operations, and then aggregates the output feature maps within each group to compute spatial attention weights.

2.3.4 SimConv

SimConv, an innovative module within the architecture of Convolutional Neural Networks (CNNs), leverages the concept of similarity among image blocks to steer the convolution process. This strategic approach is designed to elevate the caliber of feature representation while concurrently boosting the network's operational efficiency. The module's underlying philosophy hinges on the discernment of representative and uncertain areas within the input feature map, followed by the application of a tailored processing strategy. This not only preserves the integrity of the feature set but also leads to a pronounced reduction in computational demands and a streamlined parameter count. The computational procedure of SimConv can be succinctly outlined as follows:

$$y = \mathrm{Re}\,LU(BN(x \cdot \omega)) \tag{2}$$

In Equation 2, x denotes the input feature map, *www* symbolizes the convolution kernel, BN refers to Batch Normalization, and ReLU is indicative of the Rectified Linear Unit activation function. The forward_fuse technique optimizes the process by integrating Batch Normalization and the ReLU activation function directly into the convolution operation, thus achieving a decrease in both computational overhead and memory usage:

$$y = \mathrm{Re}\,LU(x \cdot \omega) \tag{3}$$

2.4 Evaluation of indicators

In this study, Precision (P), Recall (R), Average Precision (AP), and Mean Average Precision (mAP) were used as evaluation metrics to compare the effectiveness of different wheat spike detection models. where P is the probability of correctly detecting the target among all the detected targets, R is the probability of correctly recognizing the target among all the positive samples, and AP is the average value of the detector for each R case, which corresponds to the area under the PR curve. The mAP is the average of the averages computed for the APs of various categories, which summarizes the APs in terms of the category dimension to measure the performance of the multi-category target detection task. They are formulated as shown in Equations 3–7.

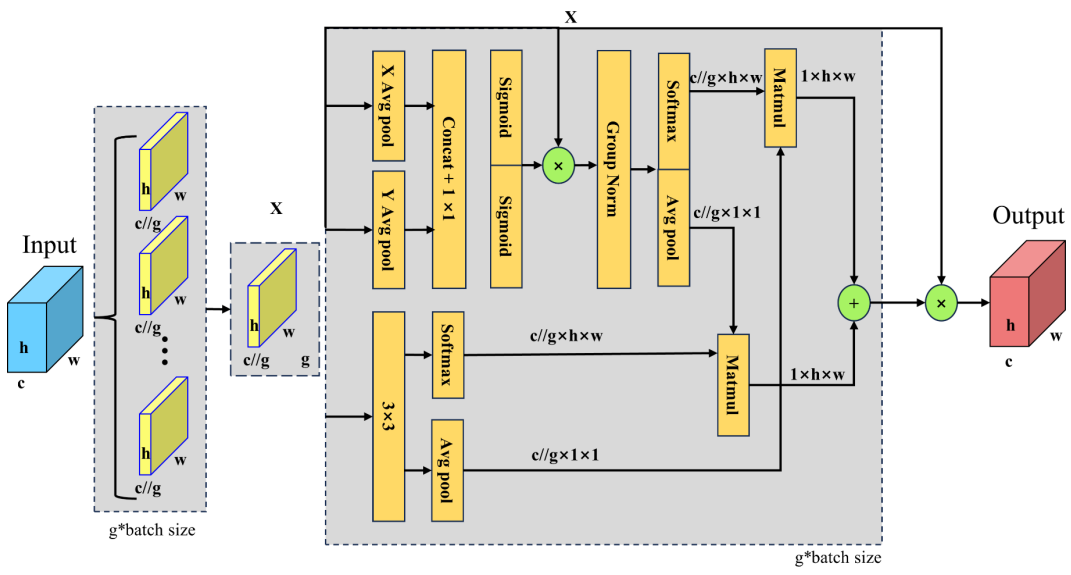


FIGURE 6
Structure of EMA attention mechanism.

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN} \tag{5}$$

$$AP = \frac{\sum P_{r_i}}{\sum r} \tag{6}$$

$$mAP = \frac{AP}{num_classes} \tag{7}$$

Where TP denotes that the predicted value is a positive sample as well as the true value. FP denotes that the predicted value is a positive sample and the true value is a negative sample. FN denotes that the predicted value is a negative sample, the true value is a positive sample, and num_classes denotes the number of classes. The P_{r_i} denotes the P-value corresponding to r - i on the PR curve, $\sum r = 1$.

In this study, a suite of evaluation metrics was used to assess the accuracy of the model. Although these enhancement strategies have generally improved the precision of the model, they have also led to an increase in Giga Floating Point Operations per Second (GFLOPs) and Params, which may elevate the hardware requirements for the model. To comprehensively assess the performance of the improved model, we introduced the global evaluation index (GPI) as a comprehensive evaluation index, as shown in Equation 8.

$$GPI = \sum_{j=1}^{4} \alpha_j (g_j - y_j) \tag{8}$$

where g_j are MAP@0.5, MAP@0.5:0.95, GFLOPs, and Params, y_j are the medians of the corresponding parameters. When g_j is MAP@0.5, α_j takes the value 0.7 and when g_j is MAP@0.5:0.95, α_j is 0.2; otherwise α_j is -0.05. MAP@0.5 is the result of calculating the AP of all images in each category when IoU is set to 0.5, and then averaged across all categories. mAP@.5:.95 indicates mean mAP at different IoU thresholds (from 0.5 to 0.95 in steps of 0.05) (0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, and 0.95). At the same time, because the network we selected was lightweight, we gave GFLOPs and Params a smaller share in the overall evaluation. The higher the GPI value or the higher the ranking, the better is the performance of the model in terms of accuracy and the combined requirements of the device.

3 Results and discussion

3.1 Experimental platforms

In this study, the experimental platform we used was a desktop computer configured with a 10th generation Intel Core i5 processor with 16 GB of memory (RAM) and an NVIDIA GeForce RTX 2060 graphics card with 6 GB of video memory. The operating system was Windows 10 Professional and CUDA version 11.8. For programming, we used two tools, Python 3.10 and PyTorch 2.1.1.

3.2 Impact of inclusion of a single-stage attention mechanism on the model

In the YOLO-FastestV2 model, the backbone network uses ShuffleNetV2 architecture, which is divided into three main phases: foundation, transition, and output (Zhang et al., 2023). The outputs of the transition and output phases were fed into the subsequent LightFPN modules. Considering that different attention mechanisms have different focuses on feature extraction and selection, this study introduced an attention mechanism module after the base and output phases to enhance the feature extraction capability of the backbone network. We denoted the attention mechanism added after the foundation stage as 1 and the attention mechanism added after the output stage as 2. The modeling accuracy of the Attention Mechanism module, which was added separately after the different stages, is presented in Figures 7 and 8.

As shown in Figure 7, the accuracy of the YOLO-FastestV2 model is P=80.81%, R=75.34%, AP=77.70%, and F1 = 77.98%. Among the models with the three attention modules added in the base stage, YOLO-FastestV2-ECA1, with the addition of the ECA attention mechanism, exhibited the highest accuracy improvement with models with P=82.64%, R=77.29%, AP=79.96%, and F1 = 79.88%. YOLO-FastestV2-LSKA1 with the addition of the LSKA attention mechanism exhibited a higher accuracy improvement (P=81.83%, R=76.22%, AP=79.31%, and F1 = 78.93%). The Yolo fastestV2_LSKA1 model showed higher accuracy gains (P=81.83%, R=76.22%, AP=78.31%, and F1 = 78.93%). The YOLO-FastestV2-EMA1 model, with the addition of attention to the EMA attentional mechanism, had smaller accuracy gains (P=81.67%, R=76.03%, AP= 78.76%, and F1 = 78.75%). As shown in Figure 8, among the models with the three attention modules added after the output stage, the YOLO-FastestV2-LSKA2 model with the addition of the

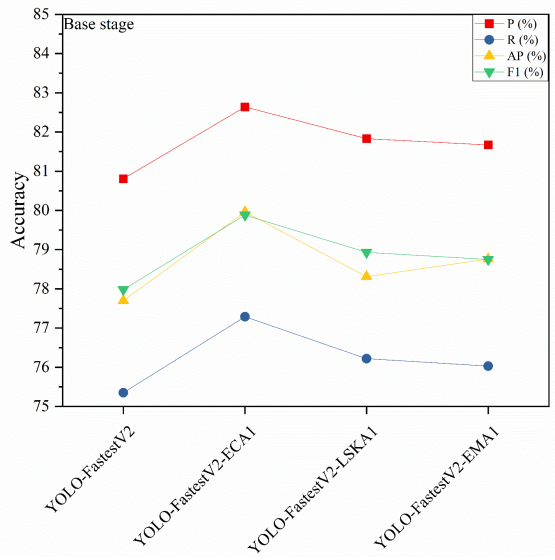


FIGURE 7
Model accuracy added by base stage attention mechanism.

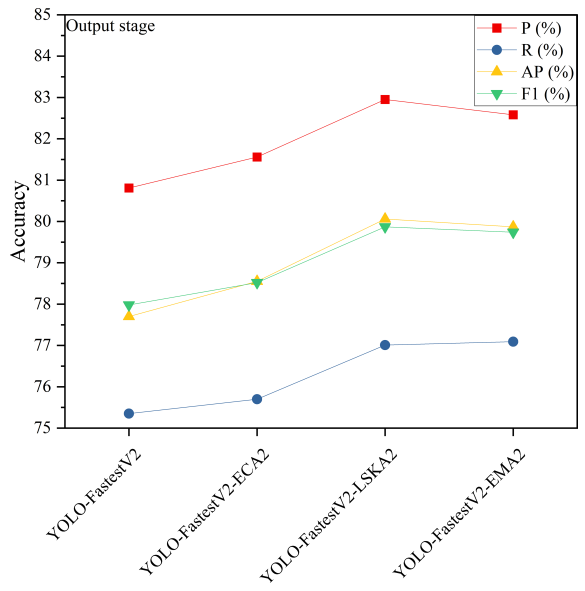


FIGURE 8
Model accuracy added by output stage attention mechanism.

LSKA attention mechanism module achieved the highest accuracy with a model accuracy of P=82.95%, R=77.01%, AP=80.06%, and F1 = 79.87%. This was followed by the YOLO-FastestV2-EMA2 model, which exhibited a greater increase in accuracy with the addition of the EMA attention mechanism module (P=82.58%, R=77.09%, AP=79.87%, and 79.74%). Lastly, the YOLO-FastestV2-ECA2 model, with the addition of the ECA attention mechanism module, exhibited the lowest accuracy improvement with P=81.56%, R=75.70%, AP=78.55%, and F1 = 78.52%.

The ECA attention mechanism improves the performance of the network by adaptively selecting a one-dimensional convolutional kernel size to adjust the coverage of local cross-channel interactions to avoid losing the target feature information during the dimensionality reduction process (Guo et al., 2021). The LSKA attention mechanism combines local self-attention with a large convolutional kernel to enable the model to focus on specific regions in the image, thereby improving its ability to recognize key features and local structures in the image (Lau et al., 2024). The EMA attention mechanism enhances the feature recognition capability of the model by reshaping some of the channel dimensions into batch dimensions and constructing local cross-channel interactions in each parallel sub-network, as well as increasing the interaction of the channel information by fusing the output feature maps of the two parallel sub-networks through a cross-space learning approach (Sun et al., 2024). The model with the ECA attention mechanism added at the base stage (YOLO-FastestV2-ECA1) showed the highest accuracy improvement, suggesting that ECA is effective in enhancing the model's recognition of the target features in the image, especially in the early stages of the model. LSKA showed a high accuracy improvement both in the base stage (YOLO-FastestV2-LSKA1) and after the output stage (YOLO-FastestV2-LSKA2). This indicated that LSKA is effective in enhancing the performance of the model at different stages, particularly in recognizing local features in the image. The model with the EMA attention mechanism added after the output stage (YOLO-FastestV2-EMA2) exhibited the higher accuracy, whereas the model with the EMA added at the base stage (YOLO-FastestV2-EMA1) exhibited a smaller accuracy gain. This suggests that EMA is more effective in the later stages of the model because it may require more contextual information to optimize feature fusion. Although the three attentional mechanisms improved the accuracy of the model to different degrees, the same attentional mechanism had different effects on the accuracy of the model at different stages of the network.

3.3 Impact of inclusion of two-stage attention mechanism on modeling

To further enhance the extraction performance of target features of the backbone network, the attention mechanisms that have higher performance enhancement of the model at different stages were selected in this study to be used in the construction of the backbone network with hybrid attention mechanisms to enhance the performance of the model. Among the results of the model enhancement by the single-stage attention mechanism modules, ECA and LSKA added after the base stage improved the accuracy of the model, and the addition of LSKA and EMA attention mechanism modules after the output stage improved the accuracy of the model. The accuracy results of the two-stage addition of the attention mechanism model are listed in Figure 9.

As can be observed from Figures 7, 8, the experimental results of combining the attention mechanisms showed that the combination of different attention mechanisms improved the accuracy of the model. From Figure 9, it can be observed that among the four models with combined attention mechanisms, the YOLO-FastestV2-ECA-EMA model with the combination of ECA and EMA attention mechanisms had the highest overall accuracy improvement (P=83.84%, R=78.11%, AP=81.42%, and F1 = 80.88%), and the LSKA and EMA attention mechanisms with the combination of YOLO-FastestV2-LSKA-EMA model had a higher accuracy improvement with a model accuracy of P=84.01%, R=77.94%, AP=81.21%, and F1 = 80.87%. Third, we used the YOLO-FastestV2-LSKA-LSKA model with the addition of the LSKA attentional mechanism in both the stages (P=83.34%, R=77.65%, AP=80.61%, and F1 = 80.39%). Finally, the YOLO-FastestV2-ECA-LSKA model with a combination of the two attentional mechanisms, ECA and LSKA, exhibited the lowest accuracy improvements of P=83.31%, R=77.40%, AP=80.53%, and F1 = 80.24%.

The YOLO-FastestV2-ECA-EMA model performed the best in terms of overall accuracy, which may be because ECA optimized the feature representation in the early stage, whereas EMA further refined the multi-scale information of the features in the later stage. The combination of both may have formed a complementary approach, which makes the model effective in extracting and utilizing features in different stages. The higher accuracy improvement of the YOLO-FastestV2-LSKA-EMA model may be because LSKA performed effective local and global structure

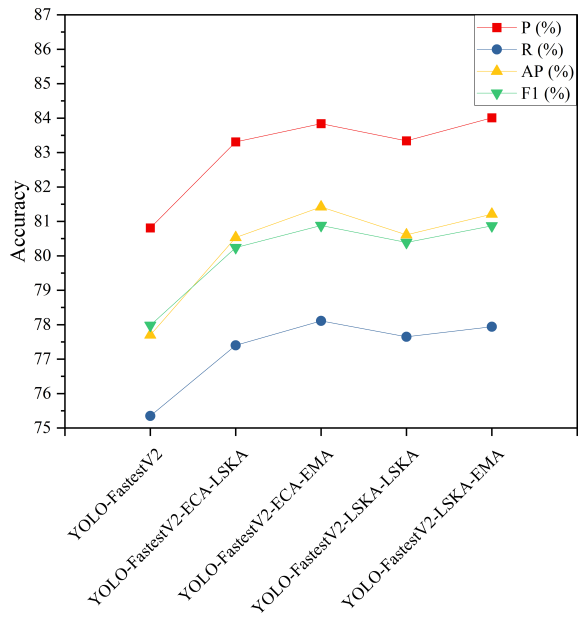


FIGURE 9
Modeling accuracy of two-stage increasing attention mechanism.

extraction of the features in the early stages, whereas EMA further refined these features in the later stages, and the combination of the two may similarly form a complementary effect. The YOLO-FastestV2-LSKA-LSKA model ranked third in terms of improvement in accuracy. This suggests that using LSKA in both phases may offer an advantage in the depth of feature extraction, but may not have the benefit of ECA or EMA in multi-scale feature capture. The YOLO-FastestV2-ECA-LSKA model showed the smallest improvement in accuracy, possibly owing to some overlap between the mechanisms. This overlap may cause the network focus to overly prioritize localized information, leading to limited improvement in model accuracy.

3.4 Impact of improved LightFPN on modeling

In the YOLO-FastestV2 algorithm, the LightFPN module was used in the detection head section to process the output results of the backbone network. Despite its lightweight design, the LightFPN encounters challenges in wheat spike detection, particularly in scenarios with uniform backgrounds and complex scene dynamics. To enhance detection accuracy in these conditions, we integrated the SimConv module to replace the traditional Conv module within the LightFPN architecture, resulting in the SimLightFPN. The SimConv module primarily consists of Conv, BatchNorm, and ReLU to accelerate the convergence of the model, which allows the direct application of convolution and activation functions without batch normalization by providing the forward_fuse method (Hu and Zhu, 2023). The accuracy of the improved model is shown in Table 1.

As can be observed from Table 1, the accuracy of the improved model was P=83.65%, R=77.54%, AP=80.51%, and F1 = 80.48%, which is an increase of 2.84% for P, 2.19% for R, 2.81% for AP, and 2.5% for F1, compared with the original model. SimLightFPN, while retaining the core advantages of LightFPN, optimized the forward propagation process of the feature fusion network through the introduction of SimConv, thereby significantly enhancing the performance of the algorithm in handling the feature fusion task (Yue et al., 2023). The results suggest that SimConv may be more efficient for feature extraction and fusion, particularly when dealing with similar backgrounds and complex scenes in whelk detection.

3.5 Multi-stage improved hybrid models

To further enhance the performance of the model, two models, YOLO-FastestV2-ECA-EMA and YOLO-FastestV2-LSKA-EMA, which have higher accuracy for the two-stage addition of the attention mechanism, were selected in this study and combined with the improved SimLightFPN module to construct a multi-stage improved hybrid model. The accuracies of the improved hybrid models are listed in Table 2.

As can be observed in Table 2, in the multi-stage improved hybrid model, the YOLO-FastestV2-SimLightFPN-LSKA-EMA model showed the highest accuracy improvement (P=84.28%, R=78.37%, AP=81.85%, and F1 = 81.22%) and the YOLO-FastestV2-SimLightFPN- ECA-EMA showed a higher accuracy improvement with models with P=83.91%, R=78.35%, AP=81.52%, and F1 = 81.03%. The YOLO-FastestV2-SimLightFPN-LSKA-EMA model performed the best in terms of accuracy improvement, which may be attributed to the advantages of the combination of the LSKA and EMA attention mechanisms in terms of an in-depth understanding of features and multi-scale information extraction. Combined with the feature fusion capability of SimLightFPN, this enabled the model to capture and analyze features at different depths better in the wheat spike detection task.

TABLE 1 Improved SimLightFPN model with original model accuracy.

Model	P (%)	R (%)	AP (%)	F1 (%)
YOLO-FastestV2	80.81	75.35	77.70	77.98
YOLO-FastestV2-SimLightFPN	83.65	77.54	80.51	80.48

TABLE 2 Accuracy of the multi-stage improved hybrid model.

Model	P (%)	R (%)	AP (%)	F1 (%)
YOLO-FastestV2	80.81	75.35	77.70	77.98
YOLO-FastestV2-SimLightFPN-LSKA-EMA	84.28	78.37	81.85	81.22
YOLO-FastestV2-SimLightFPN-ECA-EMA	83.91	78.35	81.52	81.03

Although the accuracy improvement of the YOLO-FastestV2-SimLightFPN-ECA-EMA model was slightly lower than that of the YOLO-FastestV2-SimLightFPN-LSKA-EMA model, it still exhibited high-performance improvement. The results show that the combination of SimLightFPN and the two-stage add-attention mechanism can effectively fuse and analyze the features of different depths to improve the accuracy of the model.

3.6 Comprehensive evaluation

The ranked results of the GPI for all models are listed in Table 3.

As can be observed from Table 3, among the various improved models, the YOLO-FastestV2-SimLightFPN-ECA-EMA model had the highest GPI of 0.84, the second-ranked model was the YOLO-FastestV2-SimLightFPN-ECA-EMA model with a GPI of 0.84, and the third-ranked model was the YOLO-FastestV2-ECA-EMA model with a GPI of 0.81. The highest accuracy of the YOLO-FastestV2-SimLightFPN-ECA-EMA model may be because it combines both the ECA and EMA attention mechanisms as well as the SimLightFPN module, which may have made the model more efficient in feature extraction and fusion. Although the GPI values of YOLO-FastestV2-SimLightFPN-LSKA-EMA and YOLO-FastestV2-SimLightFPN-ECA-EMA are close to each other, YOLO-FastestV2-SimLightFPN-LSKA-EMA performed better in other performance metrics (e.g., mAP@0.5, and mAP@.5:.95) is better, which suggests that it has an advantage in terms of accuracy. Compared with the YOLO-FastestV2-SimLightFPN-ECA-EMA model, the YOLO-FastestV2-SimLightFPN-LSKA-EMA model has a higher computational complexity and number of parameters, which may lead to limitations in its deployment performance on mobile platforms. The YOLO-FastestV2-ECA-EMA model had a slightly lower GPI than the other two models, but it also performed well in terms of accuracy. The most efficient model, YOLO-FastestV2-SimLightFPN-ECA-EMA, achieved a high level of accuracy while maintaining low computational requirements.

4 Conclusion

Accurate detection and measurement of the number of wheat spikes are important for the rapid prognosis of wheat yield and national food security. This study is based on fusing images of wheat spikes from multiple datasets to construct a model for wheat spike detection that can be used for preliminary estimation of wheat yield on mobile devices. The accuracy and reliability of the model were assessed by analyzing the differences between the model detection results and actual labeling results. In this study, the YOLO-FastestV2 model was used as the basis for constructing a wheat spike detection model. To improve wheat spike detection the accuracy, three attention mechanisms (ECA, EMA, and LSKA) were introduced and combined with YOLO-FastestV2 to construct a new model. In this study, the attention mechanism was added after the base and output phases of the backbone network. The YOLO-FastestV2-ECA1 model with ECA attention added after the base phase demonstrated the

TABLE 3 Value of GPI for all models.

Model	mAP@0.5 (%)	mAP@.5:.95 (%)	GFLOPs	Params	GPI	Rank
YOLO-FastestV2	67.57	55.10	0.38	0.24	0.00	14
YOLO-FastestV2-ECA1	70.41	60.00	0.38	0.24	0.60	6
YOLO-FastestV2-ECA2	67.78	55.12	0.38	0.24	0.03	13
YOLO-FastestV2-LSKA1	67.97	57.03	0.4	0.24	0.09	12
YOLO-FastestV2-LSKA2	69.93	58.90	0.39	0.28	0.42	10
YOLO-FastestV2-EMA1	68.58	56.63	0.4	0.24	0.17	11
YOLO-FastestV2-EMA2	69.77	58.76	0.39	0.24	0.44	9
YOLO-FastestV2-ECA-LSKA	70.23	59.26	0.39	0.28	0.48	8
YOLO-FastestV2-ECA-EMA	71.58	61.41	0.39	0.24	0.81	3
YOLO-FastestV2-LSKA-LSKA	70.59	59.85	0.41	0.28	0.52	7
YOLO-FastestV2-LSKA-EMA	71.07	60.68	0.41	0.25	0.66	5
YOLO-FastestV2-SimLightFPN	70.75	60.66	0.38	0.24	0.67	4
YOLO-FastestV2-SimLightFPN-LSKA-EMA	72.00	61.71	0.41	0.25	0.84	2
YOLO-FastestV2-SimLightFPN-ECA-EMA	71.76	61.46	0.39	0.24	0.84	1

highest improvement in accuracy (P=82.64%, R=77.29%, AP=79.96%, and F1 = 79.88%). Conversely, the YOLO-FastestV2-LSKA2 model with the LSKA attention mechanism added after the output phase exhibited the highest accuracy gains (P=82.95%, R=77.01%, AP=80.06%, and F1 = 79.87%). To further enhance the extraction performance of target features in the backbone network, this study selected the ECA and LSKA attention mechanisms to improve the accuracy of the model after the base phase and the LSKA and EMA attention mechanisms that improved the accuracy of the model after the output phase to construct a model with a two-stage increase in the attention mechanism. The YOLO-FastestV2-ECA-EMA model demonstrated the highest accuracy (P=83.84%, R=78.11%, AP=81.42%, and F1 = 80.88%) among the two-stage models for increasing the attentional mechanisms. In this study, SimLightFPN was proposed by introducing SimConv to simplify and improve the feature fusion capability of the LightFPN module compared with the unimproved YOLO-FastestV2 model. This resulted in an increase of 2.84% in P, 2.19 in R, 2.81% in AP, and 2.5% in F1, compared with the original model. We then constructed a hybrid wheat spike detection model by combining a two-stage model with simultaneous addition of attention mechanisms and SimLightFPN. Among the hybrid models, the YOLO-FastestV2-SimLightFPN-LSKA-EMA model exhibited the highest accuracy (P=84.28%, R=78.37%, AP=81.85%, and F1 = 81.22%). In the comprehensive evaluation of the models, the GPI value of YOLO-FastestV2-SimLightFPN-ECA-EMA was ranked No. 1 (0.84), which is the optimal model under the combined consideration of model accuracy and model complexity. Currently, we have only used one lightweight machine-learning model for improvement and evaluation, and we will explore more machine-learning models in our future research and combine different network structures and techniques to achieve a wider range of applications on diverse datasets.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SQ: Conceptualization, Methodology, Software, Writing – original draft. ZQ: Data curation, Formal analysis, Supervision, Writing – review & editing. WW: Software, Visualization, Writing – review & editing. FW: Formal analysis, Writing – review & editing. XJ: Funding acquisition, Investigation, Writing – review & editing. JJ: Funding acquisition, Visualization, Writing – review & editing. LZ: Funding acquisition, Software, Writing – review & editing. YS: Funding acquisition, Investigation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was also supported by National Key Research and Development China Project (No. 2022YFD2001205), Major Science and Technology Special Project in Henan Province (No.231100110200), Key Scientific Research Projects of Colleges and Universities in Henan Province (No. 24B416001), and Key Scientific Research Projects of Colleges and Universities in Henan Province (No.22B416002).

Acknowledgments

We thank Global Wheat Detection and PP Flying Paddle for providing publicly available datasets for wheat spike detection for this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aich, S., Josuttes, A., Ovsyannikov, I., Strueby, K., Ahmed, I., Duddu, H. S., et al. (2018). "Deepwheat: Estimating phenotypic traits from crop images with deep learning," in *2018 IEEE Winter conference on applications of computer vision (WACV)*. 323–332 (Lake Tahoe, NV, USA: IEEE). doi: 10.1109/WACV.2018.00042

Alkhudaydi, T., and De La Iglesia, B. (2022). Counting spikelets from infield wheat crop images using fully convolutional networks. *Neural Computing Appl.* 34, 17539–17560. doi: 10.1007/s00521-022-07392-1

Chen, Z., Yang, J., Chen, L., and Jiao, H. (2022). Garbage classification system based on improved ShuffleNet v2. *Resources Conserv. Recycling* 178, 106090. doi: 10.1016/j.resconrec.2021.106090

David, E., Serouart, M., Smith, D., Madec, S., Velumani, K., Liu, S., et al. (2021). Global wheat head detection 2021: An improved dataset for benchmarking wheat head detection methods. *Plant Phenomics*. doi: 10.34133/2021/9846158

Dog-qiuqiu (2021). *FastestDet: it has higher accuracy and faster speed than Yolo-fastest.*

Dong, Q., Sun, L., Han, T., Cai, M., and Gao, C. (2024). PestLite: A novel YOLO-based deep learning technique for crop pest detection. *Agriculture* 14, 228. doi: 10.3390/agriculture14020228

Ferrante, A., Cartelle, J., Savin, R., and Slafer, G. A. (2017). Yield determination, interplay between major components and yield stability in a traditional and a

contemporary wheat across a wide range of environments. *Field Crops Res.* 203, 114–127. doi: 10.1016/j.fcr.2016.12.028

Ganeva, D., Roumenina, E., Dimitrov, P., Gikov, A., Jelev, G., Dragov, R., et al. (2022). Phenotypic traits estimation and preliminary yield assessment in different phenophases of wheat breeding experiment based on UAV multispectral images. *Remote Sens.* 14, 1019. doi: 10.3390/rs14041019

Gao, M., Chen, Y., Zhang, Z., Feng, Y., and Fan, W. (2021). Classification algorithm of garbage images based on novel spatial attention mechanism and transfer learning. *Syst. Eng. Theory Pract.* 41, 498–512. doi: 10.12011/SETP2020-1645

Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. (2021). Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*. doi: 10.48550/arXiv.2107.08430

Grillo, O., Blangiforti, S., and Venora, G. (2017). Wheat landraces identification through glumes image analysis. *Comput. Electron. Agric.* 141, 223–231. doi: 10.1016/j.compag.2017.07.024

Guo, Z., Zhao, L., Yuan, J., and Yu, H. (2021). Msanet: multiscale aggregation network integrating spatial and channel information for lung nodule detection. *IEEE J. Biomed. Health Inf.* 26, 2547–2558. doi: 10.1109/JBHI.2021.3131671

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*. 2961–2969.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1904–1916. doi: 10.1109/TPAMI.2015.2389824

Hu, H., and Zhu, Z. (2023). Sim-YOLOv5s: A method for detecting defects on the end face of lithium battery steel shells. *Advanced Eng. Inf.* 55, 101824. doi: 10.1016/j.aei.2022.101824

Jin, X., Liu, S., Baret, F., Hemerlé, M., and Comar, A. (2017). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sens. Environ.* 198, 105–114. doi: 10.1016/j.rse.2017.06.007

Kamilaris, A., and Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Comput. Electron. Agric.* 147, 70–90. doi: 10.1016/j.compag.2018.02.016

Lau, K. W., Po, L.-M., and Rehman, Y. A. U. (2024). Large separable kernel attention: rethinking the large kernel attention design in CNN. *Expert Syst. Appl.* 236, 121352. doi: 10.1016/j.eswa.2023.121352

Li, J., Li, C., Fei, S., Ma, C., Chen, W., Ding, F., et al. (2021). Wheat ear recognition based on RetinaNet and transfer learning. *Sensors* 21, 4845. doi: 10.3390/s21144845

Li, L., Hassan, M. A., Yang, S., Jing, F., Yang, M., Rasheed, A., et al. (2022). Development of image-based wheat spike counter through a Faster R-CNN algorithm and application for genetic studies. *Crop J.* 10, 1303–1311. doi: 10.1016/j.cj.2022.07.007

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). "Ssd: Single shot multibox detector," *Computer VisionECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, October 1114, 2016, Proceedings, Part I 14. Springer International Publishing, 2016.

Ma, T., Yang, Z., Song, Y.-F., Liang, J., and Wang, H. (2023). DMEF-net: lightweight infrared dim small target detection network for limited samples. *IEEE Trans. Geosci. Remote Sens.* 61, 1–15. doi: 10.1109/TGRS.2023.3333378

Misra, T., Arora, A., Marwaha, S., Jha, R. R., Ray, M., Jain, R., et al. (2021). Web-SpikeSegNet: deep learning framework for recognition and counting of spikes from visual images of wheat plants. *IEEE Access* 9, 76235–76247. doi: 10.1109/ACCESS.2021.3080836

Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., et al. (2023). "Efficient multi-scale attention module with cross-spatial learning," in *ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece. 1–5. doi: 10.1109/ICASSP49357.2023.10096516

Pask, A. (2012). *Physiological Breeding: A field guide to wheat phenotyping. II. Cimmyt*.

Qiu, Z., Zhao, N., Zhou, L., Wang, M., Yang, L., Fang, H., et al. (2020). Vision-based moving obstacle detection and tracking in paddy field using improved yolov3 and deep SORT. *Sensors* 20, 4082. doi: 10.3390/s20154082

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA. 779–788. doi: 10.1109/CVPR.2016.91

Ren, S., He, K., Girshick, R., and Sun, J. (2017). "Faster R-CNN: Towards real-time object detection with region proposal networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., and LeCun, Y. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*. doi: 10.48550/arXiv.1312.6229

Slafer, G. A., Savin, R., and Sadras, V. O. (2014). Coarse and fine regulation of wheat yield components in response to genotype and environment. *Field Crops Res.* 157, 71–83. doi: 10.1016/j.fcr.2013.12.004

Sun, L., Dai, W., and Muhammad, G. (2024). Multi-level Graph Memory Network Cluster Convolutional Recurrent Network for traffic forecasting. *Inf. Fusion* 105, 102214. doi: 10.1016/j.inffus.2023.102214

Tan, M., Pang, R., and Le, Q. V. (2020). "Efficientdet: Scalable and efficient object detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 10781–10790. doi: 10.1109/CVPR42600.2020.01079

Tang, Z., Yang, J., Li, Z., and Qi, F. (2020). Grape disease image classification based on lightweight convolution neural networks and channelwise attention. *Comput. Electron. Agric.* 178, 105735. doi: 10.1016/j.compag.2020.105735

Wang, J., Cui, Z., Jiang, T., Cao, C., and Cao, Z. (2022). Lightweight deep neural networks for ship target detection in SAR imagery. *IEEE Trans. Image Process.* 32, 565–579. doi: 10.1109/TIP.2022.3231126

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. (2020). "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 11534–11542. doi: 10.1109/CVPR42600.2020.01155

Wang, Y., Qin, Y., and Cui, J. (2021). Occlusion robust wheat ear counting algorithm based on deep learning. *Front. Plant Sci.* 12, 645899. doi: 10.3389/fpls.2021.645899

Wen, C., Wu, J., Chen, H., Su, H., Chen, X., Li, Z., et al. (2022). Wheat spike detection and counting in the field based on SpikeRetinaNet. *Front. Plant Sci.* 13, 821717. doi: 10.3389/fpls.2022.821717

Wosner, O., Farjon, G., and Bar-Hillel, A. (2021). Object detection in agricultural contexts: A multiple resolution benchmark and comparison to human. *Comput. Electron. Agric.* 189, 106404. doi: 10.1016/j.compag.2021.106404

Xiong, H., Cao, Z., Lu, H., Madec, S., Liu, L., and Shen, C. (2019). TasselNetv2: in-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods* 15, 1–14. doi: 10.1186/s13007-019-0537-2

Xu, X., Li, H., Yin, F., Xi, L., Qiao, H., Ma, Z., et al. (2020). Wheat ear counting using K-means clustering segmentation and convolutional neural network. *Plant Methods* 16, 1–13. doi: 10.1186/s13007-020-00648-8

Ye, J., Yu, Z., Wang, Y., Lu, D., and Zhou, H. (2023). WheatLFANet: in-field detection and counting of wheat heads with high-real-time global regression network. *Plant Methods* 19, 103. doi: 10.1186/s13007-023-01079-x

Yue, X., Qi, K., Na, X., Zhang, Y., Liu, Y., and Liu, C. (2023). Improved YOLOv8-seg network for instance segmentation of healthy and diseased tomato plants in the growth stage. *Agriculture* 13, 1643. doi: 10.3390/agriculture13081643

Yun, J., Jiang, D., Liu, Y., Sun, Y., Tao, B., Kong, J., et al. (2022). Real-time target detection method based on lightweight convolutional neural network. *Front. Bioeng. Biotechnol.* 10, 861286. doi: 10.3389/fbioe.2022.861286

Zang, H., Wang, Y., Ru, L., Zhou, M., Chen, D., Zhao, Q., et al. (2022). Detection method of wheat spike improved YOLOv5s based on the attention mechanism. *Front. Plant Sci.* 13, 993244. doi: 10.3389/fpls.2022.993244

Zhang, J., Wang, J., and Zhang, S. (2021). An ultra-lightweight and ultra-fast abnormal target identification network for transmission line. *IEEE Sensors J.* 21, 23325–23334. doi: 10.1109/JSEN.2021.3110878

Zhang, L.-L., Jiang, Y., Sun, Y.-P., Zhang, Y., and Wang, Z. (2023). Improvements based on ShuffleNetV2 model for bird identification. *IEEE Access*. 11, 101823–101832. doi: 10.1109/ACCESS.2023.3314676

Zhang, Q., Liu, Y., Gong, C., Chen, Y., and Yu, H. (2020). Applications of deep learning for dense scenes analysis in agriculture: A review. *Sensors* 20, 1520. doi: 10.3390/s20051520

Zhao, F., Wang, K., and Yuan, Y. (2014). Study on wheat spike identification based on color features and adaboost algorithm. *Crops* 1, 1–5. doi: 10.3969/j.issn.1001-7283.2014.01.033

Zhao, J., Zhang, X., Yan, J., Qiu, X., Yao, X., Tian, Y., et al. (2021). A wheat spike detection method in UAV images based on improved YOLOv5. *Remote Sens.* 13, 3095. doi: 10.3390/rs13163095

Zhou, H., Riche, A. B., Hawkesford, M. J., Whalley, W. R., Atkinson, B. S., Sturrock, C. J., et al. (2021). Determination of wheat spike and spikelet architecture and grain traits using X-ray Computed Tomography imaging. *Plant Methods* 17, 1–9. doi: 10.1186/s13007-021-00726-5

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 02 July 2024
DOI 10.3389/fpls.2024.1425131

Check for updates

OPEN ACCESS

EDITED BY
Ting Sun,
China Jiliang University, China

REVIEWED BY
Li Jiang,
Jiangsu University, China
Xu Hao,
Shandong Academy of Agricultural Sciences, China

*CORRESPONDENCE
Changwei Tan
cwtan@yzu.edu.cn

RECEIVED 29 April 2024
ACCEPTED 18 June 2024
PUBLISHED 02 July 2024

CITATION
Hong Q, Liu W, Zhu Y, Ren T, Shi C, Lu Z, Yang Y, Deng R, Qian J and Tan C (2024) CTHNet: a network for wheat ear counting with local-global features fusion based on hybrid architecture. *Front. Plant Sci.* 15:1425131. doi: 10.3389/fpls.2024.1425131

COPYRIGHT
© 2024 Hong, Liu, Zhu, Ren, Shi, Lu, Yang, Deng, Qian and Tan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

CTHNet: a network for wheat ear counting with local-global features fusion based on hybrid architecture

Qingqing Hong[1,2], Wei Liu[1,2], Yue Zhu[1,2], Tianyu Ren[1,2], Changrong Shi[1,2], Zhixin Lu[1,2], Yunqin Yang[1,2], Ruiting Deng[1,2], Jing Qian[1,2] and Changwei Tan[1,2]*

[1]Jiangsu Key Laboratory of Crop Genetics and Physiology/Jiangsu Key Laboratory of Crop Cultivation and Physiology, Agricultural College of Yangzhou University, Yangzhou, China, [2]Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops/Joint International Research Laboratory of Agriculture and Agri-Product Safety of the Ministry of Education of China/ Jiangsu Province Engineering Research Center of Knowledge Management and Intelligent Service, College of Information Engineer, Yangzhou University, Yangzhou, China

Accurate wheat ear counting is one of the key indicators for wheat phenotyping. Convolutional neural network (CNN) algorithms for counting wheat have evolved into sophisticated tools, however because of the limitations of sensory fields, CNN is unable to simulate global context information, which has an impact on counting performance. In this study, we present a hybrid attention network (CTHNet) for wheat ear counting from RGB images that combines local features and global context information. On the one hand, to extract multi-scale local features, a convolutional neural network is built using the Cross Stage Partial framework. On the other hand, to acquire better global context information, tokenized image patches from convolutional neural network feature maps are encoded as input sequences using Pyramid Pooling Transformer. Then, the feature fusion module merges the local features with the global context information to significantly enhance the feature representation. The Global Wheat Head Detection Dataset and Wheat Ear Detection Dataset are used to assess the proposed model. There were 3.40 and 5.21 average absolute errors, respectively. The performance of the proposed model was significantly better than previous studies.

KEYWORDS

deep learning, transformer, wheat ear counting, density detection, feature fusion

1 Introduction

For the management and phenotyping of wheat production, an accurate wheat ear count is essential. Traditional techniques of counting wheat mostly rely on manual counting, which is prone to human subjectivity, leading to inaccurate identification and omission of wheat ears and low efficiency (Fernandez-Gallego et al., 2018; Zhou et al., 2018;

Sadeghi-Tehran et al., 2019). To increase the efficiency and accuracy of counting wheat, effective automatic counting techniques must be created (Patrício and Rieder, 2018).

Deep learning techniques have been frequently used in wheat counting tasks because of their strong computational and feature extraction capabilities. The current methods used for wheat ear counting mainly include three methods: image segmentation, object detection, and density estimation. For high-density wheat ear counting in complex field backgrounds, it mainly relies on object detection methods and density estimation methods. The former uses deep learning algorithms to pinpoint and identify specific wheat ears.

In order to automatically count wheat ears, (Cao et al., 2020) used the deep learning model efficientdt-d3 for target detection while counting the number of targets based on image processing and deep learning. (Li et al., 2022) used a two-stage detection method to explore the use of Faster R-CNN model to detect the number of ears per unit area of wheat based on RGB images and applied it to genotype analysis. The model successfully localized the target genotypes and successfully predicted the quantity of wheat ears with an average accuracy of 86.7%. An enhanced Yolov4 model based on the Convolutional Block Attention Module (CBAM) (Woo et al., 2018) was also proposed by (Yang et al., 2021). The model can adaptively learn and select significant spatial and channel information in the feature map, efficiently eliminate background interference, and improve the network's ability to extract features by introducing spatial and channel attention methods. (Sun et al., 2022) suggested an enhanced wheat ear counting network (WHCnet) to address the issue that wheat ears are frequently overlooked. Utilizing both the original information and the underlying knowledge of the features, adaptive pooling of original information using Augmented Feature Pyramid Network (AugFPN) (Guo et al., 2020) enhances the effectiveness of wheat ear identification. (Wen et al., 2022) introduced a weighted Bidirectional Feature Pyramid Network (BiFPN) (Tan et al., 2020) into the feature pyramid network of RetinaNet (Lin et al., 2017) to fuse multiscale features. Meanwhile, focus loss and attention modules were added to train a RetinaNet based on several optimizations, which successfully achieved effective detection and counting of wheat ears. The wheat counting algorithm based on object detection mentioned earlier uses convolutional operations to obtain local features of the image, and obtains more comprehensive and rich feature maps through multi-scale and attention mechanisms, without considering the global contextual information of the image. In addition, in order to achieve optimal performance, the high performance of object detection algorithms largely rely on clear images and accurate labeling. However, it takes a lot of time and effort to classify small objects using bounding boxes, particularly for high-density wheat images.

In recent years, scientists have given the density map-based wheat counting model a lot of attention. For counting and localization, the method uses point-labeled data, which requires less effort than box labeling. Convolutional neural networks are used to create high-quality density maps that accurately depict the distribution of wheat. The density maps may be summed to calculate the number of wheat ears, and they are ideally suited to tiny and dense situations. (Pound et al., 2017) built an hourglass network for multi-task learning based on an encoder/decoder structure to simultaneously localize ear characteristics and categorize awn phenotypes in wheat grown in greenhouses. (Khaki et al., 2022) proposed a new method for wheat ear counting called WheatNet. Accelerated the detection rate by using MobileN-etV2 (Sandler et al., 2018) as the primary feature extractor with less parameters. For wheat ear counts and localization, the model uses two parallel subnets that complement one another and boost prediction precision. (Ma et al., 2022) proposed a transfer learning method of the ground-based fully convolutional network. Filter pyramid blocks and dilated convolution are combined to successfully address the issue of counting performance degradation brought on by a reduction in ground resolution. In order to locate and count wheat ear points, (Zaji et al., 2022) developed a hybrid Unet (Ronneberger et al., 2015) structure combining a point-labeled dataset and a constant density graph generation algorithm, considerably enhancing the accuracy of wheat counting. (Xiong et al., 2019) built the tasselnetv2 local regression network for counting wheat by including a contextual information extraction module to the local patches. Without expanding the model's capacity, the accuracy of the counting was increased. (Lu and Cao, 2020) implemented a fast version of TasselNetV2, TasselNetV2+, by splitting TasselNetV2 into an encoder, a counter, and a normalizer based on a novel framework view of TasselNetV2. This fast version improves the speed by an order of magnitude compared to TasselNetV2, while maintaining the same level of counting accuracy. The wheat ear counting method based on density estimation performs better in dealing with complex backgrounds and high-density wheat ears, as it focuses on the overall distribution of wheat ears rather than the precise position of individual wheat ears. Compared to object detection methods, density estimation methods typically have higher computational efficiency and lower costs. The above research indicates that the use of object detection methods and density estimation methods in agriculture can significantly improve the automation level of agricultural production and produce better results. The current methods are mainly based on convolutional neural networks because of their powerful local feature extraction capabilities, which perform better than the traditional methods (Liu et al., 2019) based on artificial feature extraction. However, because the convolutional operation is limited to the convolutional kernel acceptance domain, it is unable to comprehend an image's global information. In most computer vision tasks, this global information and long-range feature dependence are crucial components. Because the self-attention mechanism can handle a long range of feature dependencies and has a significant advantage in extracting global context information, Transformers based on the self-attention mechanism have been widely used in vision tasks in recent years (Carion et al., 2020).

The aim of this study is to increase the diversity of features by combining local features and global context information. We introduce a new network for wheat ear counting with local-global features fusion based on hybrid architecture. The main contributions of this paper can be summarized as follows: (1) We design a new model that combines a Cross Stage Partial (CSP) (Wang et al., 2020) for extracting local features with a Pyramid Pooling Transformer (P2T) (Wu et al., 2023) for capturing global context information to acquire features at multiple scales. Our method has stronger feature extraction capability and significantly improves the wheat counting performance. (2) We propose a feature fusion module for fusing image local features and global context information. The module can effectively integrate local features and global context information. (3) We design a new hybrid loss function that combines the wheat counting loss and the attention loss to train the hybrid network and improve the counting performance of the model. (4) Our proposed CTHNet method is compared with the counting methods proposed in previous research. Experimental results on two commonly used datasets, Global Wheat Head Detection Datasets (GWHD) (David et al., 2021) and Wheat Ears Detection Dataset (WEDD) (Madec et al., 2019), show that our method achieves better performance in the wheat counting task.

The structure of the paper is as follows: the second part describes the selection of the two public wheat datasets, the data preprocessing process, and a general introduction to the proposed model. The third part includes the evaluation metrics, experimental setup, performance comparison, and a discussion of the method of this paper and other comparative methods. Finally, the fourth part summarizes the main findings of the study and looks at possible future research directions.

2 Materials and methods

2.1 Data set and processing

2.1.1 Data set

In this work, we utilized two publicly available wheat ear datasets: (1) GWHD, (2) WEDD. Example images from two datasets were shown in Figure 1. In addition, the details of each dataset are as shown in Table 1.

The GWHD is a large-scale image dataset specially used to detect wheat ears. The data set aims to promote the research and development of computer vision and machine learning in the field of agriculture. As shown in Figure 1, this dataset contains a large number of high-resolution wheat field images from all over the world. The images cover different geographical areas, different planting conditions, and different wheat varieties. The shape and color of wheat ears vary greatly in the dataset, and there is overlap between wheat ears. Accurate localization and counting of wheat ears is challenging due to the influence of many factors. Each image is professionally marked with the exact position of the wheat ears, which is represented by a bounding box, providing the accurate position and size information of the wheat ears. The images contain a total of 147,722 wheat ears samples, with an average of 43 wheat ears in each image.

The second dataset is from the public dataset WEDD provided by (Madec et al., 2019). This dataset was collected on a wheat field phenotype analysis platform using a Sony ILCE-6000 digital camera. It contains 236 high-resolution wheat images (6000 × 4000 pixels) with a total of 31,622 wheat ears. The number of ears in each image varies from 80 to 240. Table 1 presents the comprehensive statistics of the training and test data sets.

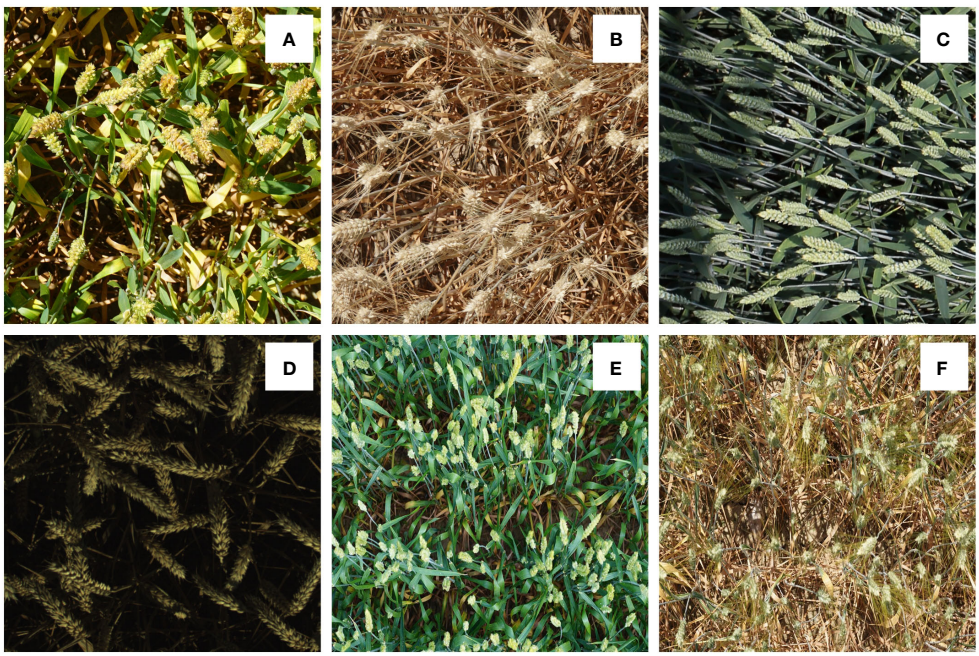


FIGURE 1
Wheat dataset diversity: **(A)** diversity in genotype, **(B)** maturity, **(C)** head orientation, **(D)** lighting, **(E, F)** density.

TABLE 1 The statistics of dataset used in this study.

Data	Dataset	Number of Image	Resolution	Min	Max	Avg	Total
GWHD	Training	3574	1024x1024	4	129	45.67	163,237
	Test	1440	1024x1024	4	179	30.79	44,331
WEDD	Training	165	6000x4000	82	245	133.85	22,085
	Test	71	6000x4000	106	176	134.32	9,537

Min, Max, Avg, and Total denote the minimum, maximum, average, and total number of annotated wheat heads, respectively.

The image in the dataset is reshaped to 1024×1024 pixels in size and include as many wheat ear samples as possible to reduce hardware pressure and unify labeling requirements.

2.1.2 Ground truth density map

This study uses a method based on density map to calculate the number of wheat ears, which needs to use point annotation to generate a real density map. Therefore, we need to use point annotation instead of box annotation to re label the image of dataset. We first use the bounding box annotation provided by dataset to calculate the centroid of each bounding box, so as to obtain the dataset of point annotation.

Let $P = \{p_1, p_2, \ldots\ldots, p_n\}$ denote the annotation set of N wheat ears. Each ear can be represented by a delta function $\delta(x - p_i)$. Therefore, we can represent the ground truth of a wheat ear image with N annotations as follows:

$$\boldsymbol{H(x)} = \sum_{i=1}^{N} \boldsymbol{\delta(x - p_i)}$$

Where p_i is the position of the ith wheat ear and x is the all-zero matrix of the same size as the labeled image. The function $\delta(x - p_i)$ purpose is to set the p_i position in the matrix x to 1.

A Gaussian function $G_\sigma(x)$ is used to smooth the discrete density map generated by $H(x)$ and transform it into a continuous function to generate the ground truth density map $F(x)$:

$$\boldsymbol{F(x)} = \sum_{i=1}^{N} \boldsymbol{\delta(x - p_i) * G_\sigma(x)}$$

Where σ represents the standard deviation, which is set to a constant in the density maps we generate. The resulting ground truth density map has the property that the sum on the density map is the same as the total number of small wheat ears in the image.

Based on the generated density map, proceed to compute the attention map using the Gaussian kernel as follows:

$$\boldsymbol{Z = F(x) * G_\sigma\ (x)}$$

$$\boldsymbol{\forall x \in Z, A(x)} = \begin{cases} \boldsymbol{0}, & \boldsymbol{x < th} \\ \boldsymbol{1}, & \boldsymbol{x \geq th} \end{cases}$$

Where th is the threshold set in our experiment, which is set to 0.001 (Zhu et al., 2019).

2.1.3 Data augmentation

To obtain high-quality datasets, additional data augmentation methods such as random cropping are added to increase the variability of the training set. Considering the distribution density of small wheat ears in wheat images, we chose an image size of 512 pixels for random cropping.

In this study, multiple enhancement techniques were randomly applied to the images. Since all augmentations are done randomly, the model cannot see exactly similar input images during training. This means that the model needs more time to train thoroughly to improve the generalization performance of the model. The different data augmentation techniques used in training are shown in Figure 2. After cropping, the enhancement process includes the hue change of the image and the flip of the picture, and these enhancements are all done randomly on the image with a probability of 0.5. This means that there is a 50% chance of applying each enhancement method to the original image.

2.2 Hybrid network model

As shown in Figure 2, we constructed a network structure in which the input images for the training process are obtained by randomly clipping the original images. Taking the input image size of 512 × 512 pixels as an example. Firstly, local features are extracted through convolution network, using four convolutional kernels with of 2 and channel dimensions of 64, 256, 512and 512 to obtain multiscale feature maps with sizes of 256 × 256 pixels, 128 × 128 pixels, 64 × 64 pixels and 32 × 32 pixels respectively. The large-scale feature map is derived from the shallow network, which has high resolution and good detail feature representation ability. On the contrary, the small-scale feature map is extracted from the deep network and contains more semantic information, but the resolution is low and the representation ability of spatial details is weak. Patch embedding is applied to extract 1×1 patch from the minimum scale feature map extracted by CNN, capturing the global contextual information of the feature maps while keeping the resolution constant. This is helps to consider local details and global information in image processing and improves the understanding of image content. Then, multi-scale feature fusion is carried out through the attention mechanism to output the prediction density map. When generating the density map, the

10.3389/fpls.2024.1425131

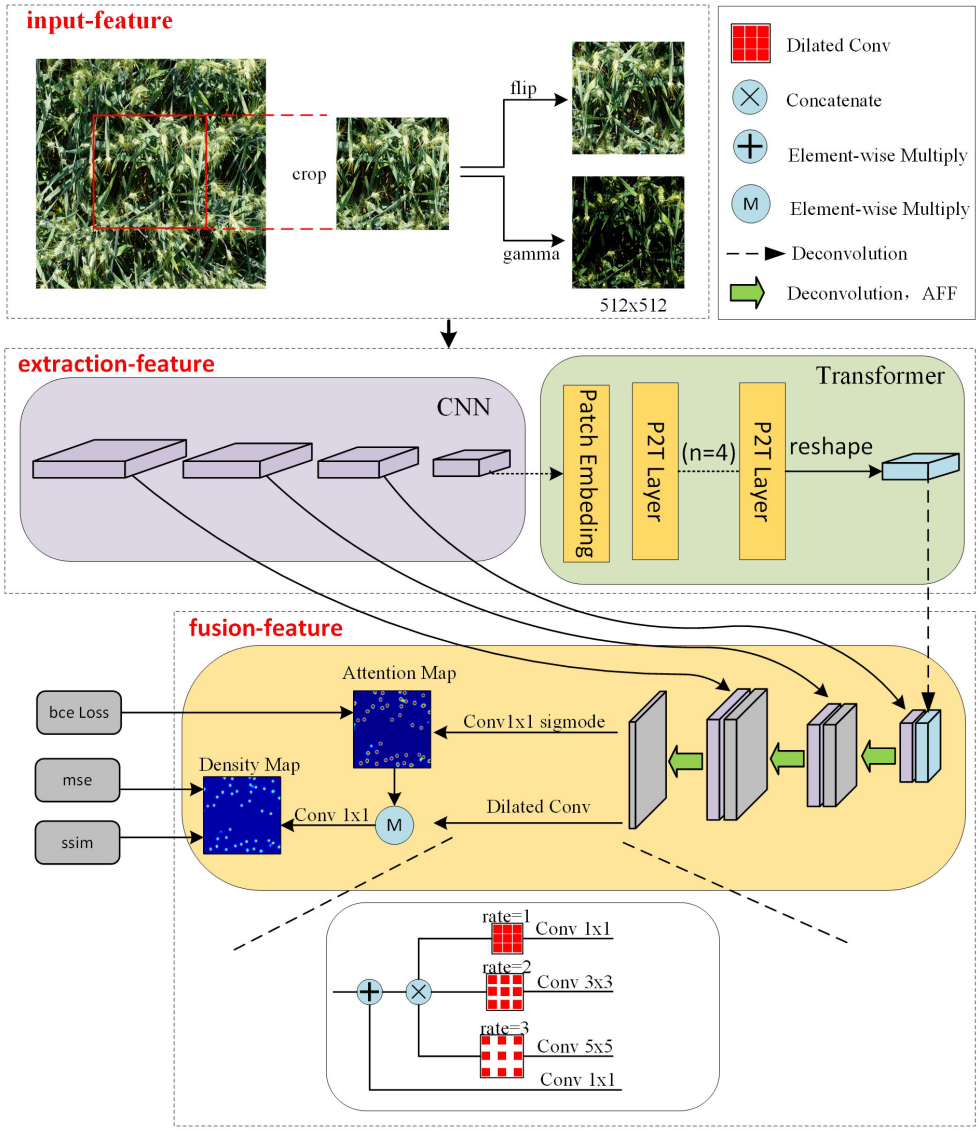


FIGURE 2
The overall framework of hybrid network includes feature extraction module, multi-scale feature aggregation module.

dilated convolution layers with different expansion rates are superimposed in parallel to make the predicted density map contain more spatial information and expand the perception domain. Taking the RGB image as input, three features are input, and the features are increased to 512 in the feature extraction stage, and then the features are reduced to two by the feature fusion, one for the loss calculation of the density map and one for the loss calculation of the attention map.

2.2.1 Feature extraction

The hybrid network is mainly composed of two parts, which realize the task of wheat ear density counting. These two parts are feature extraction and feature fusion. In the feature extraction, we combine the self-attention mechanism with CSP module to build a new network structure. This special design structure can effectively extract local features and global context. Through CNN convolution and global attention mechanism, the local details and global context information of the image can be paid attention to simultaneously in the process of feature extraction. In the process of feature fusion, we enhance the detection ability of the hybrid attention network for wheat at different scales by fusing multi-scale local and global features.

In our proposed backbone network, CSP was used to construct a CNN network to extract the local features of the image. CSP structure is a convolutional neural network structure commonly used to construct the backbone network. By dividing the feature map of the base layer into two parts, the gradient flow propagates through different network paths, finally the feature maps of the beginning and end stages of the network are integrated. It can effectively improve the learning ability and training speed of convolutional neural network.

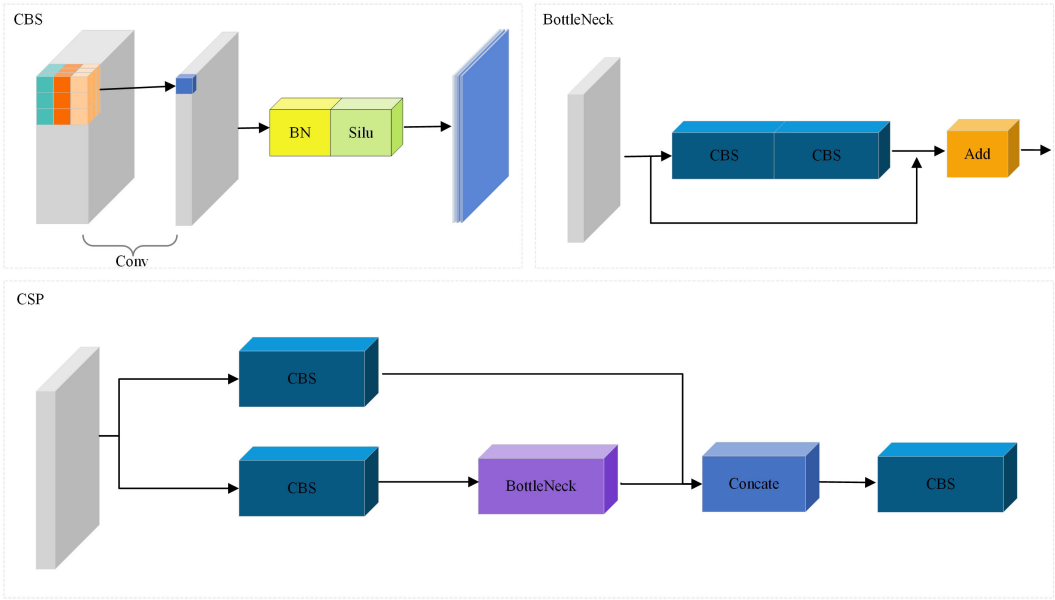


FIGURE 3
Illustration of CBS, BottleNeck and CSP.

As shown in Figure 3, the core idea of the CSP structure is to divide the input feature map into two branches. One branch is responsible for extracting low-level features, while the other branch is responsible for further processing these features to obtain high-level semantic information. Feature fusion and recombination are realized by introducing cross connection and partial connection. This design can improve the ability of feature representation and enable the network to better capture feature information at different scales.

In addition, we also use the P2T structure to capture global context information. It introduces pyramid pooling in the multi-head self-attention module of vision Transformer. The benefits are reduced sequence length, reduced computation, and the ability to simultaneously learn strong contextual representations for better semantic features. The calculation process is shown in Figure 4. The input first passes through a pooling based multi head self-attention module, and its output adds residuals and performs normalization layer. Like traditional transformer blocks, a Feedforward Network (FFN) (Vaswani et al., 2017) is used for feature projection. Finally, apply residual connections and a normalization layer again.

Our backbone network combines CSP structure and P2T structure to form a hybrid network. Firstly, we use the CSP module to perform dimensionality reduction and local feature extraction on the input RGB image. The purpose of this method is to preserve key information while reducing subsequent computational complexity. Next, we use the P2T structure to capture global contextual information. In order to fully capture the multi-scale information of the target, we designed the feature extraction network as a pyramid structure. This is conducive to the generation of multi-scale feature representations and improves the performance and robustness of tasks.

2.2.2 Feature fusion

Although the backbone network can extract multi-scale local features and global context information, the high-level features may lead to the loss and ambiguity of local information after deep

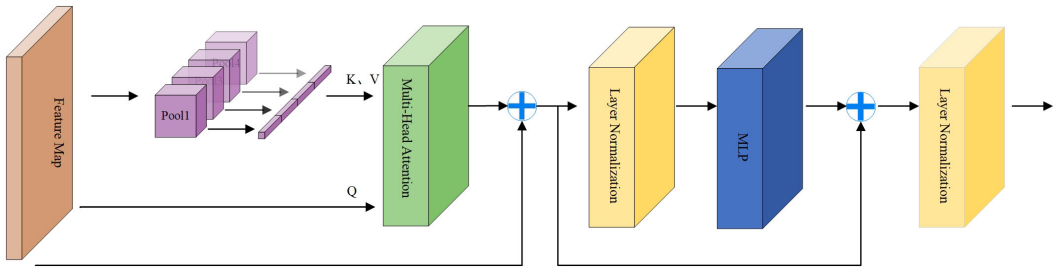


FIGURE 4
Architecture of the Multi-head Pyramid Pooling Transformer.

convolution operation. High-level features focus on capturing more abstract and richer semantic features in the image. However, it loses some important details. On the contrary, low-level features are extracted from the shallow layer of the neural network and focus on capturing the details and local features of the input data. But it lacks high-level semantic information, and its feature representation capability is limited.

In order to fully integrate multi-scale features, we use an Attention Feature Fusion module (AFF) (Dai et al., 2021). This module uses two branches with different scales to extract channel attention for global features and channel attention for local features. As shown in Figure 5, the key idea is that channel attention can be achieved at multiple scales by varying the size of spatial pooling.

The purpose of AFF module is to comprehensively use global and local feature information to enhance the ability of feature representation. Global attention helps the model focus on the semantic content of the overall image, while channel attention helps the model adaptively learn the importance of different channels, so as to further integrate the information of different features. In this way, the local features and global context information in the image can be better fused, so as to improve the performance and robustness of the model.

2.2.3 Loss function

Most of the existing methods are based on the assumption of pixel independence and use pixel-wise Euclidean loss to train the network. It is defined as follows:

$$\mathbf{L_{den}} = \frac{1}{\mathbf{N}}\sum_{\mathbf{i=1}}^{\mathbf{N}}\|\mathbf{P(x_i)} - \mathbf{G_i}\|^2$$

Where $P(x_i)$ is the predicted density map of the ith input image, G_i is the true density map of the ith input image, x_i represents the i th input image, and N is the number of input images.

In addition to the pixel-wise loss function, we add a local correlation loss to the density map to improve the quality of the predicted density map. The SSIM index is used to measure the local pattern consistency between the estimated density map and the ground truth. From three local statistics, namely mean, variance and covariance, the similarity between two images is calculated (Cao et al., 2018). The local statistics are computed by:

$$\mu_x = \frac{1}{M}\sum_{i=1}^{M} x_i$$

$$\sigma_x^2 = \frac{1}{M-1}\sum_{i=1}^{M}(x_i - \mu_x)^2$$

$$\sigma_{xy} = \frac{1}{M-1}\sum_{i=1}^{M}(x_i - \mu_x)(y_i - \mu_y)$$

Where μ_x and σ_x^2 are the local mean and variance estimates of image x, M represents the number of pixels of image x, and σ_{xy} is the local covariance estimate, so the SSIM index is calculated point by point as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Where C_1, C_2 are minimal constants, to avoid division by zero and set as (Wang et al., 2004). The SSIM value ranges from -1 to 1 and takes the value 1 when two images are identical. Therefore, the local consistency loss is as follows.

$$L_C = 1 - SSIM(x,y)$$

L_C is the local pattern consistency loss, which measures the local pattern difference between the estimation results and the ground truth.

In addition to computing the predictive density map loss, we also introduce an attention map loss function, which is a binary class entropy defined as follows.

$$L_{att} = -\frac{1}{N}\sum_{i=1}^{N}(A_i^{GT}\log(P_i) + (1 - A_i^{GT})\log(1 - P_i))$$

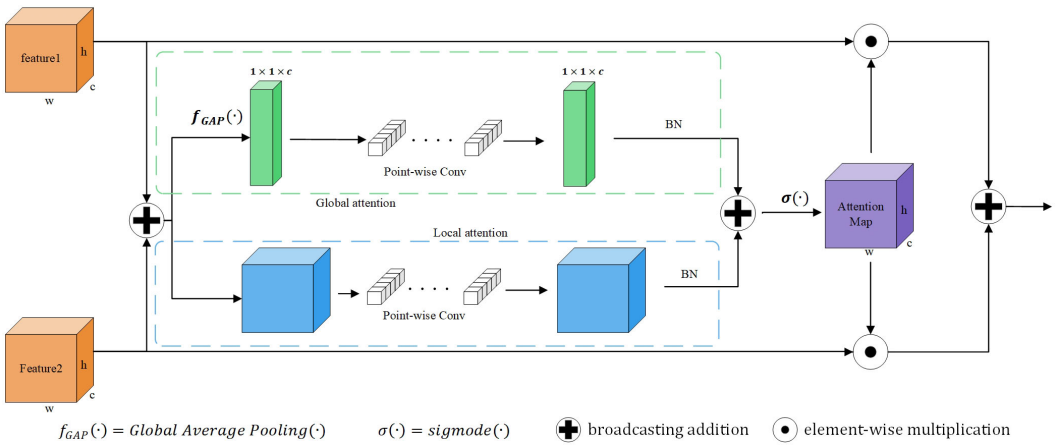


FIGURE 5
Illustration of the Attention Feature Fusion (AFF).

Where A_i^{GT} is the attention map groundtruth and P_i is the probability that each pixel in the predicted attention map is activated by the sigmoid function.

For high-density wheat ear images in complex backgrounds. Euclidean loss is used to train counting loss, while SSIM is used to train density map similarity loss. By combining these two loss functions, the aim is to improve the counting accuracy of high-density wheat ear images in complex backgrounds. The entire network is trained using the following unified loss function:

$$L = L_{den} + L_c + \alpha L_{att}$$

Where α is a weighted weight, which is set to 0.1 (Zhu et al., 2019) in the experiments. We exploit this combined loss for end-to-end training.

3 Experiment and result

This section presents the metrics used for the experimental evaluation, the training hyperparameters, and the final results. All experiments were trained on a computer equipped with a 24GB Geforce RTX 3090 graphics card. The deep learning framework used in the code is Pytorch 1.12 and the programming language is python 3.9. Adam (Adaptive Gradient Descent) method is used to optimize the learning rate during training. The initial learning rate is set to 0.0001. This configuration allows model training and evaluation to be performed efficiently.

3.1 Model performance evaluation

The performance of the model in the prediction and counting accuracy of Macintosh was evaluated using the mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) as indicators. MAE and RMSE are commonly used in previous density estimation studies to measure the error between the predicted value and the actual value. MAPE is a measure of relative error, using absolute values to avoid bias between positive and negative errors, and is one of the most commonly used metrics for evaluating prediction performance. They can be expressed as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_i - G_i|$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|P_i - G_i|^2}$$

$$MAPE = \frac{100\,\%}{N}\sum_{i=1}^{N}\left|\frac{P_i - G_i}{G_i}\right|$$

Where N is the number of test images, $\bar{P}$ is the average value of the predicted value, P_i and G_i are the predicted number and the real value in the i-th image respectively. RMSE is used to measure the average error between the predicted value and the real value of the model. The smaller the RMSE, the closer the prediction of the model to the real value. MAE is used to measure the average absolute error between the predicted value and the real value of the model. The smaller the MAE, the more accurate the model prediction and the higher the counting accuracy of the model. By calculating these indicators, we can evaluate the accuracy of the model in predicting the number of wheats ears and compare it with the actual value manually marked to measure the performance and accuracy of the model.

3.2 Ablation experiments

In this study, we used ablation experiments to explore the effects of P2T, CSP and SSIM loss on the model. As shown in Table 2, where V1 represents the model built only with CSP, V2 represents the integration of P2T module on the basis of V1, similarly, V3 represents the addition of AFF multi-scale fusion on the basis of V1, and CTHNet represents the final algorithm proposed in this paper, V4 represents the result obtained by training the model without using SSIM loss. The test results show that the integrated P2T and CSP are sufficient to ensure acceptable accuracy (MAE=5.61, RMSE=6.77). Moreover, the addition of AFF further integrates local features and global context information and leads to more accurate detection accuracy (MAE=5.21, RMSE=5.27), compared with using CSP alone. SSIM loss is used to compare the similarity between real density maps and predicted density maps. By adding SSIM loss, the model's counting performance can be effectively improved. Table 2 shows a decrease of 0.18 in MAE and 1.48 in RMSE. These findings confirm that the model effectively improves the counting accuracy of the model by fusing global context information and local features.

We showed the change in loss curves to provide a more intelligible representation of the loss change. Figure 6 depicts the variations in training loss and SSIM loss values for each improved version. It can be observed from Figure 6 that the loss value decreases rapidly at the beginning of training, and with the increase of training epochs, the training loss gradually decreases and fluctuates near a critical value. From Figure 7, it can be seen that the density map generated by integrating local and global features is clearer and more pronounced than the density map generated solely based on local features, especially in the edge and high-density areas. Notably, the loss function of the final algorithm proposed in this paper still continues to decrease over a range of 40 to 100 epochs, suggesting that the model has a good capacity for learning.

TABLE 2 Performance metrics of CTHNet before and after improvement on the WEDD.

Model	CSP	P2T	AFF	SSIM	MAE	RMSE
V1	√	–	–	√	5.67	7.26
V2	√	√	–	√	5.61	6.77
V3	√	–	√	√	5.68	7.95
V4	√	√	√	–	5.39	6.75
CTHNet	√	√	√	√	5.21	5.27

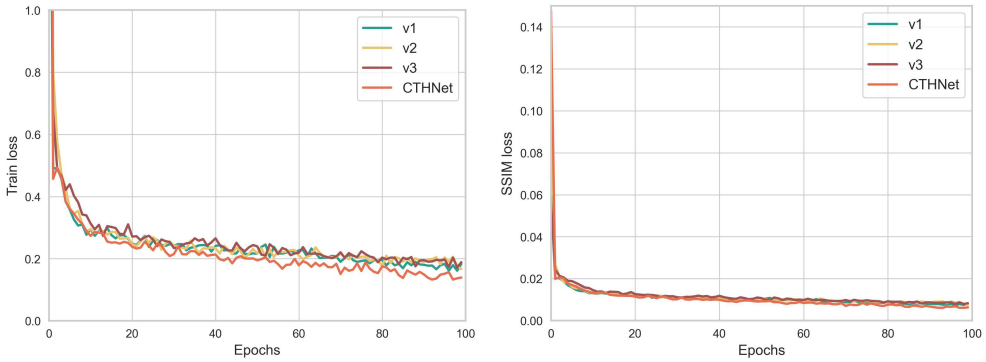


FIGURE 6
Results from the training process: Train loss and SSIM loss.

3.3 Comparison experiments

In recent years, density regression algorithms have progressed quickly and been used in a variety of counting applications. We contrast our method with various density counting methods, such as MCNN (Zhang et al., 2016), SANet (Cao et al., 2018), SCAR (Gao et al., 2019), PSNet (Cheng et al., 2020), TasseLnetv2+ (Lu and Cao, 2020) and CCTrans (Tian et al., 2021). We evaluated the counting performance of the proposed model on two public wheat ear detection datasets GWHD and WEDD (Table 3).

We perform visual analysis on two datasets to show the counting performance of the proposed method more intuitively and scientifically. We use the test images of GWHD and WEDD datasets to compare the number of true peak annotations of various counting methods (Figure 8). We also collected and compared the predicted density maps generated by different counting models (Figure 9).

Table 3, show that hybrid network counting performs better than other counting methods based on density estimation. Since the ear of wheat is a small target, its feature is easily contaminated

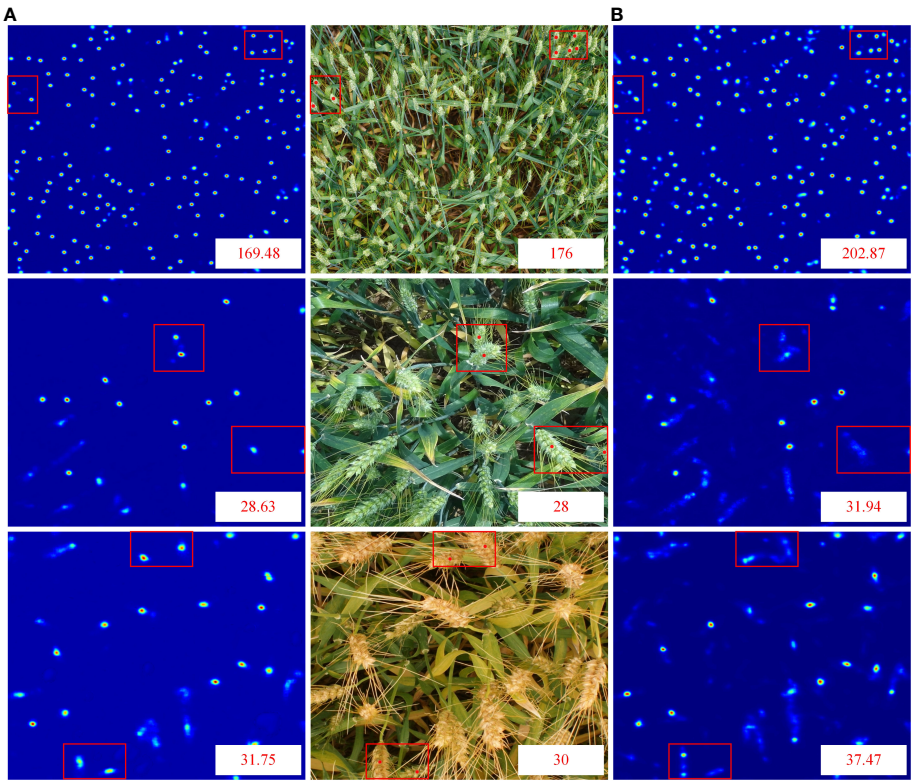


FIGURE 7
Comparison of density maps: **(A)** Density maps generated by combining local and global features **(B)** Density maps generated based on local features.

10.3389/fpls.2024.1425131

TABLE 3 Compared with other density estimation methods on the GWHD and WEDD dataset.

Model	GWHD			WEDD		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
MCNN	8.49	17.61	27.65	10.69	14.29	7.79
SANet	4.80	11.28	13.72	11.88	13.80	8.78
SCAR	4.48	8.03	15.63	6.38	8.59	4.77
TasseLnet v2+	6.44	11.97	20.50	7.93	9.97	5.97
CCTrans	5.10	11.07	15.01	6.91	8.45	5.26
PSNet	4.62	8.64	15.64	6.29	7.86	4.77
CTHNet	3.40	5.75	12.47	5.21	5.27	3.95

during the extraction process by examples like background, which makes it challenging for the network to capture the discriminative information required for following tasks (Shi et al., 2023). CNN-based counting models MCNN, SANet and TasseLnetv2+ learn high-dimensional features and reduce spatial redundancy through down-sampling operations, but this leads to the loss of representation for small objects, making accurate counting difficult. To overcome this problem, CCTrans uses a pyramid visual transformer backbone to capture global information and complete the counting task by combining low-level and high-level features. The experimental results show that CCTrans performs well in counting effect. In addition, SCAR and PSNet show better counting performance than MCNN and SANet by adding attention to enhance the extraction ability of small target features, but they fall short of our proposed model. This is because the convolution operation mainly extracts semantic features at high-level, and the potential features for small targets may be lost or compressed. The convolutional network also has some challenges in the recognition of small targets, especially when dealing with scenes with complex backgrounds. Convolutional operations and self-attention mechanisms are used to extract multi-scale local features and global context information. This lessens the detrimental effects of the convolution constraint by enabling the full utilization of the global context information. Meanwhile, a feature fusion module based on local and global attention is used to combine local features with global context information, improving the capability of feature representation.

In addition, we use dilated convolution to improve the density map's receptive field, keeping the context and small item details of the wheat image. Compared with other wheat ear number models, the network model performs well in complex background (Table 3).

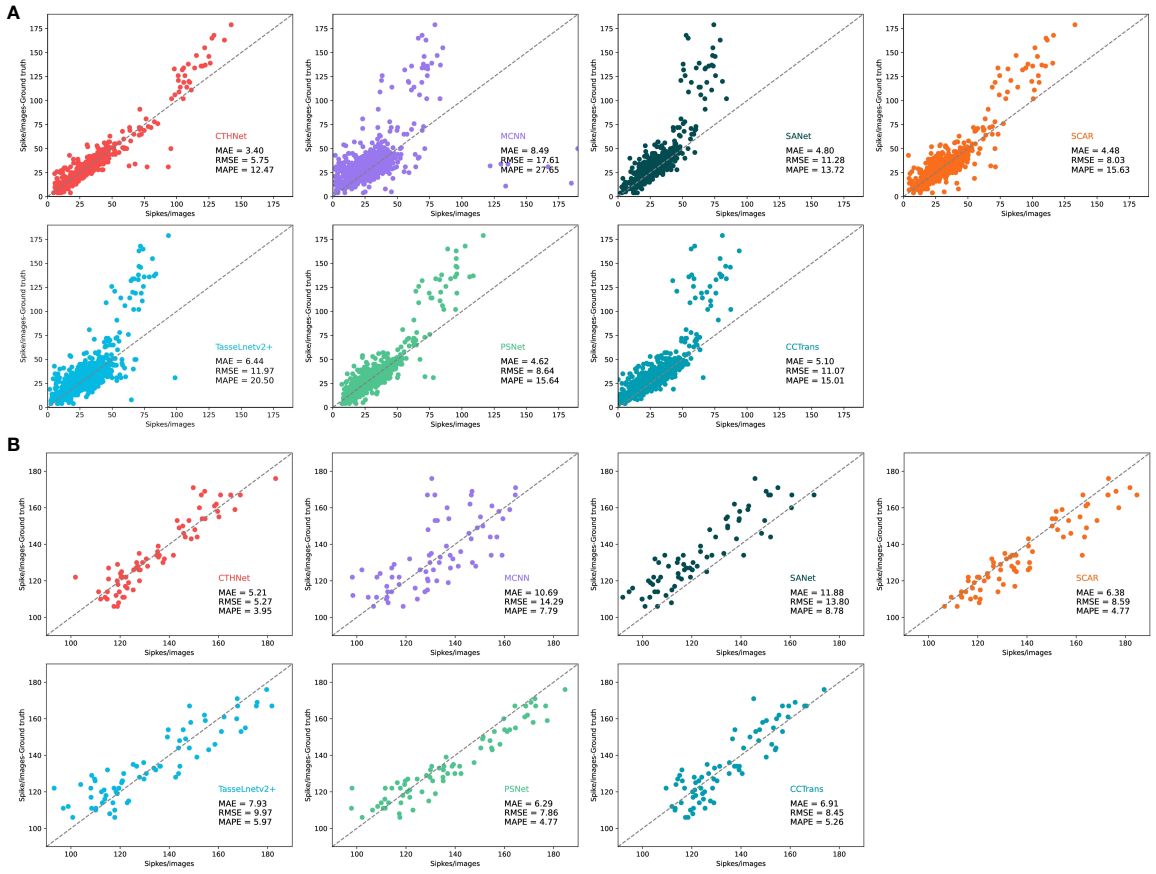


FIGURE 8
Comparison of the number of real ears annotated on the GWHD test dataset **(A)** and the WEDD test dataset **(B)** with that estimated by MCNN, SANet, SCAR, TassLnetv2+, PSNet, CCTrans and CTHNet.

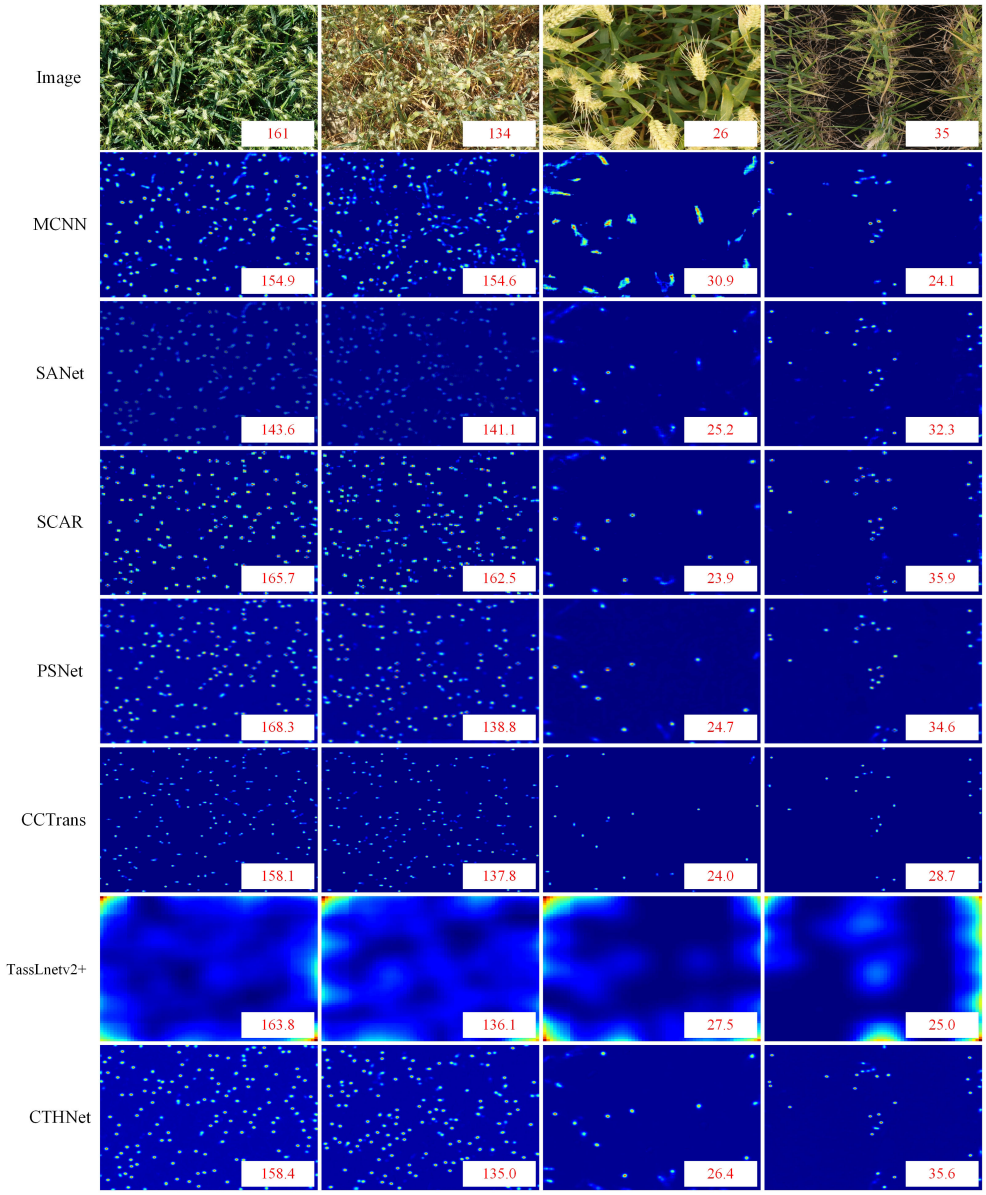


FIGURE 9
Comparison of predicted density maps generated by different models on images.

Compared with SANet, MCNN, PSNet, SCAR, TasseLnetv2+ and CCtrans. MAE improves the performance on GWHD dataset by 6.67, 5.48, 1.08, 1.17, 2.04, 1.70. On the WEDD dataset, the performance is improved by 1.40, 5.09, 1.22, 1.08, 2.72, 1.70. By looking at the density plots of the predictions of different models (Figure 9). We find that the predictive density maps generated by the hybrid network can better approximate the ground truth than the other models.

On the test images of the two datasets, we compared the number of true peak annotations for various counting methods (Figure 8). It can be found that the performance of the model on the GWHD dataset is better than the WEDD dataset. This is mainly because the wheat density in WEDD dataset is much higher than GWHD, and the occlusion between wheat ears and background noise are more likely to cause confusion interference, which increases the difficulty of feature extraction and counting. However, on the high-density WEDD dataset, the hybrid model continues to outperform the other models in terms of counting ability, indicating that the hybrid network design can successfully address the issues of misdetection and omission brought on by the overlap of wheat ear counts and the complexity of the background. The proposed method not only improves the accuracy of the model, but also has wide applicability and is able to deal with wheat counting tasks in different scenarios. This is of great significance for improving the efficiency of agricultural production, reducing resource waste, and for farmland management and crop production.

3.4 Discussion

We create a hybrid network using convolution operation and self-attention mechanism to fully extract local features and global context information of wheat ear images. Experimental results on two public datasets show that the proposed method has good counting and generalization ability.

Firstly, SCAR and PSNet with the added attention mechanism outperform the convolution based MCNN, SANet and TasseLnetv2+ in counting performance. This is because the attention mechanism highlights the key features of the wheat ears and ignores unnecessary regions by assigning different weights to different parts of the feature map. However, since the context information in the image is ignored, SCAR and PSNet with the attention mechanism are not as good as the hybrid network proposed in this study in counting performance. CCTrans utilizes the pyramid vision transformer as a backbone, making it easy to extract global context information. However, the counting results of CCTrans in GWHD and WEDD are still lower than the model proposed in this paper. In contrast, by combining the convolutional operation with the self-attention mechanism and adding the attention feature fusion to the hybrid model, the global context information and local details of the image can be effectively fused. By fusing these two kinds of information, the relationship between various spatial information can be strengthened, thus improving the recognition ability of wheat ears on the feature map.

In this study, we realized that there are some differences in features between wheat images and dense crowd images. The shape of a wheat ear is usually rectangular, while the shape of a human head is round. Therefore, the traditional annotation methods and density map generation methods for dense crowd counting tasks cannot describe the characteristics of wheat ears well. To solve this problem, we plan to design a suitable labeling method for wheat ears to accurately describe the shape and location of wheat ears. For example, we can use lines to describe the shape of the wheat ear or make point annotations on the head of the wheat ear to describe the position of the wheat ear. With this more refined labeling method, the characteristics of the wheat ear can be better captured, including the shape characteristics and location characteristics of the wheat ear.

Additionally, the phenotypic characteristics of a single wheat ear, including the spike length, spike width, and grain number, are significant indicators of the quality and growth of the crop. The traditional way of gathering phenotypic information on wheat ears includes manual measurement and counting, which is time-consuming and labor-intensive and significantly slows down wheat ear research. We will implement the batch extraction of phenotypic data from single wheat ears using the above-mentioned study methods and data combined with an example segmentation algorithm. We will next propose a scheme for extracting phenotypic information of wheat ears in a complex background in the field. Through this program, the staff can collect wheat phenotype data in the field in real time, saving data collection time and improving research efficiency, and provides strong support for agricultural breeding and phenotyping work.

4 Conclusion

In this study, we propose a hybrid network model that combines convolutional operations and a self-attention mechanism. This model is used to generate high-quality wheat ear density maps for accurately calculating the number of wheat ears in complex background images. The results demonstrate a significant improvement in the wheat ear counting task achieved by the hybrid network. By fusing local features of wheat ears and global context information, the edge and small target information in the image can be effectively preserved, and the counting performance is improved.

The hybrid network has good feature representation ability and can meet the requirements of wheat ear number under the conditions of occlusion and overlap. It provides a reliable estimate of wheat yield and provides strong support for agricultural production. In the future research, we will focus on the design of labeling method and density map generation method for wheat counting task, in order to further improve the accuracy and performance of hybrid network. We will also develop precision agriculture applications that utilize unmanned aerial vehicle (UAV) to collect wheat ear image data at different times, varieties and planting densities, further validating the performance of the model and improving its generalization ability. This will provide real-time and accurate information for agricultural production, help farmers make scientific decisions, and improve crop management and yield.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: http://www.global-wheat.com/ https://github.com/simonMadec.

Author contributions

QH: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – review & editing. WL: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. YZ: Software, Writing – review & editing. TR: Software, Writing – review & editing. CS: Supervision, Writing – review & editing. ZL: Supervision, Writing – review & editing. YY: Supervision, Writing – review & editing. RD: Supervision, Writing – review & editing. JQ: Supervision, Writing – review & editing. CT: Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The research was funded by the Key Research and Development Program of Jiangsu Province, China (BE2022337, BE2022338, BE2023302, BE2023315), the National Natural Science Foundation of China (32071902, 42201444), A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), the University Synergy Innovation Program of Anhui Province(GXXT-2023-101), the Jiangsu Agricultural Science and Technology Innovation Fund (CX(22)3149), and the Open Project for Joint International Research Laboratory of Agriculture and Agri-Product Safety of the Ministry of Education of China (JILAR-KF202102).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Cao, L., Zhang, X., Pu, J., Xu, S., Cai, X., and Li, Z. (2020). "The field wheat count based on the efficientdet algorithm," in *2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE)*. Dalian, China: IEEE. 557–561. doi: 10.1109/ICISCAE51034.2020.9236918

Cao, X., Wang, Z., Zhao, Y., and Su, F. (2018). "Scale Aggregation Network for Accurate and Efficient Crowd Counting," in *Computer Vision – ECCV*, vol. 2018 . Eds. V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss (Switzerland: Springer International Publishing), 757–773. doi: 10.1007/978-3-030-01228-1_45

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020). "End-to-End Object Detection with Transformers," in *Computer Vision – ECCV*, vol. 2020 . Eds. A. Vedaldi, H. Bischof, T. Brox and J.-M. Frahm, (Switzerland: Springer, Cham) 213–229. doi: 10.1007/978-3-030-58452-8_13

Cheng, J., Chen, Z., Zhang, X., Li, Y., and Jing, X. (2020). Exploit the potential of Multi-column architecture for Crowd Counting. *arXiv. arXiv:2007.05779v1*. abs/2007.05779, arXiv:2007.05779. doi: 10.48550/arXiv.2007.05779

Dai, Y., Gieseke, F., Oehmcke, S., Wu, Y., and Barnard, K. (2021). "Attentional feature fusion," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Waikoloa, HI, USA: IEEE. 3559–3568. doi: 10.1109/WACV48630.2021.00360

David, E., Serouart, M., Smith, D., Madec, S., Velumani, K., Liu, S., et al. (2021). Global wheat head detection 2021: An improved dataset for benchmarking wheat head detection methods. *Plant Phenomics* 2021, 2021/9846158. doi: 10.34133/2021/9846158

Fernandez-Gallego, J. A., Kefauver, S. C., Gutiérrez, N. A., Nieto-Taladriz, M. T., and Araus, J. L. (2018). Wheat ear counting in-field conditions: High throughput and low-cost approach using RGB images. *Plant Methods* 14, 22. doi: 10.1186/s13007-018-0289-4

Gao, J., Wang, Q., and Yuan, Y. (2019). SCAR: Spatial-/channel-wise attention regression networks for crowd counting. *Neurocomputing* 363, 1–8. doi: 10.1016/j.neucom.2019.08.018

Guo, C., Fan, B., Zhang, Q., Xiang, S., and Pan, C. (2020). "AugFPN: improving multi-scale feature learning for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE. doi: 10.1109/CVPR42600.2020

Khaki, S., Safaei, N., Pham, H., and Wang, L. (2022). WheatNet: A lightweight convolutional neural network for high-throughput image-based wheat head detection and counting. *Neurocomputing* 489, 78–89. doi: 10.1016/j.neucom.2022.03.017

Li, L., Hassan, M. A., Yang, S., Jing, F., Yang, M., Rasheed, A., et al. (2022). Development of image-based wheat spike counter through a Faster R-CNN algorithm and application for genetic studies. *Crop J.* 10, 1303–1311. doi: 10.1016/j.cj.2022.07.007

Lin, T. Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," in *2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy: IEEE. 2999–3007. doi: 10.1109/ICCV.2017.324

Liu, Z., Wenzhun, H., and Liping, W. (2019). Field wheat ear counting automatically based on improved K-means clustering algorithm. *Nongye Gongcheng Xuebao/Transactions Chin. Soc. Agric. Eng.* 35, 174–181. doi: 10.11975/j.issn.1002-6819.2019.03.022

Lu, H., and Cao, Z. (2020). TasselNetV2+: A fast implementation for high-throughput plant counting from high-resolution RGB imagery. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.541960

Ma, J., Li, Y., Liu, H., Wu, Y., and Zhang, L. (2022). Towards improved accuracy of UAV-based wheat ears counting: A transfer learning method of the ground-based fully convolutional network. *Expert Syst. Appl.* 191, 116226. doi: 10.1016/j.eswa.2021.116226

Madec, S., Jin, X., Lu, H., De Solan, B., Liu, S., Duyme, F., et al. (2019). Ear density estimation from high resolution RGB imagery using deep learning technique. *Agric. For. Meteorology* 264, 225–234. doi: 10.1016/j.agrformet.2018.10.013

Patrício, D. I., and Rieder, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* 153, 69–81. doi: 10.1016/j.compag.2018.08.001

Pound, M. P., Atkinson, J. A., Wells, D. M., Pridmore, T. P., and French, A. P. (2017). "Deep learning for multi-task plant phenotyping," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. Venice, Italy: IEEE. 2055–2063. doi: 10.1109/ICCVW.2017.241

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI*. Eds. N. Navab, J. Hornegger, W. M. Wells and A. F. Frangi (Switzerland: Springer International Publishing), 2015 234–241.

Sadeghi-Tehran, P., Virlet, N., Ampe, E. M., Reyns, P., and Hawkesford, M. J. (2019). DeepCount: in-field automatic quantification of wheat spikes using simple linear iterative clustering and deep convolutional neural networks. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01176

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018). "MobileNetV2: inverted residuals and linear bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE. 4510–4520. doi: 10.1109/CVPR.2018.00474

Shi, L., Sun, J., Dang, Y., Zhang, S., Sun, X., Xi, L., et al. (2023). YOLOv5s-T: A lightweight small object detection method for wheat spikelet counting. *Agriculture* 13, 872. doi: 10.3390/agriculture13040872

Sun, J., Yang, K., Chen, C., Shen, J., Yang, Y., Wu, X., et al. (2022). Wheat head counting in the wild by an augmented feature pyramid networks-based convolutional neural network. *Comput. Electron. Agric.* 193, 106705. doi: 10.1016/j.compag.2022.106705

Tan, M., Pang, R., and Le, Q. V. (2020). "EfficientDet: scalable and efficient object detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE. 10778–10787. doi: 10.1109/CVPR42600.2020.01079

Tian, Y., Chu, X., and Wang, H. (2021). CCTrans: simplifying and improving crowd counting with transformer. *ArXiv. ArXiv:2109.14483* abs/2109.14483, arXiv:2109.14483. doi: 10.48550/arXiv.2109.14483

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, CA, USA: Curran Associates Inc. 6000–6010.

Wang, C. Y., Mark Liao, H. Y., Wu, Y. H., Chen, P. Y., Hsieh, J. W., and Yeh, I. H. (2020). "CSPNet: A new backbone that can enhance learning capability of CNN," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Seattle, WA, USA: IEEE. 1571–1580. doi: 10.1109/CVPRW50498.2020.00203

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861

Wen, C., Wu, J., Chen, H., Su, H., Chen, X., Li, Z., et al. (2022). Wheat spike detection and counting in the field based on spikeRetinaNet. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.821717

Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). "CBAM: Convolutional block attention module," in *Computer Vision – ECCV*. Eds. V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss (Switzerland: Springer International Publishing), 2018 3–19. doi: 10.1007/978-3-030-01234-2_1

Wu, Y. H., Liu, Y., Zhan, X., and Cheng, M. M. (2023). P2T: pyramid pooling transformer for scene understanding. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 12760–12771. doi: 10.1109/TPAMI.2022.3202765

Xiong, H., Cao, Z., Lu, H., Madec, S., Liu, L., and Shen, C. (2019). TasselNetv2: In-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods* 15, 150. doi: 10.1186/s13007-019-0537-2

Yang, B., Gao, Z., Gao, Y., and Zhu, Y. (2021). Rapid detection and counting of wheat ears in the field using yolov4 with attention module. *Agronomy* 11, 1202. doi: 10.3390/agronomy11061202

Zaji, A., Liu, Z., Xiao, G., Bhowmik, P., Sangha, J. S., and Ruan, Y. (2022). Wheat spike localization and counting via hybrid UNet architectures. *Comput. Electron. Agric.* 203, 107439. doi: 10.1016/j.compag.2022.107439

Zhang, Y., Zhou, D., Chen, S., Gao, S., and Ma, Yi. (2016). "Single-image crowd counting via multi-column convolutional neural network," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE. 589–597. doi: 10.1109/CVPR.2016.70

Zhou, C., Liang, D., Yang, X., Yang, H., Yue, J., and Yang, G. (2018). Wheat ears counting in field conditions based on multi-feature optimization and TWSVM. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01024

Zhu, L., Zhao, Z., Lu, C., Lin, Y., Peng, Y., and Yao, T. (2019). Dual path multi-scale fusion networks with attention for crowd counting. *arXiv. arXiv:1902.01115* abs/1902.01115, arXiv:1902.01115. doi: 10.48550/arXiv.1902.01115

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 05 July 2024
DOI 10.3389/fpls.2024.1415297

Check for updates

OPEN ACCESS

EDITED BY
Ting Sun,
China Jiliang University, China

REVIEWED BY
Xiaomei Yi,
Zhejiang Agriculture and Forestry University, China
Dongmei Chen,
Hangzhou Dianzi University, China

*CORRESPONDENCE
Huiping Jiang
jianghp@muc.edu.cn

RECEIVED 10 April 2024
ACCEPTED 07 June 2024
PUBLISHED 05 July 2024

CITATION
Wang S, Jiang H, Yang J, Ma X, Chen J, Li Z and Tang X (2024) Lightweight tomato ripeness detection algorithm based on the improved RT-DETR.
Front. Plant Sci. 15:1415297.
doi: 10.3389/fpls.2024.1415297

COPYRIGHT
© 2024 Wang, Jiang, Yang, Ma, Chen, Li and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Lightweight tomato ripeness detection algorithm based on the improved RT-DETR

Sen Wang [1,2], Huiping Jiang [1,2]*, Jixiang Yang [1,2], Xuan Ma [1,2], Jiamin Chen [1,2], Zhongjie Li [1,2] and Xingqun Tang [1,2]

[1]School of Information Engineering, Minzu University of China, Beijing, China, [2]Key Laboratory of Ethnic Language Intelligent Analysis and Security Governance of MOE, Minzu University of China, Beijing, China

Tomatoes, widely cherished for their high nutritional value, necessitate precise ripeness identification and selective harvesting of mature fruits to significantly enhance the efficiency and economic benefits of tomato harvesting management. Previous studies on intelligent harvesting often focused solely on identifying tomatoes as the target, lacking fine-grained detection of tomato ripeness. This deficiency leads to the inadvertent harvesting of immature and rotten fruits, resulting in economic losses. Moreover, in natural settings, uneven illumination, occlusion by leaves, and fruit overlap hinder the precise assessment of tomato ripeness by robotic systems. Simultaneously, the demand for high accuracy and rapid response in tomato ripeness detection is compounded by the need for making the model lightweight to mitigate hardware costs. This study proposes a lightweight model named PDSI-RTDETR to address these challenges. Initially, the PConv_Block module, integrating partial convolution with residual blocks, replaces the Basic_Block structure in the legacy backbone to alleviate computing load and enhance feature extraction efficiency. Subsequently, a deformable attention module is amalgamated with intra-scale feature interaction structure, bolstering the capability to extract detailed features for fine-grained classification. Additionally, the proposed slimneck-SSFF feature fusion structure, merging the Scale Sequence Feature Fusion framework with a slim-neck design utilizing GSConv and VoVGSCSP modules, aims to reduce volume of computation and inference latency. Lastly, by amalgamating Inner-IoU with EIoU to formulate Inner-EIoU, replacing the original GIoU to expedite convergence while utilizing auxiliary frames enhances small object detection capabilities. Comprehensive assessments validate that the PDSI-RTDETR model achieves an average precision mAP50 of 86.8%, marking a 3.9% enhancement over the original RT-DETR model, and a 38.7% increase in FPS. Furthermore, the GFLOPs of PDSI-RTDETR have been diminished by 17.6%. Surpassing the baseline RT-DETR and other prevalent methods regarding precision and speed, it unveils its considerable potential for detecting tomato ripeness. When applied to intelligent harvesting robots in the future, this approach can improve the quality of tomato harvesting by reducing the collection of immature and spoiled fruits.

KEYWORDS

tomato, ripeness recognition, deep learning, RT-DETR, PConv, deformable attention, slimneck, Inner-EIoU

1 Introduction

Tomatoes are widely favored for their rich content of vitamin C, potassium, and lycopene (Story et al., 2010). During the growth of tomatoes, their color gradually shifts from green to yellow, orange, and finally red, while the firmness decreases, and the sweetness and acidity reach a balance, leading to an increase in nutritional content. The maturity level of tomatoes directly impacts their nutritional value, taste, and the timing of harvesting. Traditional manual detection methods are subjective and are often inefficient and costly, failing to meet the needs for high-efficiency maturity discrimination and harvesting (El-Bendary et al., 2015). Although sensors provide a non-contact means of detecting maturity, they sometimes struggle to accurately distinguish between closely related stages of ripeness, are significantly influenced by ecological elements like illumination and temperature, and are cost-prohibitive (Aghilinategh et al., 2020; Alam Siddiquee et al., 2020). Consequently, the creation of a lightweight, efficient, and precise algorithm for maturity detection is of great importance for the intelligent grading and harvesting of tomatoes However, the natural growth environment of tomatoes, characterized by fruit occlusion, subtle color differentiation, and variations in lighting conditions, presents challenges for the accurate identification of tomato maturity.

To enhance the quality of tomato harvests and reduce labor costs, while also accurately distinguishing fruits of different maturities for harvesting, certain conventional machine learning methods have been utilized in the maturity detection of fruits and vegetables. For example (Zhao et al., 2016), extracted Haar-like characteristics from monochrome images of tomatoes and used an AdaBoost classifier to identify potential tomato targets, but the precision and speed of their identification still require enhancement (Liu et al., 2019). utilized the Gradients of Oriented Histograms descriptor for training the classifier and introduced a coarse-to-fine scanning technique that enhanced the accuracy of tomato detection; however, the algorithm did not consider maturity grading and was limited only to the identification of ripe tomatoes, which constitutes a significant limitation. Recently (Bai et al., 2023), utilized machine learning to aid in image analysis, merging form, surface, and hue characteristics of tomatoes to achieve high-precision identification and picking point localization of clustered tomatoes. However, their model is large and requires high-end hardware. Although traditional machine learning demonstrates significant advantages over manual inspection in identifying the ripeness of tomatoes, it still suffers from issues such as the cumbersome process of manual feature extraction, high model complexity, low detection accuracy, and slow processing speeds.

In unstructured environments, the characterization of tomatoes is further complicated by factors such as variable lighting conditions, occlusion by leaves, and overlapping fruits. These challenges render conventional machine vision algorithms less effective in differentiating the maturity levels, leading to considerable limitations in their applicability. Consequently, deep learning has been employed to address these aforementioned difficulties (Afonso et al., 2020). utilized the Mask-RCNN (He et al., 2017) algorithm to identify tomato images captured in a greenhouse environment, but its ability to distinguish the ripeness of tomatoes was subpar (Ko et al., 2021). have implemented multi-stream convolutional network (Simonyan and Zisserman, 2014) for the detection of tomato ripeness and have employed probabilistic decision integration to achieve more accurate categorization outcomes of ripeness, thereby achieving commendable detection accuracy. Nonetheless, the research pertaining to the real-time identification of tomatoes remains deficient. In recent years, models from the YOLO (Redmon et al., 2016; Redmon and Farhadi, 2017) series have demonstrated exceptional effectiveness within the realm of industrial object recognition, outperforming conventional two-stage detection pipelines (Liu et al., 2020). have substituted conventional rectangular boxes for circular ones for tomato localization, introducing a modified YOLOv3-based detection model that reduces the impact of variations in lighting, overlaps, and obstructions. Nevertheless, this model does not consider ripeness information, which limits its effectiveness in detecting tomatoes across various growth stages (Lawal, 2021). have employed compact architecture integration and pyramid pooling within a modified YOLOv3 framework to enhance the identification accuracy of tomatoes (Zheng et al., 2022). have integrated the ResNet (He et al., 2016) architecture into the CSPDarknet53 backbone of YOLOv4, incorporating depthwise separable convolution blocks as residual edges to execute tomato object detection tasks at three different scales. This has enhanced the robustness under varying degrees of occlusion and lighting conditions. Additionally (Khan et al., 2023), have proposed a novel approach to tomato ripeness classification utilizing a modular convolutional transformer. This method merges the benefits of convolutional networks and transformers (Vaswani et al., 2017), potentially elevating the productivity and correctness in the processes of tomato collection, evaluation, and quality oversight. However, its detection capabilities are limited when dealing with tomatoes that are heavily occluded, cluttered, or small in size.

Deep learning supersedes traditional machine learning in automated feature extraction and handling of high-dimensional data, particularly achieving higher accuracy in image recognition tasks. In the agricultural sector, especially in fruit and vegetable harvesting, there is a growing demand for low-power embedded devices to reduce costs and improve efficiency. Hence, with algorithmic practicality in mind, it is crucial to minimize the footprint and computational demands of the model while concurrently improving the accuracy and velocity of identification. Furthermore, the algorithms are required to possess robustness to withstand interference from unstructured external elements such as changing illumination, climatic conditions, and obstructions due to vegetation. These challenges necessitate the development of innovative deep learning approaches that can effectively balance performance, efficiency, and adaptability to real-world scenarios. This study introduces a lightweight tomato maturity identification approach utilizing an enhanced RT-DETR

framework to tackle the previously mentioned possible issues. Our main contributions are as follows:

1. "Integration of residual blocks with partial convolutional": By amalgamating PConv lightweight convolutions with residual blocks into a novel PConv_Block module, enhanced residual architecture for backbone networks. This integration preserves performance while reducing computational load, thereby elevating the efficiency of feature extraction.
2. "Introduction of deformable attention mechanism": Incorporating the deformable attention mechanism into the encoder of the transformer to replace the multi-head attention mechanism results in the AIFI-DAT component. This enables the framework to grasp intricate associations among various segments of the input, providing enhanced performance in the task of fine-grained classification of tomato ripeness.
3. "Design of a lightweight Neck architecture": The novel slimneck-SSFF structure is proposed by integrating the Scale Sequence Feature Fusion (SSFF) framework with the slim-neck, which incorporates lightweight GSConv and VoVGSCSP modules. Introduced at the Neck stage, this architecture boosts the detection abilities for tiny items while preserving precision, and concurrently reduces computational demand and inference latency.
4. "Loss function optimization": The Inner-IoU is amalgamated with EIoU, introducing an auxiliary bounding box within EIoU controlled by a scale factor ratio to obtain Inner-EIoU. Employing this loss function in place of the original GIoU used by the model yields faster and more efficient regression results.
5. "Evaluation of effectiveness": Thorough assessment with the tomato maturity dataset shows that the proposed PDSI-RTDETR framework outperforms the initial RT-DETR framework regarding accuracy and speed, with reduced computational costs, and outperforms other common object detection models.

2 Materials and methods

2.1 RT-DETR network

RT-DETR is a novel real-time end-to-end target detection model (Lv et al., 2023). Compared to YOLOv8, RT-DETR demonstrates improved efficacy and better equilibrium in the same testing environments, with a shortened training period and without employing the mosaic data augmentation strategy, while maintaining detection speeds on par with the YOLO series. The RT-DETR model is partitioned into three primary segments: the backbone network, the hybrid encoder, and the decoder.

The backbone architecture harnesses the capabilities of a convolutional network to extract salient features, procuring outputs across three distinct scales with strides of 8, 16, and 32. The neck network employs an Attention-based Intra-scale Feature Interaction (AIFI) module to process high-level features from the backbone network, significantly reducing computational load and improving computational speed without compromising performance. It also uses a Cross-scale Feature-Fusion Module (CCFM) for the integration and interaction of multi-scale features. RT-DETR leverages a transformer-based denoising decoder to enhance the quality of bipartite matching samples, accelerating the convergence rate during training. The model dynamically adjusts queries based on IoU, focusing efforts on regions most relevant to the detection targets. The decoder obtains initial object queries from an IoU-aware selection mechanism and iteratively refines them. Overall, RT-DETR presents significant potential for advancement in industrial object detection. The structure of the foundational model RT-DETR selected for this article is shown in Figure 1.

2.2 Dataset collection

The main objective of this research is to ascertain the maturity level of tomatoes cultivated in outdoor settings, thus improving the robustness of our model. To enhance data diversity, images of tomato ripeness were acquired in two batches. The first batch of photographs was taken during the daytime in a tomato picking garden located in Fengtai District, Beijing, China (longitude 116° 12'3.7548"E, latitude 39°47'26.8332"N), using a Xiaomi 9 smartphone equipped with a SONY IMX586 lens (48 MP, f/1.7 aperture, 26mm equivalent focal length). The images were captured under various lighting conditions, ranging from bright sunlight to overcast skies, with temperatures between 25°C and 30°C in early September. The captured images exhibit varying conditions, such as strong light, shadow occlusion, and overlapping fruits. The second batch of images was sourced from 112 tomato images in the publicly available Fruits and Vegetables Image Recognition Dataset (Seth, 2020) on Kaggle. These images underwent data augmentation processing, including mean blur (with kernel sizes ranging from 5×5 to 15×15), random cropping (cropping a random portion of the image within 40-80% of the original image size), and random rotation (rotation by a random angle between -60° and 60°), to enhance the diversity of the dataset. All augmented images were resized to a consistent size of 640x640 pixels. After data augmentation, a total of 3,251 images were obtained. Examples of data augmentation are shown in Figure 2.

To meet the experimental requirements, the dataset was divided into three parts using random splitting. First, the dataset was randomly shuffled to ensure the randomness of the samples. Then, according to the predefined ratios, the shuffled dataset was divided into three subsets: 2,275 images (70%) were used for the training dataset, 325 images (10%) for the validation dataset, and 651 images (20%) were designated for the testing dataset. This random splitting approach effectively assigns the data to different subsets, reducing the bias introduced by the division of the dataset and improving the generalization ability of the model.

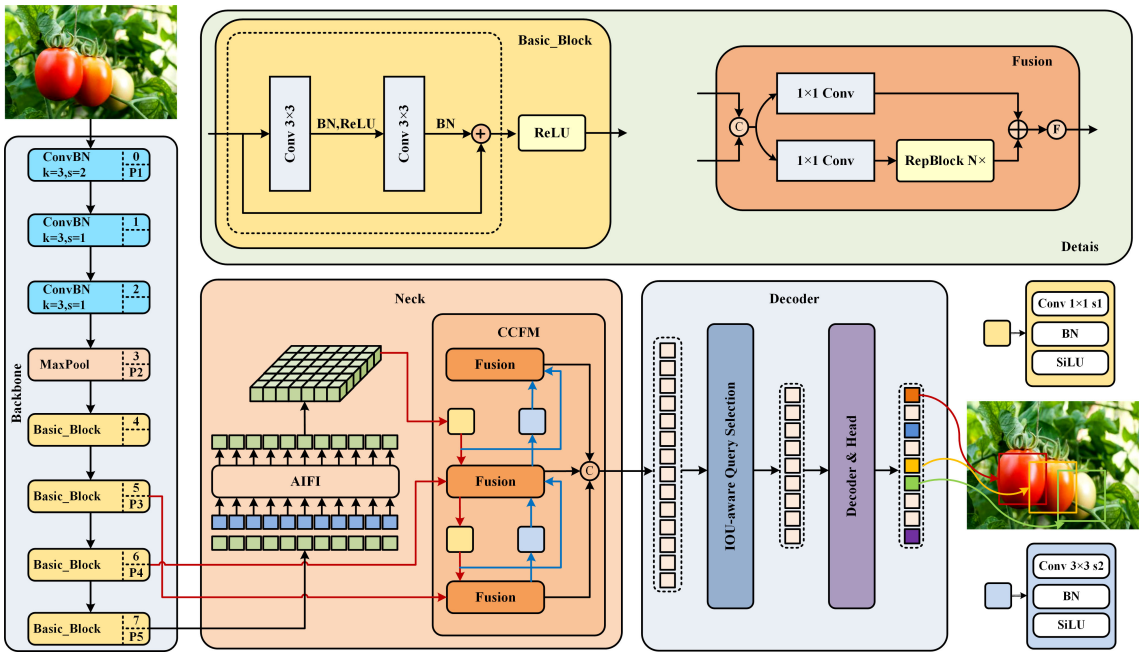


FIGURE 1
RT-DETR network structure diagram.



FIGURE 2
Examples of data augmentation techniques applied to the tomato images.

2.3 Dataset annotation and processing

Given that the principal emphasis of this paper lies in the analysis of computer vision algorithms in assessing tomato ripeness, visual differentiation is performed based on the appearance differences in coloration, hue saturation, size and shape, as well as the degree of spoilage. Based on these characteristics, tomatoes are categorized into five classes: unripe, half-ripe, ripe, overripe, and rotten. Unripe tomatoes exhibit a bright green hue, smaller size, low sugar content, acidic taste, and generally have a total soluble solids TSS content below 5%. Half-ripe tomatoes transition to yellow or pink while retaining green areas, with a sugar content of around 8%, a pH level of approximately 4.25, and a slightly increased TSS of about 7%. Ripe tomatoes are uniformly red or deep red, larger in volume, and vivid in color, with a pH typically ranging between 4.2 and 4.5, and a higher TSS content, likely between 10-12%. Overripe tomatoes are deep red, beginning to lose their luster, showing slight shrinkage or skin relaxation. Rotten tomatoes may develop irregular brown or black spots, lose their normal skin sheen, exhibit significant shrinkage, soften, or even burst. Utilizing the Labelme annotation tool, the tomatoes in the images are manually labeled according to the aforementioned characteristics. Detailed information about the dataset annotation is presented in Table 1 as follows:

2.4 Training and experimental comparison platform

The network experimental environment is based on Ubuntu 20.04, Python 3.10.12, and Pytorch 2.0.1, with relevant hardware configurations and model parameters detailed in Supplementary Table 1. The batch size was set to 4, with a training duration of 100 epochs, and a learning rate selected at 0.0001. An adaptive image size of 640×640 was selected for the experiments.

3 Algorithm design

3.1 The PDSI-RTDETR model architecture

The advent of RT-DETR has filled a void in the DETR (Carion et al., 2020) series for real-time monitoring applications, offering a superior balance of precision and velocity relative to the YOLO series. On the other hand, the detection of tomato ripeness in natural environments necessitates not only high accuracy and rapid processing but also confronts the challenge of model lightweighting to mitigate issues such as hardware costs. This study introduces the lightweight PDSI-RTDETR model to address these concerns. The PConv_Block module, proposed in this paper, merges partial convolutional (PConv) (Chen et al., 2023) with residual blocks, optimizing the backbone network for efficient feature extraction with reduced computational burden. Deformable attention mechanisms (Xia et al., 2022) are introduced to the encoder, enhancing fine-grained classification through the AIFI-DAT module. The proposed slimneck-SSFF structure combines the Scale Sequence Feature Fusion framework (Kang et al., 2023) with a slim-neck design, featuring GSConv and VoVGSCSP modules (Li et al., 2022a), to improve small object detection with reduced computational cost and lower inference latency. Optimization of the loss function is achieved by integrating Inner-IoU (Zhang et al., 2023a) with EIoU (Zhang et al., 2022) to form Inner-EIoU, improving regression efficiency. Comprehensive evaluations confirm that the PDSI-RTDETR model outperforms the baseline RT-DETR and other prevalent object detection methods in accuracy, speed, and computational efficiency. Figure 3 illustrates the PDSI-RTDETR model architecture.

3.2 Backbone network improvement

To circumvent the computational redundancies of complex models in simple tasks, which result in diminished detection velocities, this paper utilizes the comparatively light ResNet-18 (He et al., 2016) as the baseline for the backbone network. Moreover, we substitute the conventional convolutions within the Basic_Block modules with partial convolutions to bolster feature extraction while simultaneously achieving heightened model lightness. Partial convolution judiciously utilizes filters across a chosen subset of input channels, thereby conserving the remainder, which culminates in lower Floating Point Operations Per Second (FLOPs) than standard convolution. This method secures an elevated operational speed on an extensive spectrum of devices without detracting from the precision of the task. The structural principle of PConv is illustrated in Supplementary Figure 1.

The PConv employs conventional convolution on a specific portion of the input channels to derive spatial features, preserving the rest of the channels unchanged. To ensure efficient memory access, the calculation engages the first or last sequence of c_p consecutive channels as a proxy for the computational demand across all feature maps. Maintaining methodological uniformity, it's assumed that the count of channels remains the same for both input

TABLE 1 Tomato ripeness dataset labeling information.

Types	Number	half_ripe	over_ripe	ripe	rotten	unripe
Training	2275	721	1285	2417	980	1408
Validation	325	146	237	337	101	222
Test	651	216	328	627	265	489
Total	3251	1083	1850	3381	1346	2119

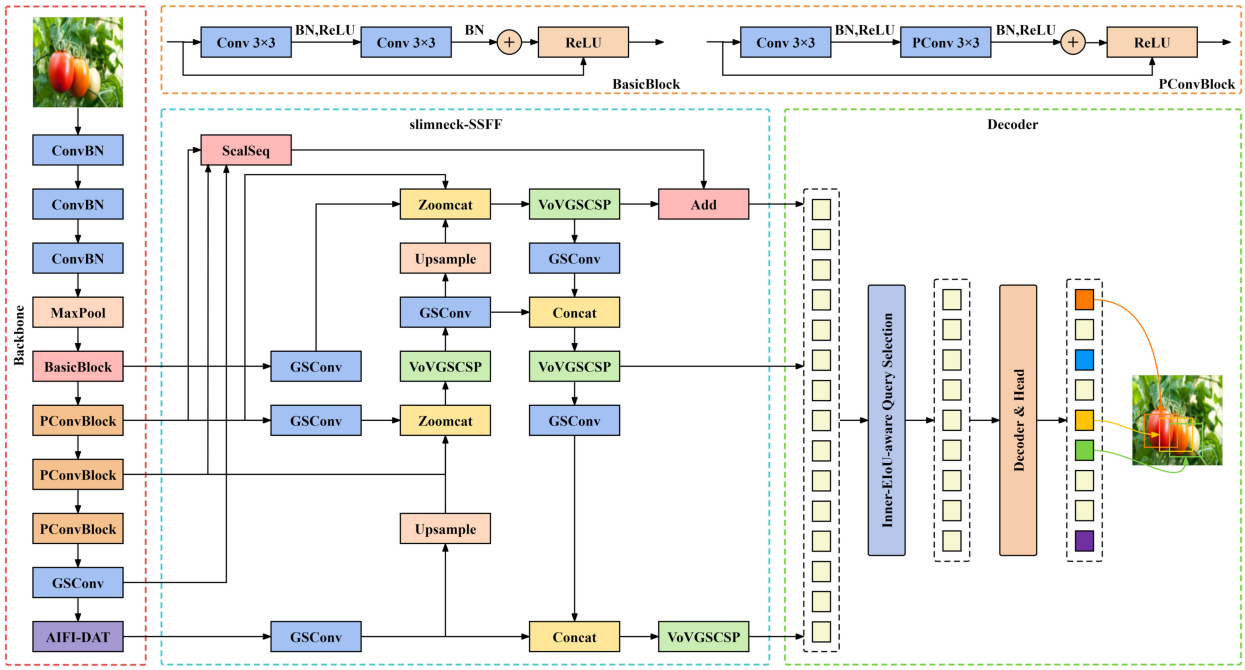


FIGURE 3
The structure diagram of improved RT-DETR(PDSI-RTDETR).

and output feature maps. Therefore, the FLOPs of a PConv are only, as detailed in Equation (1):

$$h \times w \times k^2 \times c_p^2 \tag{1}$$

Within the formula, h and w denote the dimensions of the feature map, k indicates the size of the convolution kernel, and c_p is the count of channels utilized by the standard convolution. Typically, with $r = c_p/c = 1/4$, the FLOPs for PConv are just $1/16$ of a regular Conv. Additionally, the scenario of diminished memory access for PConv includes, as specified in Equation (2):

$$h \times w \times 2c_p + k^2 \times c_p^2 \approx h \times w \times 2c_p \tag{2}$$

Memory access for PConv is merely a quarter compared to a typical convolution, as the remaining $c \sim c_p$ channels do not participate in the computation and therefore do not require memory access. Incorporating PConv into the feature extraction network markedly decreases computational and memory demands, streamlining the backbone network and enhancing model inference speed. The backbone network incorporating the proposed PConv_Block module is shown in Figure 4.

3.3 Improvements to the efficient hybrid encoder

3.3.1 Deformable attention

The deformable self-attention module determines the locations of key and value pairs inside the self-attention framework based on the data. This method facilitates targeted attention on significant areas, enhancing the extraction of meaningful features. Additionally, it addresses the issue of excessive memory and computational costs associated with dense attention. Deformable attention shares shifted keys and values for each query, reducing spatial complexity and avoiding significant information loss that can result from down-sampling techniques. An illustration of deformable attention is presented in Figure 5.

As shown above, for an input feature map $x \in \mathbb{R}^{H \times W \times C}$, a uniform grid of points $p \in \mathbb{R}^{H_G \times W_G \times 2}$ is established as a baseline. To calculate the offsets for each grid point, the feature map undergoes a linear transformation into query tokens $q = xW_q$, which are then processed by a specialized lightweight network θ_{offset} to produce the offsets $\Delta p = \theta_{offset}(q)$. Following this, sampling of features occurs at these adjusted points to form keys and values, culminating in the creation of the projection matrix, as described in Equations (3) and (4):

$$q = xW_q, \tilde{k} = \tilde{x}W_k, \tilde{v} = \tilde{x}W_v \tag{3}$$

$$\Delta p = \theta_{offset}(q), \tilde{x} = \varnothing(x; p + \Delta p) \tag{4}$$

$\tilde{k}$ and $\tilde{v}$ respectively represent the deformed key and value embeddings, and are transformed into bilinear interpolation using the sampling function $\varnothing(\cdot;\cdot)$, as defined in Equation (5):

$$\varnothing\left(z; \left(p_x, p_y\right)\right) = \sum_{(r_x, r_y)} g(p_x, r_x) g(p_y, r_y) z[r_y, r_x, :] \tag{5}$$

where $g(a, b)$ and the coordinates (r_x, r_y) span every position on $z \in \mathbb{R}^{H \times W \times C}$. Multi-head attention is performed on q, k, v, incorporating relative positional displacements R. The result from the attention heads is articulated as follows in Equation (6):

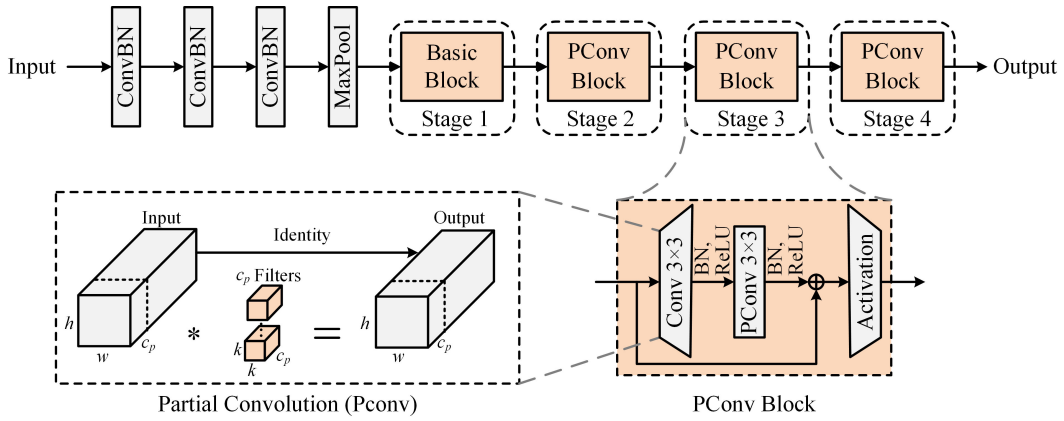


FIGURE 4
Lightweight feature extraction backbone network structure incorporating PConv.

$$Z^{(m)} = \sigma(q^{(m)}\tilde{k}^{(m)T}/\sqrt{d} + \varnothing(\hat{B}; R))\tilde{v}^{(m)} \quad (6)$$

where $\sigma(\cdot)$ represents the softmax operation and d is the dimensionality of each attention head. $Z^{(m)}$ signifies the output embedding from the m-th attention head. The outputs from all heads are merged via concatenation and then transformed through W_o to produce the final output z, as described in Equation (7):

$$Z = Concat(z^{(1)}, \cdots, z^{(M)})W_o \quad (7)$$

3.3.2 Introduce deformable attention into the AIFI module

Within the neck architecture of the model, a single transformer encoder layer is dedicated to processing the S_5 features emanating from the backbone network. Leveraging the rich semantic attributes of high-level features, this methodology significantly curtails computational demands and augments processing speed without sacrificing performance robustness. This optimized hybrid encoder orchestrates intra-scale feature interaction, morphing multi-scale features into a serialized array of image feature sequences. The replacement of the conventional multi-head self-attention with deformable attention facilitates adaptive sampling of pivotal feature locations, diminishing memory usage and circumventing the severe information loss inherent to down-sampling techniques, thereby elevating the computational efficiency and prowess in feature capture of the model. The computational process, as delineated in Equations (8) and (9), is as follows:

$$Q = K = V = FlDatten(S_5) \quad (8)$$

$$F_5 = Reshape(DAttn(Q, K, V)) \quad (9)$$

where *DAttn* represents the deformable attention, and *Reshape* represents the restoration of the feature's shape to match that of S_5, which is the inverse operation of *FlDatten*.

3.4 Improved neck

The role of the neck network in the model is to coordinate and bolster the feature representation at different levels in order to enhance the precision in identifying targets across various sizes. The neck network of the RT-DETR model uses the AIFI module to

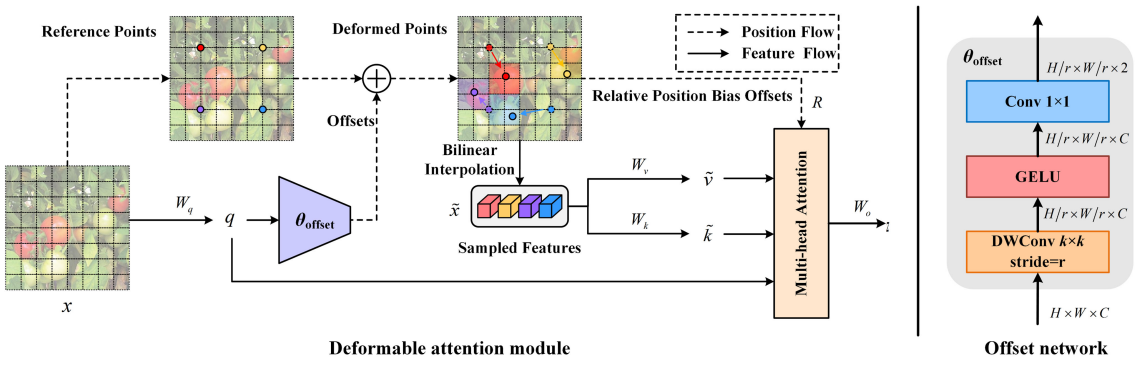


FIGURE 5
Deformable attention mechanism structure, which adjusts reference points and applies relative position bias through an offset network to optimize feature transformation.

process the high-level features, and then the CCFM module is used for the interaction and fusion of multi-scale features. Its compared with YOLO the number of parameters and computation of this neck structure of the network has risen, and considering the necessity to detect a significant volume of targets within an individual image and the large color difference between different ripeness of tomatoes, the original model loses the small target information in the process of convolution and down-sampling. Therefore, this paper introduces the SSFF module, GSConv and Slim-neck technology, and proposes the slimneck-SSFF feature fusion architecture, which reduces the complexity and computation of the model on the basis of improving the accuracy.

As shown in Figure 6, the GSConv module combines conventional convolution with separable convolution and uses the Shuffle procedure to integrate the features generated by both, ensuring inter-channel information exchange while effectively reducing computational costs.

As illustrated in Figures 6B, C, GSbottleneck consists of two GSConv modules and one DWConv module, with input features fed into each module and their outputs summed. Based on GSbottleneck, VoVGSCSP is constructed using a one-off aggregation approach, effectively decreasing the count of parameters and floating-point computations.

To distinguish between the ripeness stages of tomatoes across different sizes, we employ the SSFF module to boost the network proficiency in capturing scale-diverse features. The structure of the SSFF module is depicted in Figure 6D. The SSFF module treats feature maps of different sizes as a scale space, adjusts the effective feature maps of different resolutions to the same resolution for concatenation, then horizontally stacks the feature maps of different scales, and utilizes three-dimensional convolution to extract their scale sequence features. In contrast to existing literature that merely adopts summation or concatenation methods to fuse pyramid features, the SSFF module can better integrate the high-dimensional information from deep feature maps with the detailed information from shallow feature maps. This provides more comprehensive and refined feature descriptions for objects of different dimensions, thereby enhancing the network's ability to capture multi-scale features.

3.5 The improved loss function

In the research conducted, the original GIoU of the model is substituted with Inner-EIoU, which offers quicker convergence, enhanced accuracy in assessment, and supplementary edges. The use of smaller auxiliary borders to compute the loss during the model training process has a gain effect on the regression of the high IoU samples, and the opposite is true for the low IoU samples. Employing the scale factor ratio to manage the creation of various scales of auxiliary edges for loss calculation yields quicker regression outcomes across diverse scenarios. Applying Inner-IOU to the EIOU loss function is calculated as follows, according to Equations (10)–(14):

$$inter = (\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)) * (\min(b_b^{gt}, b_b) - \max(b_t^{gt}, b_t)) \tag{10}$$

$$union = (w^{gt} * h^{gt}) * (ratio)^2 + (w * h) * (ratio)^2 - inter \tag{11}$$

$$IoU^{inner} = \frac{inter}{union} \tag{12}$$

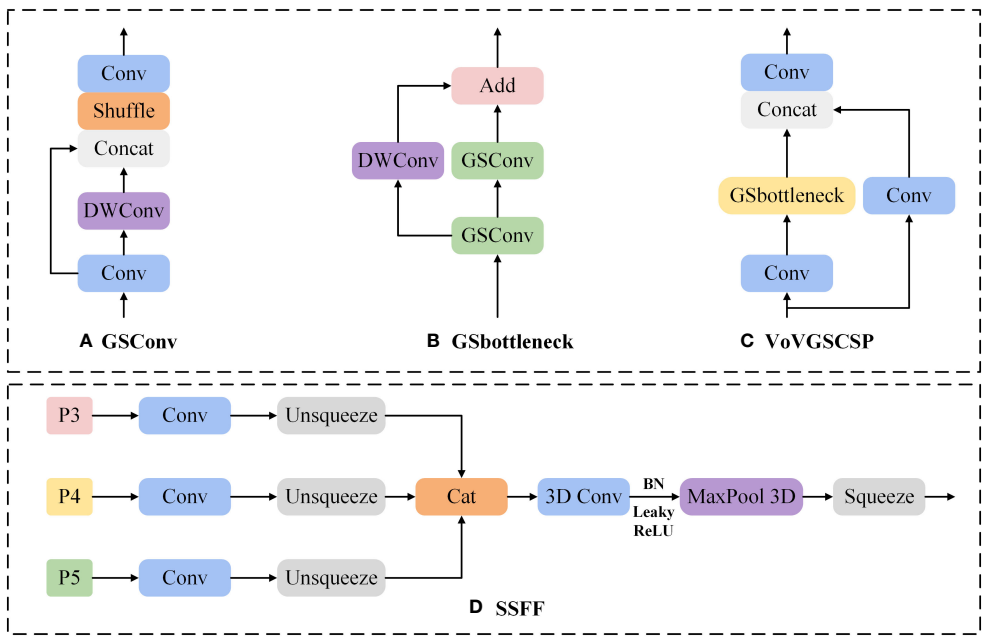


FIGURE 6
(A) GSConv, **(B)** GSbottleneck, **(C)** VoVGSCSPC, and **(D)** SSFF module structure diagrams.

$$\mathcal{L}_{EIoU} = \mathcal{L}_{Iou} + \mathcal{L}_{dis} + \mathcal{L}_{asp} \tag{13}$$

$$\mathcal{L}_{Inner-EIoU} = \mathcal{L}_{EIoU} + IoU - IoU^{inner} \tag{14}$$

where the ground truth (GT) box and anchor are respectively represented as B^{gt} and B. The width and height of the GT box are denoted by w^{gt} and h^{gt}, while the width and height of the anchor box are represented by w and h. The *ratio* is an auxiliary factor that controls the size of the helper box. The EIOU loss function is composed of three components: the overlap loss $\mathcal{L}_{Iou}$, the center distance loss $\mathcal{L}_{dis}$, and the width and height loss $\mathcal{L}_{asp}$. Inner-EIoU, focusing more accurately on the box's center, is an enhanced bounding box regression loss that quickens convergence through loss calculation with scale-modified support boxes and shows efficacy in detecting smaller objects.

4 Evaluation indicators

The paper assesses the algorithmic performance by comparing the disparities in image detection by the network model before and after enhancements, under identical experimental settings. The study employs precision (P), recall (R), mean average precision (mAP), F1 score, GFLOPs, and frames per second (FPS) as evaluative criteria.

Precision represents the proportion of true positives in the predictions classified as positive by the model. It gauges the adeptness of the model in distinguishing negative samples. Higher precision denotes greater reliability of the model in predicting positive cases. The formula for precision is presented as in Equation (15).

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

Recall quantifies the fraction of true positive samples accurately detected as positive by the model, showcasing the proficiency in identifying positive instances. An increased recall suggests the model identifies more true positives. The formula for calculating recall is presented in Equation (16).

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

The F1 score, serving as the harmonic mean between precision and recall, aims to consolidate these metrics into one indicator. This score varies between 0 and 1, where scores nearing 1 denote a stronger model. The calculation for the F1 score is detailed in Equation (17).

$$F1 = \left(\frac{2}{Recall^{-1} + Precision^{-1}}\right) = 2\frac{Precision \cdot Recall}{Precision + Recall} \tag{17}$$

The Average Precision (AP) measures the mean highest precision at various recall levels per category. The mean Average Precision (mAP) averages these APs across categories, assessing object detection models' overall performance. It is a comprehensive metric suitable for multi-category object detection tasks, as illustrated in the following formula, Equations (18) and (19).

$$AP = \int_0^1 P(r)dr \tag{18}$$

$$mAP = \frac{\sum_{j=1}^{S} AP(j)}{S} \tag{19}$$

where S denotes the total category count, and the divisor is the aggregate of AP values across all categories.

Moreover, GFLOPs measures the computational complexity, allowing for efficiency comparisons among models. FPS, indicating images processed per second, assesses real-time performance. The formula involving t_{avg} represents FPS calculation, as shown in Equation (20).

$$FPS = \frac{1}{t_{avg}} \tag{20}$$

5 Results and analysis

5.1 Backbone network comparative experiment

The backbone network of the RT-DETR model, employing ResNet-18 for feature extraction, encompasses four Basic_Block modules. To explore the appropriate position for enhancing the backbone network architecture, the introduced PConv_Block module is utilized to substitute each Basic_Block module, and the performance of the enhanced model is examined.

As indicated in Table 2, the introduction of the PConv_Block module leads to a reduction in the computational cost of the model, while the highest mAP50 also increases by 1.3%. However, within the backbone network structure, it has been found that the number of PConv usages is not the more the better, after all 4 modules are improved mAP50 instead decreased by 0.5%, and experiments found that it is more effective to use it in the middle feature layer. This may be due to PConv requiring both location and semantic information, which are relatively balanced in the middle feature layers. Based on the experimental results, this paper replaces the last three Block modules in the backbone of the baseline model with PConv_Block modules, lightening the model structure while improving detection accuracy.

Furthermore, the challenge in accurately detecting tomato ripeness involves minimizing computational demands while enhancing precision and speed. To corroborate the efficiency of employing partial convolution, several leading convolutional networks have been chosen for benchmarking in comparative studies, as illustrated in Table 3.

As can be seen from Table 3, compared with the original model, the introduction of lightweight PConv_ Block not only significantly reduces the model parameters and computational cost, but also achieves the improvement of P, R, and mAP, of which the mAP50 obtains a sizable improvement of 1.3 percentage points. The integration of Deformable Convolution v2 (Zhu et al., 2019) has resulted in an increase of 1.4% in mAP50, thus yielding enhanced accuracy; however, this enhancement comes at the cost of a 2.2%

TABLE 2 Comparison of model performance after improving the backbone network.

Model	P(%)	R(%)	mAP50(%)	mAP50:95(%)	GFLOPs(G)
RT-DETR	81.7	78.7	82.9	69.8	57.3
RT-DERT+1 PConv_Block	84.2	77.9	82.5	69.3	53.8
RT-DERT+2 PConv_Block	84.8	78.5	84.1	70.5	50.2
RT-DERT+3 PConv_Block	85.1	80.9	84.2	71.3	46.7
RT-DERT+4 PConv_Block	82.9	77.6	82.4	68.8	43.2

increase in the parameter count of the model. Additionally, the current study has incorporated DySnakeConv (Qi et al., 2023) into the original Basic_Block module. While DySnakeConv demonstrates impressive performance on the mAP50 metric, improving by 1.1% over the original model, it has substantially increased both the parameter count and computational demand of the model, with GFLOPs experiencing a 6.8% surge. In the final analysis, compared to the original model, the backbone integration of DualConv (Zhong et al., 2022) and AKConv (Zhang et al., 2023b), despite reducing both parameters and computational load, results in respective reductions of 0.3% and 2.3% in mAP50:95, which does not satisfy the high-precision requirements for tomato ripeness detection. Upon comparison of the performance of several prevalent operators, the PConv_Block proposed in this article exhibits exceptional performance.

5.2 Verifying the role of AIFI-DAT module

To confirm and explore the factors contributing to enhanced tomato ripeness detection performance through the integration of the intra-scale feature interaction module with deformable attention, this study utilized XGrad-CAM (Fu et al., 2020) for generating and contrasting heatmaps pre and post the incorporation of the attention mechanism. This comparative evaluation is depicted in Figure 7.

From the above figures, it can be observed that the RT-DETR model without the AIFI-DAT module neglects certain tomato targets, while attending to irrelevant locations such as tree leaves. Upon integration of the AIFI-DAT module, the focus of the model becomes more concentrated and accurate, mitigating attention towards interfering elements like foliage. Additionally, it enables precise identification of the ripeness of smaller tomatoes.

5.3 Verifying the effectiveness of the slimneck-SSFF structure

As seen in the ablation experiments in Table 4, by comparing four different configurations of the model, the specific effects of various neck network improvement strategies on the model performance can be observed. Model 2, which introduces slimneck, enhances the precision and recall by 1.2% and 2.5%, in comparison with the RT-DETR model, while the calculation complexity and the parameter count are diminished to 53.3 GFLOPs and 19.30 M, illustrating the significant effect of the slimneck strategy in boosting the computational efficiency of the model. Conversely, Model 3, employing the SSFF strategy, markedly improves the mAP50 and mAP50:95 by 2.1% and 3.1%, yet modestly elevates the computational complexity and the number of parameters, highlighting the additional resource requirement in aiming for higher detection accuracy. Ultimately, Model 4, merging both slimneck and SSFF strategies, attains equilibrium, raising the precision and recall to 85.4% and 82.4% while enhancing mAP50 and mAP50:95 by 2.1% and 2.6%, and trimming the number of parameters by 1.8%, thus achieving the optimization objective of managing the consumption of computational resources while preserving high performance.

5.4 Validity of loss function improvements

To substantiate the effectiveness of the proposed Inner-EIoU loss function, experiments were conducted by selecting different ratio values to alter the size of the auxiliary bounding box. Comparative analyses were performed against established loss functions such as GIoU, SIoU, EIoU, and Shape-IoU (Rezatofighi et al., 2019; Gevorgyan, 2022; Zhang et al., 2022; Zhang and

TABLE 3 Comparative experiment results of different backbone networks.

Backbone network	P(%)	R(%)	mAP50(%)	mAP50:95(%)	Params(M)	GFLOPs(G)
Basic_Block	81.7	78.7	82.9	69.8	19.97	57.3
DualConv_Block	84.3	80.1	83.6	69.5	16.02	50.2
AKConv_Block	82.2	77.6	81.5	67.5	15.44	49.4
DCNv2_Block	81.8	81.9	84.3	70.5	20.41	47.4
DySnakeConv_Block	83.8	80.6	84.0	68.2	27.96	61.2
PConv_Block	85.1	80.9	84.2	71.3	14.17	46.7

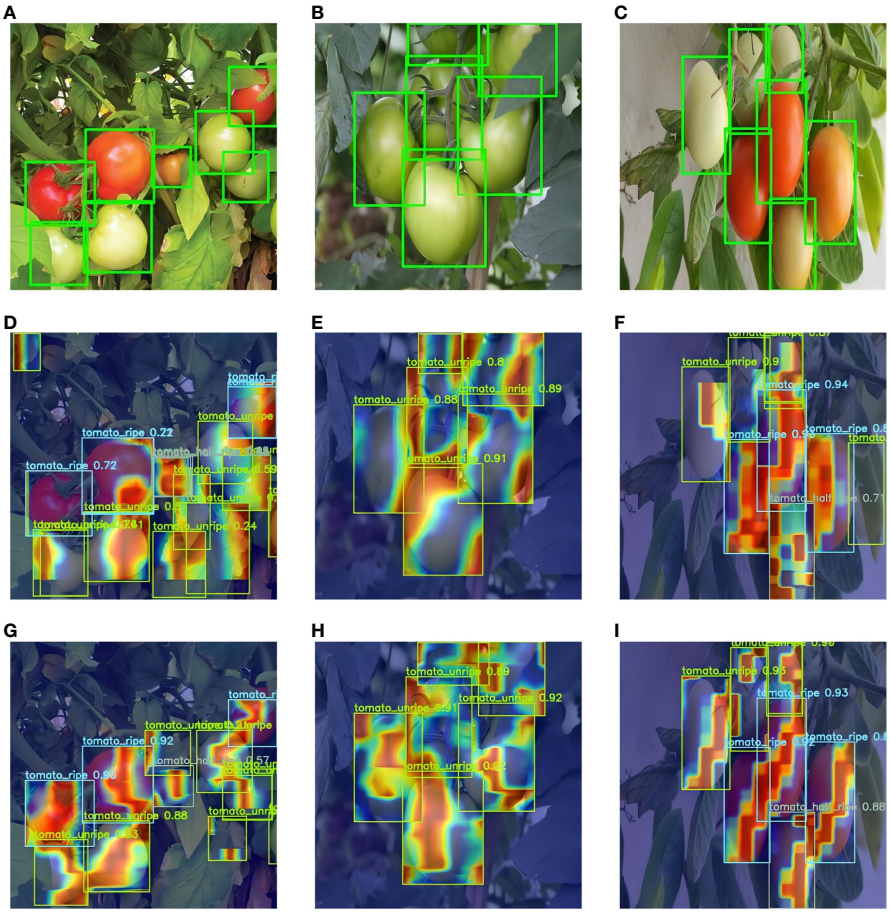


FIGURE 7
Comparison of feature visualization before and after adding the AIFI-DAT module. **(A–C)** image correctly labeled boxes, **(D–F)** feature heatmaps of the original RT-DETR, and **(G–I)** feature heatmaps with the addition of the AIFI-DAT module.

Zhang, 2023). According to the data illustrated in Table 5, the model incorporating the Inner-EIoU loss function with a ratio of 0.7 manifested superior detection precision. Relative to the baseline model employing the GIoU loss function, there was an enhancement of 0.9% in the metric mAP50 and a 0.7% increment in mAP50:95. This suggests that the use of the Inner-EIoU loss function can lead to more stable regression on bounding boxes and higher prediction accuracy.

In models for detecting objects, the goal is to accurately determine the position and dimensions of objects with maximum precision. Therefore, a lower bounding box loss indicates a smaller discrepancy between the predicted bounding boxes and the true annotations, signifying enhanced model performance. From Supplementary Figure 2, it is evident that the Inner-EIoU markedly outperforms other loss functions, facilitating accelerated descent velocity and reducing the convergence time of the model.

5.5 Ablation study

To verify the enhancement effect of the proposed improvement modules on the model, eight sets of ablation experiments were designed. On the basis of the RE-DETR network, the following modifications were made: the Basic_Block in the feature extraction

TABLE 4 Performance comparison of models after neck improvements.

Model	P (%)	R (%)	mAP50 (%)	mAP50:95 (%)	GFLOPs (G)	Params (M)
1.RT-DETR	81.7	78.7	82.9	69.8	57.3	19.97
2.RT-DERT+slimneck	82.9	81.2	83.9	70.3	53.3	19.30
3.RT-DERT+SSFF	84.4	81.6	85.0	72.9	61.5	20.16
4.RT-DERT+slimneck+SSFF	85.4	82.4	85.0	72.4	57.6	19.61

TABLE 5 Performance comparison of models with improved loss functions.

Loss function	P(%)	R(%)	mAP50(%)	mAP50:95(%)
GIoU	86.3	82.8	85.9	73.2
SIoU	85.0	83.5	85.5	72.8
Shape-IoU	86.7	82.4	85.5	73.0
EIoU	84.2	83.8	86.2	73.7
Inner-EIoU(ratio=0.70)	86.9	84.1	86.8	73.9
Inner-EIoU(ratio=0.75)	85.4	81.1	84.6	70.6
Inner-EIoU(ratio=0.80)	86.0	82.6	85.9	73.3
Inner-EIoU(ratio=1.10)	86.1	83.1	85.4	73.0
Inner-EIoU(ratio=1.13)	86.0	82.9	85.5	73.1
Inner-EIoU(ratio=1.15)	86.7	83.4	85.7	73.2

network was replaced with the lightweight PConv_Block module, the intra-scale feature interaction module equipped with a deformable attention mechanism was added, and the feature fusion network was optimized using slimneck and SSFF. Additionally, the loss function was switched to Inner-EIoU. The experiments were conducted by successively adding each improvement module, and the results are presented in Table 6.

The ablation study results presented in Table 6 elucidate that substituting the backbone with a module incorporating lightweight partial convolution engenders enhancements of 1.3%, 1.5%, and 22.5% across the mAP50, mAP50:95, and FPS metrics respectively. This evidences that refining the backbone network concurrently elevates accuracy and expedites detection velocity, while also effectuating a diminution in computational load and parameter count by 18.5% and 29.1%, respectively. The incorporation of the AIFI-DAT layer into the improved model, relative to its predecessor, yielded ascensions of 0.9% and 14.7% in mAP50 and FPS, respectively. Moreover, the integration of a deformable attention mechanism slightly augmented the computational and parameter requisites of the model, yet significantly bolstered its performance in discerning tomatoes of disparate shapes. The amalgamation of the slimneck-SSFF architecture within the neck network substantially uplifted mAP50 and mAP50:95 by 2.1% and 2.6%, correspondingly, while maintaining stability in other metrics. This attests to the efficacious enhancement of the fusion and articulation of tomato feature information by the slimneck-SSFF framework. Subsequent to adopting the Inner-EIoU loss function, the model manifested incremental enhancements of 0.9%, 1.4%, and 7.5% in mAP50, mAP95, and FPS, respectively.

Additionally, a comparison between Experiment 1 and Experiment 6 reveals that the model integrating the PConv_Block module and AIFI-DAT layer exhibited enhancements of 1.9%, 2.2%, and 22.3% in mAP50, mAP50:95, and FPS, respectively, alongside a substantial reduction in computational and parameter volume. Further analysis of Experiments 1, 6, and 7 shows that, following the incorporation of the slimneck-SSFF architecture in Experiment 7, there was a marginal increase in computational demand compared to Experiment 6, yet the precision and speed of the model were augmented. Finally, a comparison between Experiments 1 and 8 demonstrates that after implementing four model improvements, the proposed PDSI-RTDETR model, relative to the baseline model, achieved respective increases in mAP50, mAP50:95, and FPS metrics by 3.9%, 4.1%, and 38.7%, with a decrease of 17.6% in GFLOPs and a reduction of 6.16M in

TABLE 6 Results of ablation experiments.

Methods	PConv_ Block	AIFI-DAT	slimneck-SSFF	Inner-EIoU	mAP50 (%)	mAP50:95 (%)	GFLOPs (G)	Params (M)	FPS (f/s)
1.base	—	—	—	—	82.9	69.8	57.3	19.97	78.3
2	√	—	—	—	84.2	71.3	46.7	14.17	95.9
3	—	√	—	—	83.8	71.1	57.5	19.98	89.8
4	—	—	√	—	85.0	72.4	57.6	19.61	79.9
5	—	—	—	√	83.8	71.2	57.0	19.88	84.2
6	√	√	—	—	84.8	72.0	46.9	14.17	95.8
7	√	√	√	—	85.9	73.2	47.2	13.81	106.9
8.ours	√	√	√	√	86.8	73.9	47.2	13.81	108.6

Where "√" symbol represents the improvement of the structure of the ordinate corresponding to this symbol.

parameter count. A comprehensive analysis of Table 6, the results of the ablation study substantiate the efficacy of each proposed improvement module.

To more explicitly showcase the superiority of the proposed PDSI-RTDETR model, the training curves for mAP50 and mAP50:95 of both the original model and the improved model were visualized and compared, with blue representing the RT-DETR model and yellow denoting the PDSI-RTDETR model, as illustrated in Figure 8. It is evident from the graph that, across various epochs of training, the enhanced model consistently surpasses the original model in both mAP50 and mAP95 metrics, demonstrating a heightened precision and improved effectiveness in detecting the ripeness of tomatoes.

5.6 Comprehensive analysis of the improved model

Table 7 delineates the performance of the enhanced model in detecting tomatoes at various stages of ripeness. Examining the overall performance, the model exhibits high precision and recall rates of 86.9% and 84.1%, respectively, achieving an mAP50 of 86.8% and an mAP50:95 of 73.9%, thereby affirming its effectiveness in adjudicating the ripeness of tomatoes comprehensively. Notably, in the detection of overripe and rotten tomatoes, the model demonstrates exceptionally high precision rates of 96.1% and 87.7%, respectively, along with superior scores in mAP50, reaching 93.4% and 89.7%, respectively. This signifies the robust capability of the model to recognize prominent feature alterations during the ripening process, such as intensified coloration and structural changes in the fruit. However, in identifying tomatoes at the half-ripe stage, the precision and recall rates of the model are marginally lower, with notably diminished mAP50 and mAP50:95 scores of 67.4% and 70.4%, respectively. This can be attributed to the subtler feature transitions characteristic of this ripening phase, presenting challenges to the detection capabilities of the model. Although the precision in detecting unripe tomatoes is the lowest at 80.6%, a higher recall rate of 90.1% suggests the efficiency of the model in not overlooking unripe tomatoes. Overall, the PDSI-RTDETR model demonstrates exemplary performance in tomato ripeness detection, particularly distinguishing itself in identifying overripe and rotten tomatoes, showcasing significant advantages.

To visually demonstrate the capability of the model in recognizing tomatoes at varying stages of ripeness, we visualized several evaluation metrics, as depicted in Figure 9. The comparison of performance curves for tomatoes of each ripeness category distinctly illustrates the detection abilities of the improved model across different maturation phases. Within Figure 9 respectively represent the precision, recall, mAP50, and F1 comparison curves for tomatoes at each level of ripeness.

Figure 10 presents the confusion matrices for the RT-DETR and PDSI-RTDETR models at their respective optimal performances. From the comparative visualization, it is intuitively observable that the PDSI-RTDETR model surpasses the original model in classification performance across all categories. Based on the analysis provided, the PDSI-RTDETR model proposed in this paper demonstrates exceptional performance in the task of tomato ripeness detection.

5.7 Comparison of different detection models

The performance of PDSI-RTDETR was compared with several other object detection models, including Faster-RCNN (Ren et al., 2016), SSD (Liu et al., 2016), YOLOv5, YOLOv6 (Li et al., 2022b), YOLOv8, and the RT-DETR series.

As indicated in Table 8, compared to Faster-RCNN, SSD, YOLOv5, YOLOv6, and YOLOv8, the mean average precision with IoU of 0.5 was found to be higher by 10.0%, 7.9%, 3.5%, 1.7%, and 1.4%, respectively. Furthermore, the detection time was recorded at 4.2 milliseconds per image, satisfying the criteria for

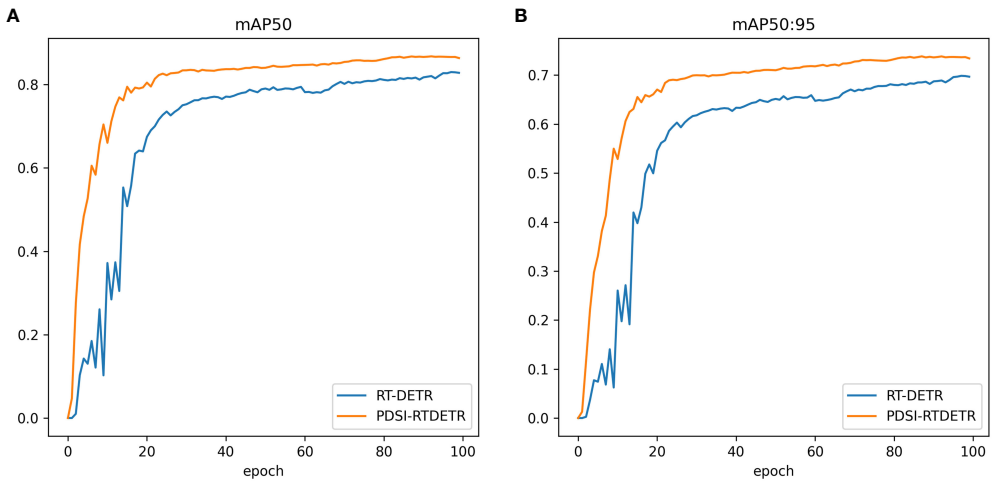


FIGURE 8
Comparison of mean Average Precision between RT-DETR and PDSI-RTDETR. **(A)** mAP50 comparison curve, **(B)** mAP50:95 comparison curve.

TABLE 7 Performance of the improved model in detecting tomatoes of different ripeness levels.

Class	Instances	P(%)	R(%)	mAP50(%)	mAP50:95(%)
all	1925	86.9	84.1	86.8	73.9
tomato_half_ripe	216	85.3	67.4	70.4	60.0
tomato_overripe	328	96.1	92.8	93.4	81.5
tomato_ripe	627	85.0	84.1	88.7	72.9
tomato_rotten	265	87.7	86.1	89.7	81.0
tomato_unripe	489	80.6	90.1	91.6	74.1

real-time detection. Additionally, in comparison to the RT-DETR-L, RT-DETR-R34, and RT-DETR-R50 models, the precision of the proposed model was enhanced by 1.8%, 5.6%, and 3.8%, respectively. Compared to the original RT-DETR model, the mAP50 and mAP50:95 were increased by 3.9% and 4.1%, respectively, while F1 score increased by 5%. Consequently, in comparison to other object detection networks, PDSI-RTDETR demonstrated significant improvements in accuracy and speed for the detection of tomato ripeness within natural environments.

The PDSI-RTDETR model introduced in this study outshines the baseline model in detecting tomato ripeness and surpasses several prevalent models for object detection, as shown in Supplementary Figure 3. An increase of 3.9% in mAP50 is noted when compared with the baseline RT-DETR model, alongside a decrease in image processing time by 4.2 milliseconds. These findings jointly suggest enhancements in both precision and processing efficiency by the model.

5.8 Visual analysis

In natural settings, tomatoes are subjected to a variety of lighting conditions, including uneven illumination, potential obstructions by leaves or branches, and even instances of fruit overlap. To address these challenges, we evaluated the performance of our enhanced model across diverse scenarios. As illustrated in Figure 11 , we present examples of tomato ripeness detection under various conditions. These findings demonstrate the exceptional robustness of the model in managing changes in lighting and in identifying overlapping fruits.

Figure 12 displays the comparative results of tomato ripeness prediction between the PDSI-RTDETR and the original RT-DETR models under varying environmental conditions.

Through these images, we can clearly observe that PDSI-RTDETR surpasses the original model in detection accuracy. Notably, a comparison presented in Figure 12 vividly showcases

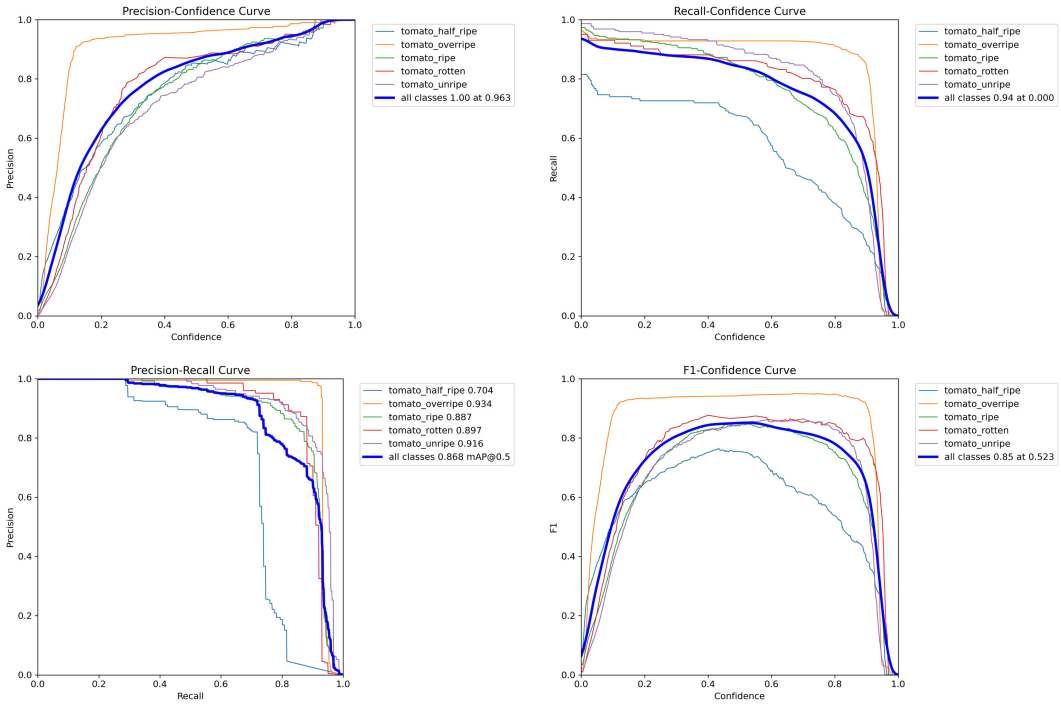


FIGURE 9
The precision, recall, mAP50, and F1 comparison curves for tomatoes at each ripeness level using the PDSI-RTDETR model.

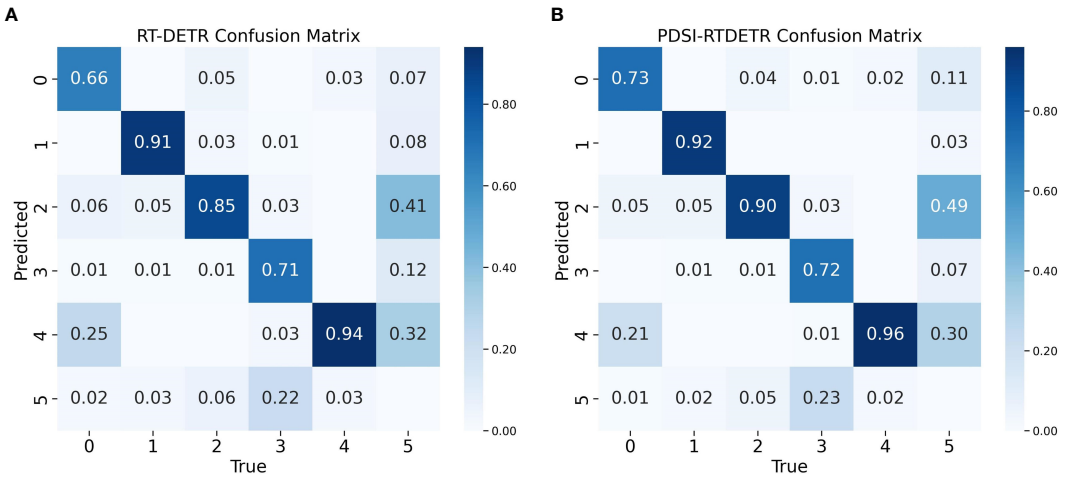


FIGURE 10
Comparison of model confusion matrix before and after improvement. **(A)** Confusion matrix for RT-DETR, **(B)** Confusion matrix for PDSI-RTDETR.

the ability of the enhanced model to accurately distinguish the ripeness of tomatoes in dense areas. Moreover, Figure 12D reveal that the original model mistakenly identified some leaves as unripe tomatoes, and Figure 12F shows the misidentification of rotten tomatoes as unripe ones. In contrast, the improved PDSI-RTDETR model correctly detected the conditions, as shown in Figures 12G–I.

6 Discussion

The experimental results demonstrate that the proposed PDSI-RTDETR model achieves significant improvements in both accuracy and efficiency for tomato ripeness detection in natural environments. The ablation studies confirm the effectiveness of each proposed improvement module, including the lightweight PConv_Block, the AIFI-DAT module, the slimneck-SSFF structure, and the Inner-EIoU loss function.

The introduction of the PConv_Block module in the backbone network not only reduces the computational cost and parameter count but also improves the detection accuracy, which can be attributed to the ability of partial convolution to extract more discriminative features by focusing on informative regions while suppressing irrelevant background information. The AIFI-DAT module further enhances the feature representation by capturing multi-scale contextual information and adaptively adjusting the receptive field based on the object scale, as revealed by the visualization of attention maps using XGrad-CAM, which demonstrates that the AIFI-DAT module enables the model to focus more on the target tomatoes while suppressing the interference from background elements such as leaves. In the neck network, the slimneck-SSFF structure effectively balances the trade-off between accuracy and efficiency, with the slimneck strategy reducing the computational complexity and parameter count, while the SSFF strategy improves the feature fusion and

TABLE 8 Comparison of performance of different models.

Model	P(%)	R(%)	mAP50(%)	mAP50:95(%)	F1	Time(ms)
Faster R-CNN	76.8	75.7	76.8	62.5	0.76	4.1
SSD	79.2	75.8	78.9	64.5	0.77	5.9
YOLOv5	82.3	80.2	83.3	70.2	0.81	8.2
YOLOv6	85.8	81.8	85.1	69.6	0.83	5.6
YOLOv8	84.3	81.2	85.4	71.4	0.83	7.9
RT-DETR	81.7	78.7	82.9	69.8	0.80	8.5
RT-DETR-L	84.6	81.8	85.0	71.7	0.83	5.4
RT-DETR-R34	80.8	76.9	80.6	66.2	0.79	4.4
RT-DETR-R50	82.6	81.0	84.2	69.9	0.82	8.2
PDSI-RTDETR	86.4	84.1	86.8	73.9	0.85	4.2

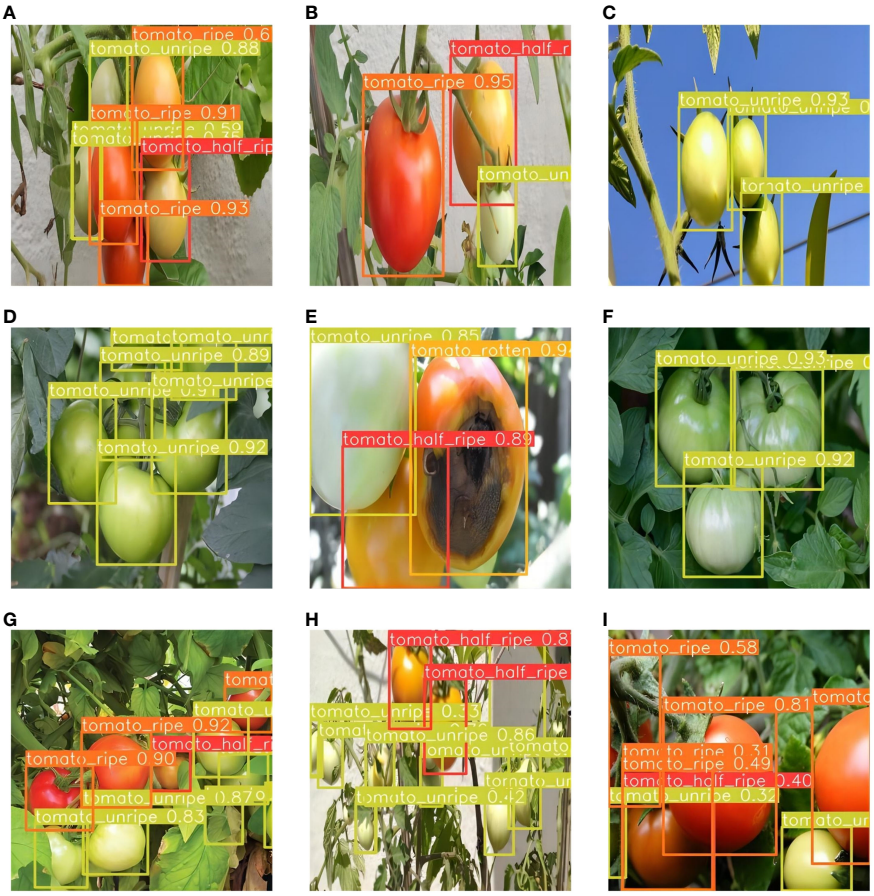


FIGURE 11
Examples of tomato ripeness detection results under different conditions. **(A–C)** under sunlight, **(D–F)** under shade, **(G–I)** under dense occlusion.

representation capability, as demonstrated by the ablation experiments, which show that the combination of slimneck and SSFF achieves the best performance, indicating the complementary nature of these two strategies. Furthermore, the Inner-EIoU loss function promotes more precise bounding box regression by penalizing the mismatch between the predicted and ground-truth boxes, and the comparison with other state-of-the-art loss functions validates the superiority of Inner-EIoU in terms of convergence speed and detection accuracy.

A comprehensive analysis of the improved model reveals its strong capability in detecting tomatoes at different ripeness stages, particularly for unripe, overripe, and rotten tomatoes. The high precision and recall rates demonstrate that the model can capture the unique features associated with each ripeness stage. Comparisons with other object detection models, including Faster-RCNN, SSD, YOLOv5, YOLOv6, YOLOv8, and the RT-DETR series, highlight the superior performance of PDSI-RTDETR in terms of accuracy and speed. Visual analysis under various environmental conditions proves the model's robustness in handling challenges such as uneven illumination, occlusion, and fruit overlap. When applied to intelligent harvesting robots in the future, this approach has the potential to enhance the quality of tomato harvesting by reducing the collection of immature and spoiled fruits.

7 Conclusion

This study proposes a fine-grained identification method for assessing tomato ripeness, which can provide more accurate decision support for tomato grading intelligent harvesting and management. In this paper, an efficient lightweight tomato ripeness detection model, PDSI-RTDETR, is proposed based on RT-DETR architecture. Substituting the original Basic_Block module of the backbone network with a streamlined PConv_Block module renders the model more compact, lowering computational requirements and the total parameter count. Then the deformable attention module is combined with AIFI module to enhance the ability of extracting detailed features of tomato and improve the accuracy and efficiency of the model. In addition, the slimneck-SSFF feature fusion architecture is proposed as the feature fusion network structure of the model, which utilizes the GSConv and VoVGSCSP modules to enhance the adaptability of the detector to tomatoes at different scales, generating more fine-grained semantic information while reducing the network computation. Ultimately, the original GIoU loss function is substituted by the Inner-EIoU loss function, which utilizes auxiliary frames to improve the ripeness determination accuracy for overlapping and small target tomatoes.

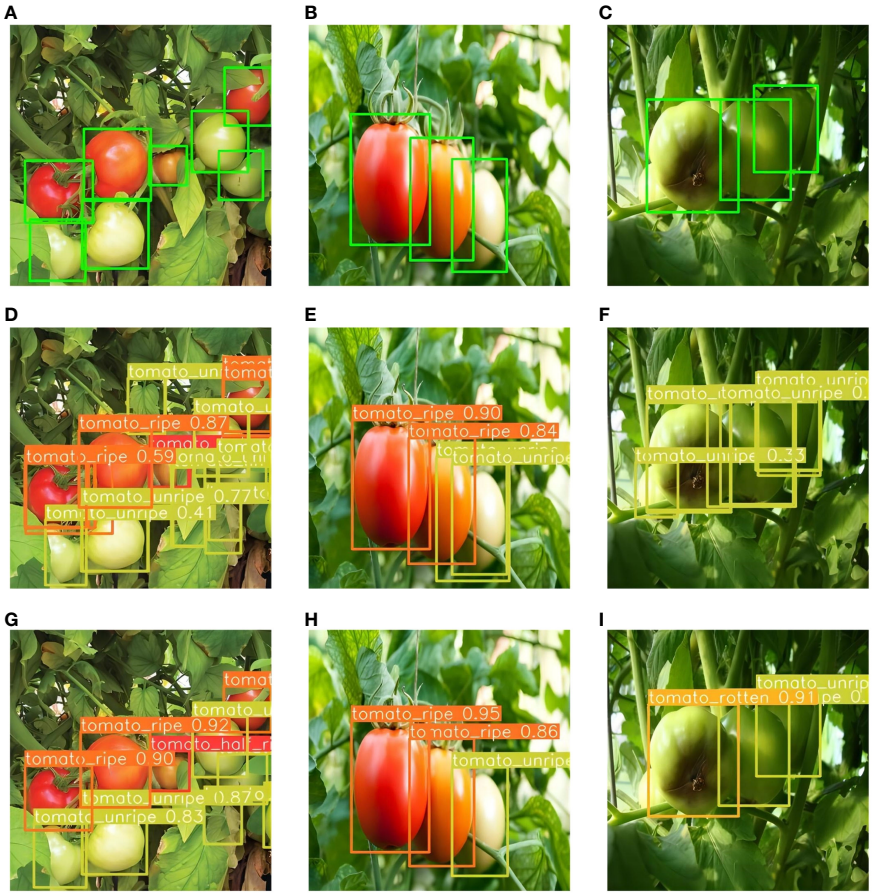


FIGURE 12
Partial detection results of two models. **(A–C)** Ground truth bounding boxes, **(D–F)** Predicted images from RT-DETR model, **(G–I)** Predicted images from PDSI-RTDETR.

Extensive tests were performed to validate the efficacy of PDSI-RTDETR. Comparative results demonstrate that PDSI-RTDETR achieves enhancements in Precision, Recall, mAP50 and mAP50:95 by 4.7%, 5.4%, 3.9% and 4.1% respectively, compared to the RT-DETR. Concurrently, GFLOPs and parameters were decreased by 17.6% and 30.8% respectively, with the detection time per image being merely 4.2 milliseconds. Furthermore, the experiments confirmed the robustness of PDSI-RTDETR, evidencing its effective detection of tomato ripeness across various scenes involving different lighting and occlusion conditions. Future research will focus on integrating the model into a practical intelligent tomato harvesting robot for on-site validation. The harvesting system will be deployed in real tomato farming environments, and its performance will be evaluated under various field conditions. Furthermore, we plan to expand our research to explore ripeness detection methods for other fruits and agricultural products. By adapting the model architecture and training methodology, we aim to develop similar lightweight and efficient ripeness detection systems for a wider range of crops. Through these efforts, we strive to contribute to the development of intelligent and sustainable agricultural practices, improve the quality of harvested crops, reduce waste, and enhance agricultural efficiency.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SW: Conceptualization, Methodology, Project administration, Software, Writing – original draft, Writing – review & editing. HJ: Funding acquisition, Resources, Writing – review & editing. JY: Conceptualization, Validation, Writing – original draft. XM: Formal analysis, Investigation, Writing – review & editing. JC: Data curation, Visualization, Writing – review & editing. ZL: Writing – original draft, Writing – review & editing. XT: Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the National Natural Science

Foundation of China (No. 61773416); Graduate Research and Practice Projects of Minzu University of China (No. SJCX2024021).

Acknowledgments

Our sincere thanks to the Key Laboratory of Ethnic Language Intelligent Analysis and Security Governance of MOE for their generous support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1415297/full#supplementary-material

References

Afonso, M., Fonteijn, H., Polder, G., and Wehrens, R. (2020). Tomato fruit detection and counting in greenhouses using deep learning. *Front. Plant Sci.* 11, 571299. doi: 10.3389/fpls.2020.571299

Aghilinategh, N., Dalvand, M. J., and Anvar, A. (2020). Detection of ripeness grades of berries using an electronic nose. *Food Sci. Nutr.* 8, 4919–4928. doi: 10.1002/fsn3.1788

Alam Siddiquee, K. N. E., Islam, M. S., Dowla, M. Y. U., Rezaul, K. M., and Grout, V. (2020). Detection, quantification and classification of ripened tomatoes: a comparative analysis of image processing and machine learning. *IET Image Process.* 14, 2442–2456. doi: 10.1049/iet-ipr.2019.0738

Bai, Y., Mao, S., Zhou, J., and Zhang, B. (2023). Clustered tomato detection and picking point location using machine learning-aided image analysis for automatic robotic harvesting. *Precis. Agric.* 24, 727–743. doi: 10.1007/s11119-022-09972-6

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020). "End-to-end object detection with transformers," in *European conference on computer vision*. 213–229 (Cham, Switzerland: Springer).

Chen, J., Kao, S.-h., He, H., Zhuo, W., Wen, S., Lee, C.-H., et al. (2023). "Run, Don't walk: Chasing higher FLOPS for faster neural networks," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Piscataway, NJ, USA: IEEE), 12021–12031.

El-Bendary, N., El Hariri, E., Hassanien, A. E., and Badr, A. (2015). Using machine learning techniques for evaluating tomato ripeness. *Expert Syst. Appl.* 42, 1892–1905. doi: 10.1016/j.eswa.2014.09.057

Fu, R., Hu, Q., Dong, X., Guo, Y., Gao, Y., and Li, B. (2020). Axiom-based grad-cam: Towards accurate visualization and explanation of cnns. *arXiv preprint* arXiv:2008.02312.

Gevorgyan, Z. (2022). SIoU loss: More powerful learning for bounding box regression. *arXiv preprint* arXiv:2205.12740.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*. (Piscataway, NJ, USA: IEEE), 2961–2969.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Piscataway, NJ, USA: IEEE), 770–778.

Kang, M., Ting, C.-M., Ting, F. F., and Phan, R. C.-W. (2023). ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation. *arXiv preprint* arXiv:2312.06458.

Khan, A., Hassan, T., Shafay, M., Fahmy, I., Werghi, N., Mudigansalage, S., et al. (2023). Tomato maturity recognition with convolutional transformers. *Sci. Rep.* 13, 22885. doi: 10.1038/s41598-023-50129-w

Ko, K., Jang, I., Choi, J. H., Lim, J. H., and Lee, D. U. (2021). Stochastic decision fusion of convolutional neural networks for tomato ripeness detection in agricultural sorting systems. *Sensors* 21, 917. doi: 10.3390/s21030917

Lawal, M. O. (2021). Tomato detection based on modified YOLOv3 framework. *Sci. Rep.* 11, 1–11. doi: 10.1038/s41598-021-81216-5

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., et al. (2022b). YOLOv6: A single-stage object detection framework for industrial applications. *arXiv preprint* arXiv:2209.02976.

Li, H., Li, J., Wei, H., Liu, Z., Zhan, Z., and Ren, Q. (2022a). Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. *arXiv preprint* arXiv:2206.02424.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands. (Cham, Switzerland: Springer), 21–37.

Liu, G., Mao, S., and Kim, J. H. (2019). A mature-tomato detection algorithm using machine learning and color analysis. *Sensors* 19, 2023. doi: 10.3390/s19092023

Liu, G., Nouaze, J. C., Touko Mbouembe, P. L., and Kim, J. H. (2020). YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3. *Sensors* 20, 2145. doi: 10.3390/s20072145

Lv, W., Xu, S., Zhao, Y., Wang, G., Wei, J., Cui, C., et al. (2023). Detrs beat yolos on real-time object detection. *arXiv preprint* arXiv:2304.08069.

Qi, Y., He, Y., Qi, X., Zhang, Y., and Yang, G. (2023). "Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. (Piscataway, NJ, USA: IEEE), 6070–6079.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Piscataway, NJ, USA: IEEE), 779–788.

Redmon, J., and Farhadi, A. (2017). "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Piscataway, NJ, USA: IEEE), 7263–7271.

Ren, S., He, K., Girshick, R., and Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. (2019). "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Piscataway, NJ, USA: IEEE), 658–666.

Seth, K. (2020). *Fruits and vegetables image recognition dataset* (Kaggle). Available at: https://www.kaggle.com/kritikseth/fruit-and-vegetable-image-recognition.

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Story, E. N., Kopec, R. E., Schwartz, S. J., and Harris, G. K. (2010). An update on the health effects of tomato lycopene. *Annu. Rev. Food Sci. Technol.* 1, 189–210. doi: 10.1146/annurev.food.102308.124120

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30.

Xia, Z., Pan, X., Song, S., Li, L. E., and Huang, G. (2022). "Vision transformer with deformable attention," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Piscataway, NJ, USA: IEEE), 4794–4803.

Zhang, Y.-F., Ren, W., Zhang, Z., Jia, Z., Wang, L., and Tan, T. (2022). Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* 506, 146–157. doi: 10.1016/j.neucom.2022.07.042

Zhang, X., Song, Y., Song, T., Yang, D., Ye, Y., Zhou, J., et al. (2023b). AKConv: convolutional kernel with arbitrary sampled shapes and arbitrary number of parameters. *arXiv preprint* arXiv:2311.11587.

Zhang, H., Xu, C., and Zhang, S. (2023a). Inner-iou: more effective intersection over union loss with auxiliary bounding box. *arXiv preprint* arXiv:2311.02877.

Zhang, H., and Zhang, S. (2023). Shape-ioU: more accurate metric considering bounding box shape and scale. *arXiv preprint arXiv:2312.17663*.

Zhao, Y., Gong, L., Zhou, B., Huang, Y., and Liu, C. (2016). Detecting tomatoes in greenhouse scenes by combining AdaBoost classifier and colour analysis. *Biosyst. Eng.* 148, 127–137. doi: 10.1016/j.biosystemseng.2016.05.001

Zheng, T., Jiang, M., Li, Y., and Feng, M. (2022). Research on tomato detection in natural environment based on RC-YOLOv4. *Comput. Electron. Agric.* 198, 107029. doi: 10.1016/j.compag.2022.107029

Zhong, J., Chen, J., and Mian, A. (2022). DualConv: Dual convolutional kernels for lightweight deep neural networks. *IEEE Trans. Neural Networks Learn. Systems*. 34, 9528–9535.

Zhu, X., Hu, H., Lin, S., and Dai, J. (2019). "Deformable convnets v2: More deformable, better results," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Piscataway, NJ, USA: IEEE), 9308–9316.l

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 25 July 2024
DOI 10.3389/fpls.2024.1415884

Check for updates

OPEN ACCESS

EDITED BY
Yuntao Ma,
China Agricultural University, China

REVIEWED BY
Yingpu Che,
Chinese Academy of Agricultural Sciences (CAAS), China
Tongyu Hou,
Shihezi University, China
Teng Miao,
Shenyang Agricultural University, China
Meiyan Shu,
Henan Agricultural University, China
Qian Wu,
Jiangsu Academy of Agricultural Sciences (JAAS), China

*CORRESPONDENCE
Xuliang Duan
duanxuliang@sicau.edu.cn

†These authors have contributed equally to this work

RECEIVED 11 April 2024
ACCEPTED 11 July 2024
PUBLISHED 25 July 2024

CITATION
Pan F, Hu M, Duan X, Zhang B, Xiang P, Jia L, Zhao X and He D (2024) Enhancing kiwifruit flower pollination detection through frequency domain feature fusion: a novel approach to agricultural monitoring. *Front. Plant Sci.* 15:1415884. doi: 10.3389/fpls.2024.1415884

COPYRIGHT
© 2024 Pan, Hu, Duan, Zhang, Xiang, Jia, Zhao and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Enhancing kiwifruit flower pollination detection through frequency domain feature fusion: a novel approach to agricultural monitoring

Fei Pan[1,2†], Mengdie Hu[1,3†], Xuliang Duan[1,2*], Boda Zhang[1,3], Pengjun Xiang[1,3], Lan Jia[1,3], Xiaoyu Zhao[1,3] and Dawei He[1,3]

[1]College of Information Engineering, Sichuan Agricultural University, Ya'an, China, [2]Ya'an Digital Agricultural Engineering Technology Research Center, Sichuan Agricultural University, Ya'an, China, [3]Agricultural Information Engineering Higher Institution Key Laboratory of Sichuan Province, Sichuan Agricultural University, Ya'an, China

The pollination process of kiwifruit flowers plays a crucial role in kiwifruit yield. Achieving accurate and rapid identification of the four stages of kiwifruit flowers is essential for enhancing pollination efficiency. In this study, to improve the efficiency of kiwifruit pollination, we propose a novel full-stage kiwifruit flower pollination detection algorithm named KIWI-YOLO, based on the fusion of frequency-domain features. Our algorithm leverages frequency-domain and spatial-domain information to improve recognition of contour-detailed features and integrates decision-making with contextual information. Additionally, we incorporate the Bi-Level Routing Attention (BRA) mechanism with C3 to enhance the algorithm's focus on critical areas, resulting in accurate, lightweight, and fast detection. The algorithm achieves a $mAP_{0.5}$ of 91.6% with only 1.8M parameters, the AP of the Female class and the Male class reaches 95% and 93.5%, which is an improvement of 3.8%, 1.2%, and 6.2% compared with the original algorithm. Furthermore, the Recall and F1-score of the algorithm are enhanced by 5.5% and 3.1%, respectively. Moreover, our model demonstrates significant advantages in detection speed, taking only 0.016s to process an image. The experimental results show that the algorithmic model proposed in this study can better assist the pollination of kiwifruit in the process of precision agriculture production and help the development of the kiwifruit industry.

KEYWORDS

YOLO, frequency domain feature fusion, BRA, kiwifruit flower, precision agriculture deleted

1 Introduction

Kiwifruit, one of the most widely consumed fruits globally, is prominent in the fruit market due to its distinctive characteristics and unique medicinal properties (Gökdoğan, 2022). Kiwifruit is a dioecious vine whose sex is recognized only at flowering and requires cross-pollination for fruiting. However, in practical orchards, natural biological pollination is often less efficient, which can adversely affect both fruit quality and the economic viability of production (Castro et al., 2021). Artificial pollination is commonly used to improve kiwifruit pollination rates in kiwifruit plantations. However, this method comes with high labor costs and the inability to accurately determine the success of pollination, leading to redundant work and increased planting expenses. In recent years, there has been rapid growth in precision agriculture and smart industry, with automated robots increasingly utilized across various agricultural domains (Yépez-Ponce et al., 2023), including robotic intelligent harvesting (Wang et al., 2022b), plant protection (Chen et al., 2021), and now, robotic pollination (Yang et al., 2023a). Robotic pollination is also expected to improve kiwifruit pollination efficiency and reduce labor costs (Gao et al., 2023a). The crux of this technology lies in accurately recognizing kiwifruit flower status and sex, particularly in complex orchard settings. Target detection, one of machine vision's basic and challenging tasks, has been widely applied in agriculture by integrating deep learning techniques. These include fruit ripeness detection (Lu et al., 2022), pest and disease identification (Qi et al., 2022), weed detection (Hasan et al., 2021), fruit and vegetable quantity counting (Wang et al., 2022a), and pesticide residue detection (Liu et al., 2019). Deep learning and target detection techniques offer potential technical support for automating kiwifruit pollination.

Target detection algorithms based on deep learning mainly include two-step and one-step frameworks. Two-stage algorithms (e.g., the Faster R-CNN (Ren et al., 2017)) achieve object detection through candidate region generation and classification. However, they are computationally intensive and slow, limiting their applicability in real-world agricultural scenarios. In contrast, single-stage algorithms (e.g., the YOLO (Redmon et al., 2016)) extract features directly from the image for target classification and localization. This approach balances speed and accuracy, making it well-suited for real-time applications and capable of meeting industrial-grade detection standards. The advancement of these algorithms continues to drive forward computer vision technology, offering effective tools for a wide range of practical applications.

The YOLO algorithm stands out as one of the most popular methods in the field of target detection, widely adopted in artificial intelligence and computer vision for its speed and accuracy. In the domain of flower recognition, Wu (Wu et al., 2020) and colleagues developed a real-time apple blossom detection method based on YOLOv4 and CSPDarknet53 and optimized the model structure by channel pruning algorithm. The model reduces the number of parameters by 96.74% while maintaining high accuracy ($mAP_{0.5}$ of 97.31%). However, their study primarily focused on detecting flower clusters, lacking a comprehensive exploration of extracting overall features of individual apple blossoms and failing to address detection challenges posed by fuzzy influences. Similarly, Xu et al. (Xu et al., 2022) introduced FlowerYolov5, a tomato bloom detection method based on the YOLOv5s architecture. This method incorporates a novel feature fusion layer. It integrates the Convolutional Block Attention Module (CBAM) to identify the bud, bloom, and fruiting stages of tomato flowers, leading to a notable improvement in $mAP_{0.5}$ by 7.8%, with a parameter count of 23.9M. However, the method neglected the detection of flowers in dense situations with a large number of parameters. Mithra (Mithra and Nagamalleswari, 2023) et al. employed UAV images to discern the sex of cucurbit plant flowers using deep transfer learning and a YOLOv4-based approach to support research on autonomous pollination. They manually labeled sex characteristics to identify and classify flowers in complex contexts, including interference, overlap, and ambiguous situations. This method achieves a $mAP_{0.5}$ of 91.2%, demonstrating its capability to quickly and effectively recognize small target objects in dense environments.

Researchers are also committed to identifying kiwifruit flowers to address the issue of autonomous pollination in kiwifruit. Williams et al. (Williams et al., 2020) first proposed a kiwifruit pollination robot and achieved an accuracy of 79.5% in flower recognition using convolutional neural networks. Lim et al. (Lim et al., 2020) utilized the Faster R-CNN Inception V2 model to achieve an accuracy of 91.9% in kiwifruit flower detection to enhance detection accuracy further. However, the Williams (Williams et al., 2020) and Lim et al. (Lim et al., 2020) teams only annotated female flowers, overlooking the potential occurrence of multi-class flowers during kiwifruit pollination. Consequently, Li et al. (Li et al., 2022b) emphasized the significance of identifying kiwifruit flowers and buds for robotic pollination, establishing a dataset that includes both flowers and buds. They achieved a $mAP_{0.5}$ of 91.49% with YOLOv4. Subsequently, in another study (Li et al., 2022a), the team introduced a novel method for multi-class detection and distribution recognition of kiwifruit flowers using YOLOv5l and Euclidean distance, further refining the classification of kiwifruit flowers. Their YOLOv5l model achieved a of 91.60%. In the agricultural sector, numerous researchers (Gai et al., 2023; Tian et al., 2023; Yu et al., 2023; Zeng et al., 2023) are refining the YOLO algorithm for the best performance in natural agricultural environments. Despite prior studies demonstrating commendable performance in detection accuracy, practical application issues such as model size and detection speed have been overlooked. The previously proposed models are larger and not applicable to real production environments.

Although the YOLO algorithm demonstrates advantages in target detection tasks, its performance in detecting dense targets within complex environments may be hindered by interference. The frequency domain features of the image contain essential information about the structure and content of the image, which helps to solve the dense detection problem of the YOLO algorithm in complex environments. Qin et al. (Qin et al., 2021) studied the channel attention problem by combining frequency domain features with a channel attention mechanism that compresses the channel using a customized Discrete Cosine Transform (DCT) but only utilizes a portion of each low-frequency message. Su et al. (Su et al., 2021) proposed a complete frequency channel attention network

based on the attention mechanism in remote sensing. This approach enhances Qin et al.'s (Qin et al., 2021) weight acquisition method by attenuating high-frequency details and maximizing the utilization of low-frequency components. However, this network may lose some high-frequency details during noise filtering, leading to the omission of certain features. Duan et al. (Duan et al., 2023) introduced a detection method for infrared small targets integrating frequency-domain clutter suppression and spatial-domain feature extraction using deep learning techniques. They applied DCT to convert spatial-domain infrared images into the frequency domain. Following this transformation, the frequency domain information was filtered using an attention mechanism to adaptively enhance the target suppression background. Ultimately, they reversed the processed frequency-domain data to the spatial domain to facilitate feature extraction and fusion. The method was a lossy transform in which the image in the frequency domain does not retain all the details of the original image, resulting in the loss of important information. Through meticulous analysis and adept utilization of frequency domain information within an image, researchers can finely adjust feature extraction methodologies to accommodate diverse frequency ranges. This meticulous approach significantly enhances the precision with which critical image features are captured within deep learning frameworks. Nevertheless, the aforementioned methods do not fully exploit all features of the frequency domain and each incurs some degree of loss in frequency domain information.

Research models for flower recognition currently face challenges, especially in identifying kiwifruit flowers, such as oversized model dimensions and Ambiguous categorization of flower categories. This study proposes a lightweight algorithm called KIWI-YOLO in response to these challenges. KIWI-YOLO effectively reduces the parameters of the kiwifruit flower detection model and improves the accuracy of flower detection. The main contributions of this study are as follows:

(1) The kiwifruit flower's whole stage image data set is constructed. In order to solve the problem of multi-category recognition of kiwifruit flowers in the natural environment, this study collected four kinds of kiwifruit flower image data covering spore, female flower, male flower, and pollination success state in real orchard environment. The original image is screened and labeled to form a full-stage image data set of kiwifruit flowers.

(2) A frequency domain feature fusion module is proposed. In order to improve the detection accuracy of the model in complex and dense environments, this study proposes a frequency domain feature fusion module (FDFF). This module enhances the model's feature integration ability by extracting the high-frequency components of the original image and fusing them with the original image.

(3) BiFormerBlock is introduced to improve the model. In this study, by introducing BiFormerBlock into the C3 module of YOLOv5n, the perception ability of the model to fine-grained features is improved, and the accurate detection of smaller targets is realized.

Experimental data demonstrate that with 1,816,147 parameters, KIWI-YOLO achieves a $mAP_{0.5}$ of 91.6% and an F1-score of 86.5%. Compared to YOLOv5n, KIWI-YOLO shows an improvement of 3.8% in $mAP_{0.5}$ and 3.1% in F1-score. The experimental results show that KIWI-YOLO maintains a high recognition accuracy and significantly improves the feature expression ability on the basis of lightweight. It can quickly and accurately detect kiwifruit flowers and promote the further development of industrial automatic pollination technology.

2 Materials and methods

2.1 Data collection and processing

The experimentally collected data on all stages of the kiwifruit flower were obtained from Yucheng kiwifruit, a Chinese National Geographical Indication product, cultivated in Yucheng District, Ya'an City, Sichuan Province, China. This data was gathered during the 2023 flowering season in Longquan Village, Yucheng District, Ya'an City, Sichuan Province (103.03305, 30.00531). The angles of different growth states in the stage of kiwifruit flower and fruit are relatively consistent, all facing the ground. Therefore, when shooting, the shooting angle was from bottom to top in the vertical direction. The camera used was an iPhone 13 Pro Max with a resolution of 4032*3024, and data collection occurred under natural lighting conditions at 10:00 a.m., 2:00 p.m., and 6:00 p.m. The light during image acquisition includes low light, normal light, and overexposed high light to increase the richness of the data, as shown in Figure 1.

In the orchard environment, kiwifruit flowers, especially bracts, are usually densely distributed. The data set of this study covers such small and dense targets to ensure the effectiveness of the proposed algorithm in practical applications. The kiwifruit flower dataset consists of four stages: bud, female flower, male flower, and successful pollination, as depicted in Figure 2. The collection time of bud data was in early April 2023. Subsequently, according to the growth cycle of kiwifruit flowers, the collection time was from mid-April to mid-May. The data of male and female flowers were collected at the pollination stage after flowering, and the data of the pollination success category were collected one week after pollination. The dataset consists of 1594 images, which were divided into training sets and test sets at a ratio of 9: 1. Labelimg was used to manually annotate the collected data. The annotation file is in YOLO format. The total number of tags is 13450, covering the four characteristic classes of kiwifruit flowers, including Female (female flower), Bud (bud), Male (male flower), and Success (successful pollination state). Their distribution ratio is about 3:3:2:2.

Data augmentation was implemented on the collected data using an online approach to enhance the model's robustness and generalization during training. Online enhancement preprocesses the data without altering the dataset size. In the online data augmentation approach, Mosaic (Bochkovskiy et al., 2020) randomly selects four images from the dataset, applies random cropping and scaling to each, and then arranges and stitches them together to create a single image. Mosaic enriches the dataset by

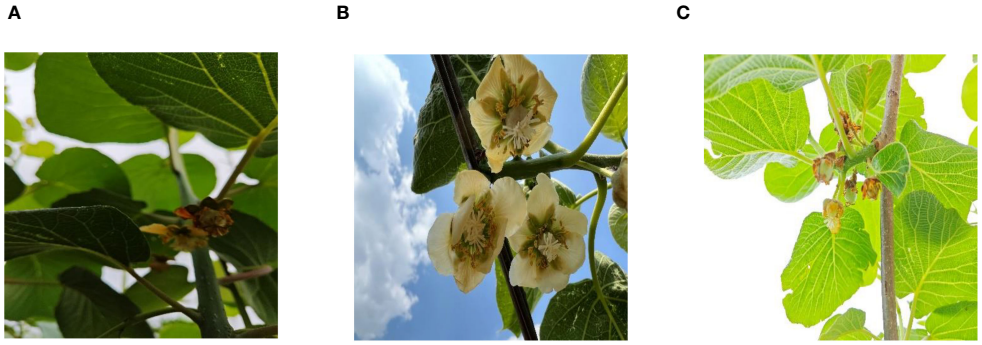


FIGURE 1
Examples of the different light. Panel **(A)** depicts the scenario under low light intensity, panel **(B)** demonstrates moderate light intensity, and panel **(C)** showcases high light intensity.

adding small target samples. On the other hand, a mixup superimposes the two images according to their weights to generate a new image. Mixup makes the model focus more on the commonality of the data, which effectively improves the generalization ability and robustness of the model. This study combines Mosaic and Mixup to enhance the original data online.

2.2 YOLOv5 algorithm

The YOLOv5 architecture comprises three main components: Backbone, Neck, and Head. The Backbone aggregates image features across different scales using modules like Focus, BottleneckCSP, and SPP. It generates five feature layers from the

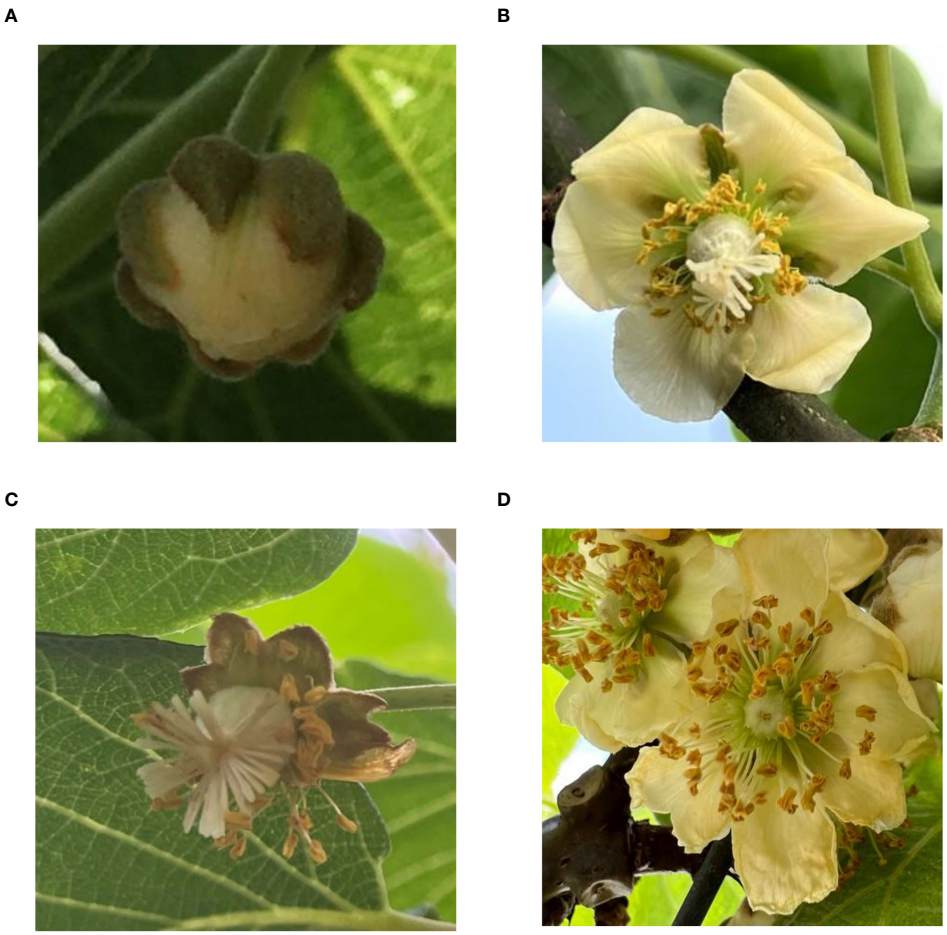


FIGURE 2
Four stages of kiwifruit flowers. Panel **(A)** is the bud stage of kiwifruit flower, panel **(B)** is the female flower of kiwifruit flower to be pollinated, panel **(C)** is the female flower of kiwifruit flower to be pollinated successfully, and panel **(D)** is the male flower of kiwifruit flower.

input image, which are fused with feature maps from the Neck layer to enhance feature extraction. The Neck, adopting the FPN (Feature Pyramid Network)+PAN (Pixel Aggregation Network) structure, combines semantic and localization features. It integrates mixed image features and forwards them to the prediction layer. The Head is the classifier and regressor of YOLOv5. It determines the object corresponding to the feature based on the feature layer, recursively derives the target result of the detection, generates the bounding box, and predicts the category.

2.3 The proposed FDFF: frequency domain feature fusion module

In the whole stage of kiwifruit flower pollination, no matter whether it is the bud, female, male, or successful pollination kiwifruit flower state, the difference in color characteristics is small. When the neural network extracts the image features, the features provided by the color information are not enough to support the neural network in distinguishing the above four categories. In the actual production process, farmers distinguish the above four categories in the following ways:

(1) Shape: The bud exhibits a nearly circular shape, clearly distinct from the other three types.

(2) Presence of stigma: Distinguish male and female flowers by whether there is a stigma.

(3) Set fruit: Successful pollination of kiwifruit flowers is characterized by set fruit, and the main feature of this stage is whether the petals are in a state of shedding, and its shape is also significantly different from the other stages.

The above three points can be summarized as the result of judging the kiwifruit flower by the contour information. In order to make the model extract the contour information more effectively, this study proposes a frequency domain feature fusion module (FDFF). FDFF transforms the feature map from the spatial domain to the frequency domain for processing and obtains the high-frequency component of the kiwifruit flower image (the high-frequency component correlates the part of the image with a faster gray change, videlicet the edge and contour of the image). The high-frequency component processed by the module is fused with the original feature map that retains the low-frequency component. After adding the FDFF module, the model can enhance the ability to obtain the contour information of each stage of kiwifruit flowers. The FDFF module consists of the Fast Fourier Transformation (FFT) layer, a MaxPooling layer, a Depthwise Separable Convolution (DSC) (Howard et al., 2017) layer, and a Spatial Attention (Woo et al., 2018) layer. The overall module flowchart of the FDFF module is shown in Figure 3A.

The FDFF module performs a two-dimensional Fourier transform of the input image through the FFT layer to convert the feature information in the spatial domain to the feature information in the frequency domain to extract the frequency domain features. Suppose the given tensor is $X_S \in R^{H\times W\times C}$, which is converted to X_F By FFT module. The conversion process is given by Equation 1:

$$X_F=\sum_{n=0}^{N-1} X_s e^{\frac{-j2\pi Fn}{N}}, F=0,1,\ldots,N-1 \quad (1)$$

where X_s denotes spatial-domain feature information, X_F denotes frequency-domain feature information, and j denotes the imaginary part; F denotes the subscript of the frequency-domain sampling point, ranging from 0 to N-1; and n denotes the subscript of the spatial-domain sampling point, ranging from 0 to N-1.

A two-dimensional complex matrix is obtained by FFT, with the real part representing the magnitude of the information in the frequency domain and the imaginary part representing the phase information. In this study, the improvement effect of the frequency domain data model is determined by comparing the use of only the

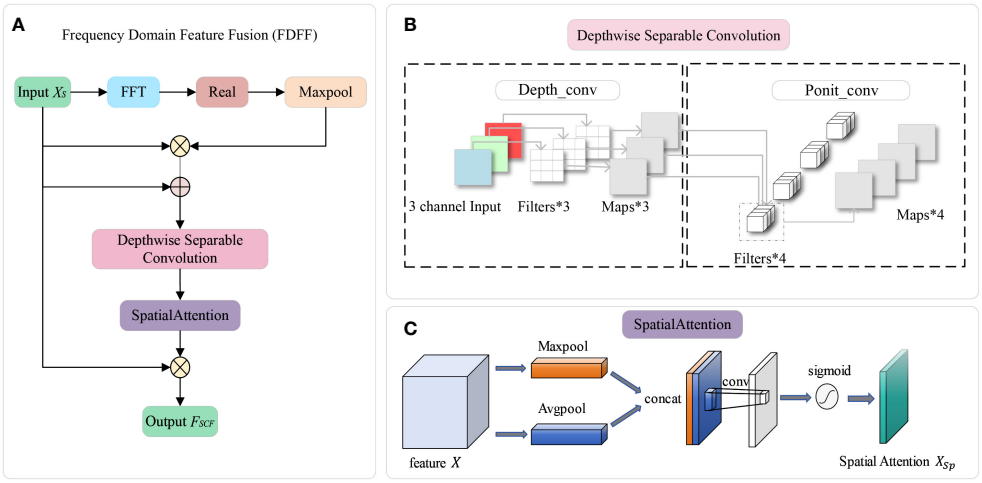


FIGURE 3
Panel **(A)** shows the overall flowchart of the FDFF module, which mainly includes the following modules: FFT layer, MaxPooling layer, DSC layer, and SpatialAttention layer. Panels **(B, C)** are the critical modules in FDFF. Panel **(B)** is the DSC structure diagram. Panel **(C)** is a spatial attention diagram.

real part, only the imaginary part, and the combination of the real part and the imaginary part. The experimental results show that the $mAP_{0.5}$ using only the real part is 91%, using only the imaginary part is 88.9%, and using both the real and imaginary parts is 90.01%. It is speculated that the high-frequency component corresponds to the important features of the image, and the amplitude is invariant. However, the interpretation and utilization of phase is more complex, the model is difficult to understand and use the relevant information, and the performance enhancement is limited. The task of this study is more important for the perception of global structure and texture. Therefore, this study uses the amplitude and spectrum and ignores the phase information by taking the real part of the operation inside the FDFF, which simplifies the problem and improves efficiency.

The max-pooling layer in this module aims to achieve frequency domain filtering and reduce the dimensionality of the data. The max-pooling selects the maximum value in each local region in the frequency domain to extract the main high-frequency features, which help identify the salient frequency domain components in the image, including texture features.

After the max-pooling layer operation, the output feature map X_M is obtained. Before the data enters the DSC layer, these frequency domain components are fused in the frequency domain to strengthen the frequency domain feature information. The multiplication of X_M with the original input data X_S is performed as shown in Equation 2.

$$\boldsymbol{X_{MS}=X_M*X_S} \tag{2}$$

where X_{MS} is the output.

After completing the fusion of frequency domain components, the FDFF module splices the obtained feature map containing high-frequency contour information with the input feature map to ensure the integrity of the spatial-frequency domain feature information so that the model can make full use of the relevant features. In order to reduce the complexity of the algorithm, the FDFF module uses DSC to perform feature fusion and tensor dimension transformation on the obtained feature map. DSC divides the convolution operation into two sub-operations: Depthwise Convolution and Pointwise Convolution. Compared with ordinary convolution, DSC has the advantages of small parameters and small computation, which can improve the speed of convolution operation.

As shown in Figure 3B, Depth_conv uses a convolution kernel for each channel separately, and the output channel of a single channel is one after the convolution operation. For the three-channel RGB map, Depth_conv outputs three feature maps with channel 1. Then, the three feature maps are sequentially spliced to get an output feature map with channel N. Through Ponit_conv processing, the convolution kernel is 1*1*M(M is the number of channels in the previous layer). Based on the previous map, a new feature map is generated by a weighted combination in the depth direction, which is used to reduce the dimension of the input feature map while maintaining the feature information's invariance. The purpose of independent convolutional computation for each channel is to reduce the number of parameters, reduce the computational effort, and increase the computational efficiency. The number of parameters for conventional convolution is Equation 3, and the number of parameters for DSC is Equation 4. The comparison of the formulas shows that the number of parameters and computation of the depth separable convolution is one-third of the conventional convolution.

$$\boldsymbol{parameter_{conv}=W_F\times H_F\times C_I\times C_O} \tag{3}$$

$$\boldsymbol{parameter_{dsc}=W_F\times H_F\times C_I+1\times 1\times C_I\times C_O} \tag{4}$$

where W_F is the width of the convolution kernel, H_F is the height of the convolution kernel, C_I is the number of input channels and C_O is the number of output channels.

The role of the spatial attention layer is to allow the task network to focus more on finding helpful information related to the current output to improve the quality of the output. Using spatial attention can act on the pixels of the feature map to weigh each pixel and fully use each pixel's information, as shown in Figure 3C. After the weights of spatial attention have been extracted, they are multiplied by the original input X_S. At this point, the spatial attention will focus more on the position of the kiwi flower to be recognized in the picture, and then the feature map F_{SCF} is obtained.

2.4 C3BiF based on dynamic sparse attention mechanism

The attention mechanism aids the model in focusing on the most relevant parts of the input data, thereby capturing key information and establishing correlations between inputs more efficiently. In this study, the C3BiF module is proposed to replace part of the C3 modules in YOLOv5 for feature extraction. The C3BiF module introduces a BiFormerBlock module containing Bi-level Routing Attention (Zhu et al., 2023) by utilizing the constructor of the C3 module, as shown in Figure 4. The module computes attention in two stages: first, coarse-grained attention is performed. The image is divided into blocks to control sparsity, and self-attention is performed at this level. Inter-block correlations are computed using Q and K to form a relation matrix A. Next, A is sparsified, retaining the largest elements and identifying pairs of blocks that require further attention. Finally, based on the sparse matrix from the first stage, fine-grained self-attention is performed, and each patch only performs the attention computation on patches in other blocks related to its block.

The calculation of attention mechanism (Woo et al., 2018; Hu et al., 2018; Wang et al., 2020) weights requires the joint participation of the target and all elements, which is the attention mechanism between input and output. VIT (Dosovitskiy et al., 2020) uses the self-attention mechanism (Vaswani et al., 2017) to capture long-term dependencies for the problem of establishing the correlation between multiple inputs in the visual field. However, the self-attention mechanism often imposes a substantial computational burden and memory overhead. In this paper, BRA is used to address the above application bottlenecks. BRA leverages sparsity to reduce computation and memory usage, enabling dense matrix multiplication suitable for

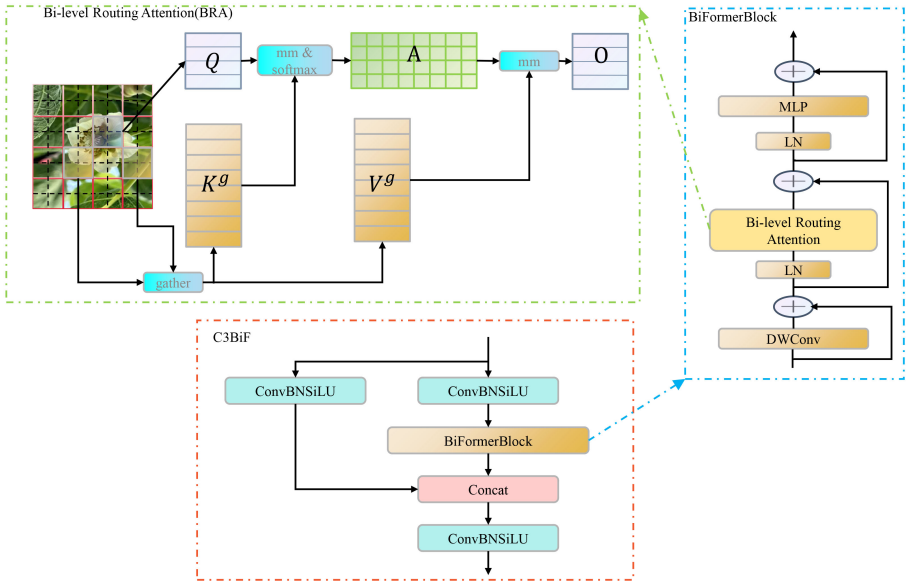


FIGURE 4
Components of the overall C3BIF architecture. BRA is the core of the C3BIF module. In the architecture of the BRA mechanism, Q is used to compute the keywords that are weighted in relevance to the query. K denotes the keyword or identifier used to provide information or for matching purposes. V is the value associated with the query result and the keyword information. A is the adjacency matrix denoting the semantic relevance between two regions. O is the output of the attention mechanism.

GPUs. This approach demonstrates excellent performance and computational efficiency, particularly in dense prediction tasks. A series of studies (Gao et al., 2023b; Yang et al., 2023b; Zhang et al., 2023) have indicated that, since BRA is based on sparse sampling rather than downsampling, it can retain fine-grained detail information. So, the BiFormerBlock containing the BRA module is suitable for detecting small objects.

2.5 The improved KIWI-YOLO algorithm

Both YOLOv8n and YOLOv5n are considered excellent nano models. In the preliminary experiments of this study, the number of parameters was 1764577 for YOLOv5n and 3302624 for YOLOv8n. The number of parameters for YOLOv8n was 1.8 times that of YOLOv5n, and $mAP_{0.5}$ was only improved by 0.03 points. Compared with the exponential increase in the number of parameters, the detection performance does not increase exponentially but leads to an increase in model complexity. Considering the subsequent application architectures of the study, the lightweight variant of YOLOv5, known as YOLOv5n, optimized for Nanodevices, has been chosen. It provides detection accuracy suitable for edge devices while maintaining high speed. The model has a channel number factor of 0.25 and a depth factor of 0.33.

This study was based on the improvement of the YOLOv5n network structure. In order to improve the detection accuracy of the algorithm, the fourth layer of the network (the next layer of P3) was replaced by the FDFF structure mentioned in the article. The FDFF improves the ability to extract and express network features. In addition, the overall detection efficiency is affected by the presence of small targets during the whole stage of pollination of kiwifruit flowers. Therefore, C3BiF containing BiFormer structure was introduced in the Head part. The overall KIWI-YOLO algorithm model is shown in Figure 5.

2.6 Training environment and evaluation indicators

In this study, the hardware was NVIDIA Quadro RTX 5000 Graphics Processing Unit (GPU) with 16GB of graphics memory and 128GB of RAM, and the Central Processing Unit (CPU) was i9-10900K, 3.7GHz. The system was Windows Server 2019 Standard, and PyCharm built the network environment with the deep learning framework PyTorch 2.0.0, the Python version of which was 3.8. For training, all images were uniformly resized to 640*640. The batch size for input images during training was set to 16, and the number of training rounds was 300. The optimizer employed was the SGD optimizer to optimize the network parameters. The initial learning rate (lr0) was set to 0.01, with an initial learning rate momentum of 0.937 and a weight decay of 0.0005.

The YOLO algorithm assesses its object detection system based on two core metrics: detection accuracy speed and detection accuracy. Detection speed refers to the time taken to detect one image. Detection accuracy is evaluated using metrics such as Precision (P), Recall (R), F1-score, Average Precision (AP), and mean Average Precision (mAP), calculated as described in Equations 5–9. In target detection, Intersection over Union (IoU) quantifies accuracy by comparing detection overlap and true labeled boxes. Typically, a ratio greater than 0.5 indicates confident detection. Moreover, the value of mean average precision, denoted as $mAP_{0.5}$, calculated at IoU = 0.5, is crucial for assessing

10.3389/fpls.2024.1415884

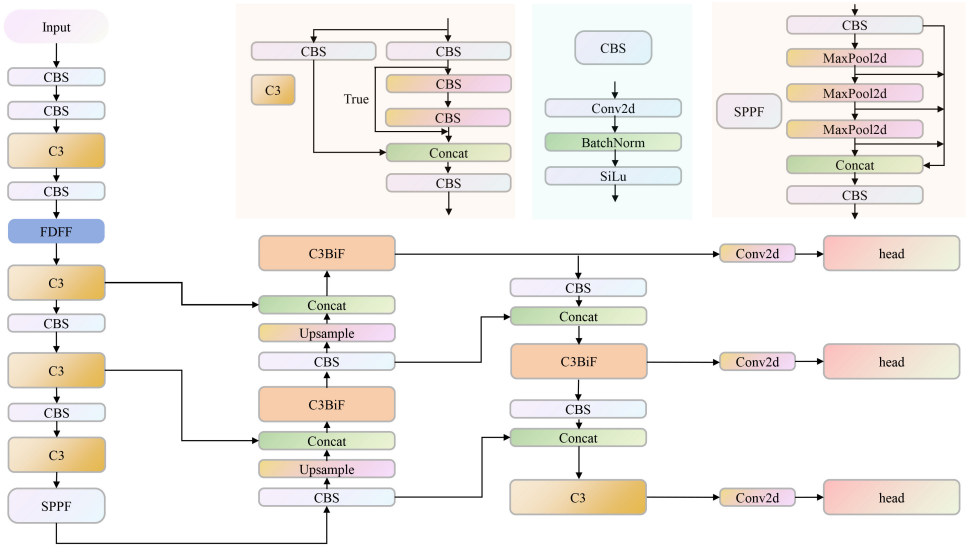


FIGURE 5
The structure of the KIWI-YOLO model. KIWI-YOLO comprises CBS, C3, SPPF, FDFF, C3BIF, and other modules. CBS encapsulates the combination of convolution, batch normalization, and activation function. C3 contains multiple CBS modules and a Bottleneck module. SFFP is a fast spatial pyramid pooling containing multiple maximum pooling layers and CBS modules. FDFF and C3BIF are the innovative modules proposed in this study.

the overall detection accuracy of the model. A higher $mAP_{0.5}$ signifies better detection accuracy.

$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{5}$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{6}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{7}$$

$$AP = \int_0^1 P(R)dR \tag{8}$$

$$mAP = \frac{1}{M} \sum_{m=1}^{M} AP_m \tag{9}$$

In the above Equation, TP denotes True Positive, TN denotes True Negative, FP denotes False Positive, FN denotes False Negative, M is the number of target classes to be detected, and AP_m is the average precision of the mth class of targets.

3 Experiments and results

3.1 FDFF module analysis

In this study, the insertion position of FDFF was experimentally compared to confirm the optimal insertion position of the FDFF module. The role of FDFF is to extract and fuse the features, so it is considered to be inserted into the backbone part of YOLOv5n to enhance the performance of the model's multilevel feature extraction. FDFF is inserted into the $P_n + 1$ layer respectively for comparative experiments to find the best effect. The experimental findings are summarized in Table 1. According to the experimental results, when the FDFF insertion is $P_3 + 1$, the effect is the best, the parameter is 1790195, the $mAP_{0.5}$ reaches 91%, and the $mAP_{0.5}$ is 3.2% higher than that of the non-insertion. The increase of $mAP_{0.5}$ inserted into the later layer of P1, P2, P4 and P5 was 2.5%, 2.7%, 2.7% and 2.5% respectively.

Analyzing the FDFF module, it can be seen that the module obtains the high-frequency information in the frequency domain while retaining the low-frequency information, fuses the frequency

TABLE 1 Results of FDFF module experiments; [P_n] corresponds to the nth Conv layer in the YOLOv5 backbone feature extraction network.

FDFF Adds Location	$mAP_{0.5}$(%)	Recall(%)	Parameters	Time(s)
Base	87.8	78.4	1764577	0.013
P_1+1	90.3(+2.5)	82.6(+4.2)	1766899	0.015
P_2+1	90.5(+2.7)	82.7(+4.3)	1772707	0.027
P_3 + 1	**91.0(+3.2)**	**83.6(+5.2)**	1790195	0.016
P_4 + 1	90.5(+2.7)	84.6(+6.2)	1882819	0.013
P_5+1	90.3(+2.5)	82.1(+3.7)	2230339	0.016

$P_n + 1$ corresponds to the next layer of P_n.
The bold values in the table emphasize the experimental results of the methods used in this study.

domain with the spatial domain, and strengthens the contour information of the kiwifruit flower image. The shallow layer of the feature extraction network mainly learns some relatively simple features, which are not abstract and complex enough for the task. In the deeper layers of the network, the neural network can learn more abstract and complex features, but at the same time, it faces the problems of training difficulty and gradient. FDFF performs better in the middle layer, which makes it easier for the gradient to be propagated in the network. In this position, it can better balance the abstraction, computational efficiency, and training effect.

3.2 C3BIF module analysis

In this study, in order to demonstrate the effect of the number of C3BIFs and the position of adding them on the network results, experiments were conducted on models that have been added with the FDFF module. Table 2 presents the experimental results. As can be seen from the table, the best effect is achieved when the first three C3 layers of the Head layer are all replaced with C3BIF and the $mAP_{0.5}$ reaches 91.6%, which is a 0.6% growth compared to the model with no replacements, and the number of parameters is 1816147. However, the growth effect of other replacement schemes is not obvious, and it will even cause a decrease of $mAP_{0.5}$.

Analyzing the C3BIF module, it is observed to contain the BiFormerBlock featuring Bi-level Routing Attention, which exhibits enhanced recognition capabilities for smaller targets. Since the fourth C3 in the head layer primarily deals with smaller feature maps for detecting large-scale targets, the foremost requirement for enhancement lies within the first three C3 layers of the Head. These initial C3 layers contribute significantly to the detection of small and medium-scale targets. Therefore, replacing three initial C3 layers with C3BIFs at the same time produces better results.

TABLE 2 Results of C3BIF module experiments.

N	$mAP_{0.5}$ (%)	Recall (%)	Parameters	Time (s)
1	90.9	82.3	1801395	0.018
2	90.3(-0.6)	83.4(+1.1)	1793747	0.013
3	90.6(-0.3)	82.7(+0.4)	1801395	0.012
1、2	91.0(+0.1)	83.5(+1.2)	1804947	0.018
2、3	90.1(-0.8)	83.2(+0.9)	1804947	0.017
1、3	91.1(+0.2)	82.3(+0.0)	1812595	0.014
1、2、3	**91.6(+0.7)**	**83.9(+1.6)**	**1816147**	**0.016**

N represents the Nth C3 in the Head layer and is replaced with C3BIF.
The bold values in the table emphasize the experimental results of the methods used in this study.

3.3 Ablation experiments of the KIWI-YOLO network

Ablation experiments were conducted on each module to assess the effectiveness of the enhanced KIWI-YOLO network model in kiwifruit flower detection. The experimental results are summarized in Table 3, and the visualization curve is shown in Figure 6.

As depicted in Table 3, the utilization of the FDFF module alone yields a significant improvement in mAP, with $mAP_{0.5}$ surging by 3 points, surpassing the 90% mark. Furthermore, $mAP_{0.5:0.95}$ experiences an increase of nearly 5 points. The Recall and F1 parameters also exhibit a notable improvement of approximately 3 points each. Remarkably, the detection time per image only marginally increases by 0.003 seconds. When solely incorporating the C3BIF module into the original YOLOv5 network, the enhancement is evident but not as pronounced. However, when the FDFF module is combined with the C3BIF module, $mAP_{0.5}$ notably rises to 91.6%, marking a 3.8-point increase compared to the base network. Similarly, Recall improves to 83.9%, reflecting a 5.5-point enhancement. Despite this improvement, the increase in the number of parameters remains minimal. Additionally, the detection time per image also improves by 0.003 seconds. These findings underscore the significant detection enhancements achieved by the improved network at a minimal additional cost.

3.4 Multi-category recognition accuracy analysis

The model distinguishes among four growth stages of kiwifruit flowers that occur during kiwifruit flower pollination. The average accuracies of the improved model were 95%, 89.6%, 93.5%, and 88.2% for the four categories: Female, Bud, Male, and Success, respectively. Compared to the original YOLOv5n network, the improvements were 1.7%, 3.9%, 6.2%, and 3.2%, respectively. A comparison of the results is presented in Table 4.

TABLE 3 Results of ablation experiments.

Baseline	FDFF	C3BIF	$mAP_{0.5}$ (%)	$mAP_{0.5:0.95}$ (%)	Recall (%)	F1-score (%)	Parameters	Time (s)
YOLOv5n			87.8	60.7	78.4	83.4	1764577	0.013
	√		90.8(+3.0)	65.1(+4.4)	82.2(+3.8)	86.1(+2.7)	1790195	0.016
		√	88.6(+0.8)	60.0(-0.7)	79.4(+1.0)	83.7(+0.3)	1790529	0.021
	√	√	**91.6(+3.8)**	**65.8(+5.1)**	**83.9(+5.5)**	**86.5(+3.1)**	**1816147**	**0.016**

$mAP_{0.5}$, $mAP_{0.5:0.95}$, Recall, F1-score, parameter size, and detection time for ablation experiments with the improved model.
The bold values in the table emphasize the experimental results of the methods used in this study.

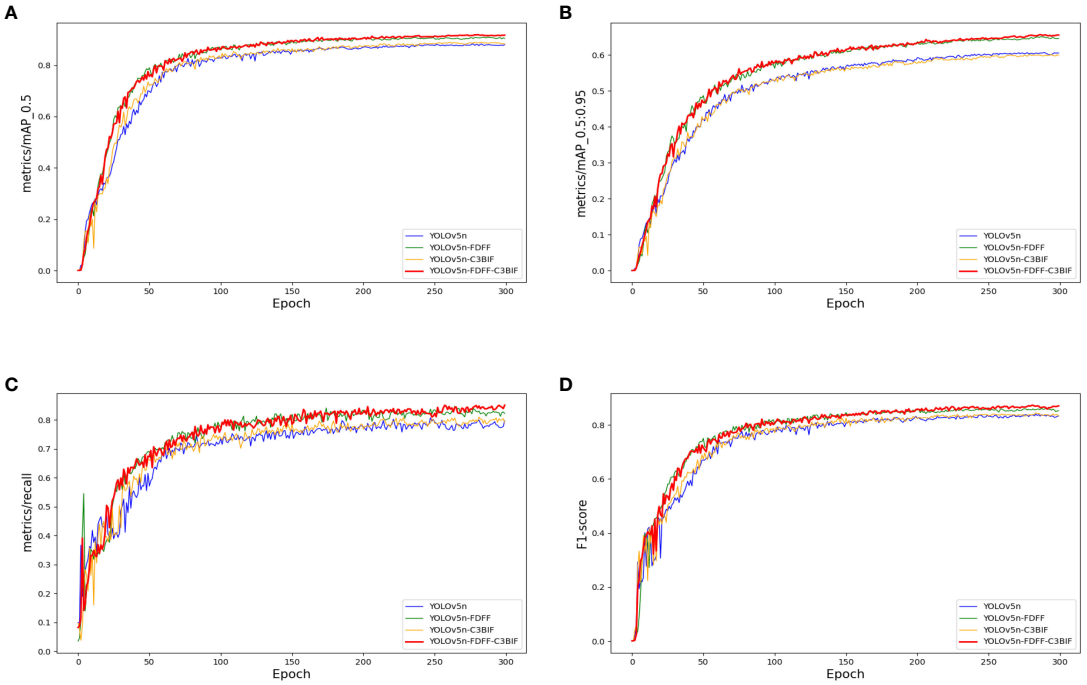


FIGURE 6
Comparison of ablation experimental results. **(A)**. Panels **(A-D)** are the result curves of $mAP_{0.5}$, $mAP_{0.5:0.95}$, recall and F1-score, respectively.

The enhanced model exhibited more noticeable improvements in detecting Bud and Male stages compared to Female and Success. Nevertheless, the average detection accuracy exceeded 93% for both the Female and Male categories. The successful class represents the successfully pollinated kiwifruit flowers, which is challenging for this type of detection. This difficulty arises from the slight morphological differences between this class and Female in the early stages of successful pollination, making them hard to distinguish. Furthermore, in the later stages, when the petals are fully removed, the morphology closely resembles that of Female without petals, resulting in a slightly weaker detection performance. Figure 7 illustrates the detection efficacy under various conditions, including different types, shading scenarios, and variations in lighting intensity.

3.5 Mainstream algorithm comparison

To evaluate the effectiveness of the KIWI-YOLO model proposed in this paper for detecting all stages of kiwifruit flower pollination, we compared its performance with current mainstream target detection models. The chosen algorithms include the two-stage target detection model Faster R-CNN (Ren et al., 2017), the one-stage target detection model SSD (Liu et al., 2016), YOLOv5s, YOLOv5n, YOLOv8s, and YOLOv8n, as well as the excellent lightweight target detection models at this stage, YOLOX-nano (Ge et al., 2021), YOLOX-tiny (Ge et al., 2021), and YOLOv7- tiny (Wang et al., 2023). The performance comparison of these detection models is shown in Table 5.

TABLE 4 Multi-category recognition effect.

Classification	Original YOLOv5	Improved model
Female	93.3%	**95.0%(+1.7)**
Bud	85.7%	**89.6%(+3.9)**
Male	87.3%	**93.5%(+6.2)**
Success	85.0%	**88.2%(+3.2)**

The average accuracy of the four categories Female, Bud, Male, and Success on the original YOLOv5 network and the improved KIWI-YOLO, respectively. The bold values in the table emphasize the experimental results of the methods used in this study.

Through analyzing the test results in this table, it can be concluded that the $mAP_{0.5}$ of the KIWI-YOLO model proposed in this paper is higher than these several detection models, 21.48, and 10.55 percentage points higher than Faster R-CNN, SSD. Compared to the current popular one-stage models, it is 3.8, 0.9, 2.4, 1.9 and 3.5 percentage points higher than YOLOv5n, YOLOv5s, YOLOv7-tiny, YOLOv8s and YOLOv8n, respectively. Compared to YOLOv5s, the superiority in $mAP_{0.5}$ is small, but in the model's size, it can be seen that KIWI-YOLO reduces the number of parameters by almost 75% compared to YOLOv5s. In comparing YOLOv5n, although the number of parameters increased by 0.05M in model parameters, it increased by 3.8% in $mAP_{0.5}$ and 3.01% in F1-score. For the lightweight YOLOX model, KIWI-YOLO is weaker than YOLOX-nano in terms of the advantage of the number of parameters, but $mAP_{0.5}$ improves by 12.48 percentage points and by 7.69 percentage points over YOLOX-tiny.

Among them, the excellent one-stage target detection algorithms YOLOv5n, YOLOv5s, YOLOv7-tiny, YOLOv8s, and YOLOv8n at the present stage are selected and compared with the KIWI-YOLO algorithm in this study and their average accuracy curves and Recall curves are compared with the line graphs as shown in Figure 8.

10.3389/fpls.2024.1415884

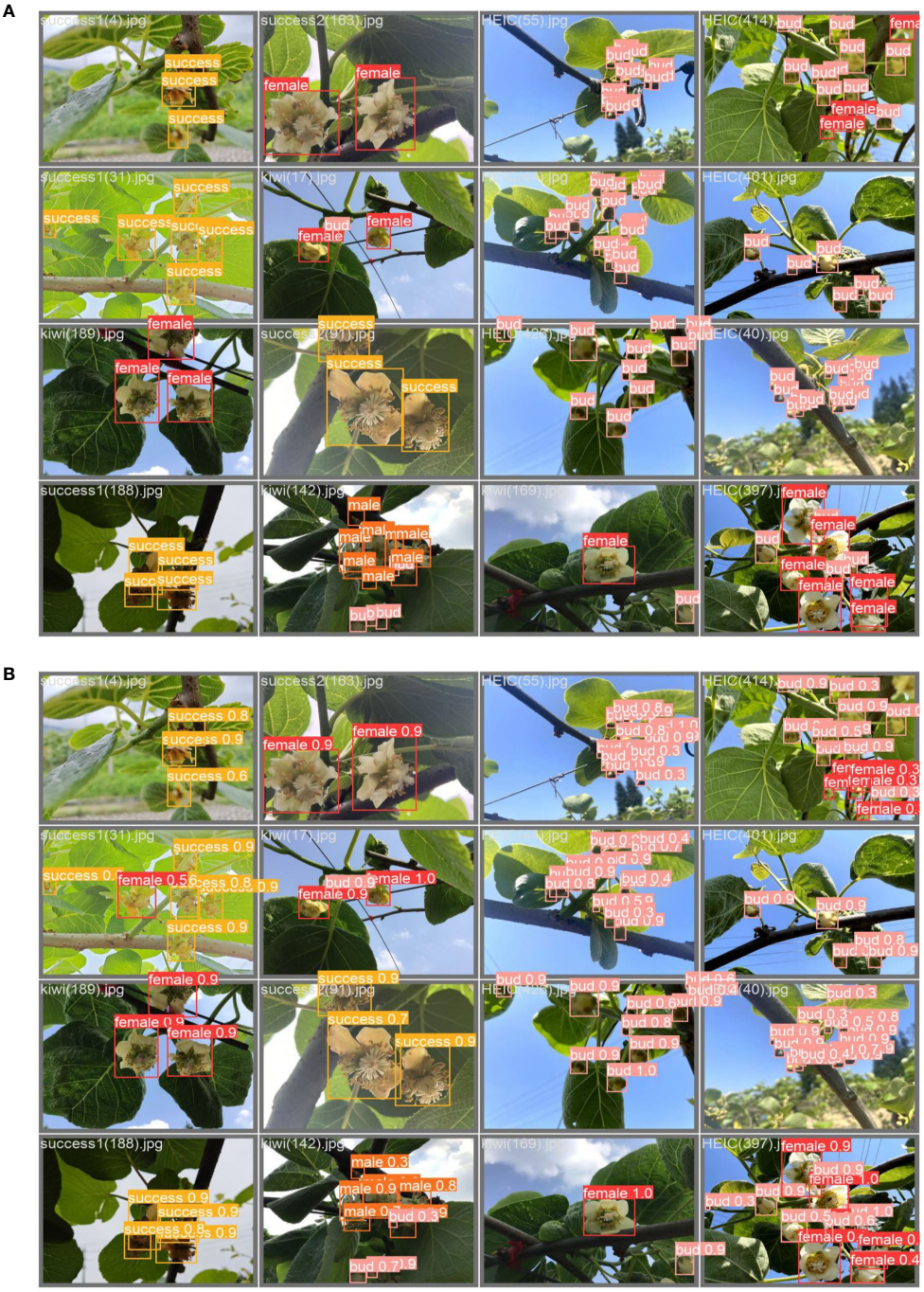


FIGURE 7
Detection effect diagram: Panel **(A)** shows the real label, and panel **(B)** shows the predicted label, i.e., the model detection result.

The experimental results show that the improved model KIWI-YOLO proposed in this study has a certain degree of superiority in terms of comprehensive performance compared with the currently popular target detection models. The model can accurately and quickly realize the detection task of kiwifruit pollination in all stages. In addition, the model can basically meet the requirements of embedded devices in terms of size and is lightweight.

3.6 The Grad-CAM analysis

Grad-CAM (Selvaraju et al., 2017) can visualize the part of the neural network that contributes the most to the prediction result. In this paper, the decision-making process of the neural network is visualized by the Grad-CAM technique, and the help of the improved module on the detection task is illustrated. Figure 9 displays original images of four different categories alongside heat maps generated for

TABLE 5 Results of mainstream algorithm comparison experiment.

Model	$mAP_{0.5}$(%)	F1-score(%)	Parameters	Average precision for each category (%)			
				Female	Bud	Male	Success
Faster R-CNN	70.12	62.25	137098724	81.28	55.87	71.56	71.77
SSD	81.05	78.00	26285486	83.26	77.97	85.31	77.66
YOLOX-nano	79.12	76.00	912159	81.61	78.27	81.47	75.12
YOLOX-tiny	83.91	76.34	5033739	84.94	83.26	87.69	79.75
YOLOv5n	87.80	83.49	1764577	93.30	85.70	87.30	85.00
YOLOv5s	90.70	86.14	7020913	94.80	88.70	91.60	87.90
YOLOv7-tiny	89.20	86.09	6015714	93.90	86.10	90.90	85.70
YOLOv8n	88.10	83.13	3302624	92.80	87.30	88.40	83.80
YOLOv8s	89.70	85.94	11127132	92.90	88.40	91.40	86.00
KIWI-YOLO	**91.60**	**86.50**	**1816147**	**95.00**	**89.60**	**93.50**	**88.20**

$mAP_{0.5}$, F1-score, number of parameters, and average accuracy of each category for ten models on the kiwifruit flower dataset.
The bold values in the table emphasize the experimental results of the methods used in this study.

each improved key layer using Grad-CAM. In these heat maps, the redder the color, the more the region contributes to the model results.

The heat map in Figure 9 shows that through the frequency domain feature fusion in the FDFF module, the network model gives more importance to the target contour but also pays attention to the overall contour of the branches and leaves. This is due to the overall frequency domain features of the image being extracted in the FDFF module, and the context information in the frequency domain is obtained by fusion with the global frequency domain features. From the comparison between columns B and C, it can be seen that the model with FDFF can combine the global frequency domain information, make full use of the context information to understand the relationship between the environment and the detection target, and better identify the target. Observing the D and E columns of Figure 9, it can be seen that before optimization by the C3BIF module, the focus of the network is more dispersed, and there are still some red areas that represent the focus of attention outside the target. After optimization with the C3BIF module, the network is more focused on the target. However, in complex environments, the model still remains weakly focused on the blade due to the similarity of the blade and target colors. From the heat map, the improved KIWI-YOLO model has achieved a more accurate detection of the four flower states during the whole stage of kiwifruit flower pollination.

4 Discussion

At present, in the field of kiwifruit flower recognition, researchers aim to improve the recognition accuracy of related models without considering the actual situation of agricultural environment. Most of the models have problems such as fuzzy categories (Lim et al., 2020; Williams et al., 2020) and excessive size (Li et al., 2022a), which limit the effective application of the model on embedded devices. In this paper, we summarize the current problems of kiwifruit flower recognition into the following three points:(1)The classification of kiwifruit flowers is fuzzy. (2)Existing model size is too large. (3)Small size model accuracy is low. Therefore, this paper has carried out relevant research on these three issues.

In this study, the dataset was manually collected and labeled to solve the fuzzy problem of kiwifruit flower classification. Based on the actual situation of the kiwifruit pollination period, the flowers are mainly divided into four categories: Female(female flower), Male(male flower), Bud(bud), and Success(successful pollination). Aiming to address the problem of the model size being too large, the existing mainstream models are compared. The two-stage detection models, such as Faster R-CNN (Ren et al., 2017) and SSD (Liu et al., 2016), require large memory and are inaccurate in localization. In the one-stage detection model, YOLOv7 (Wang et al., 2023) introduces an efficient layer aggregation network to improve performance. However, it performs poorly on the small target detection task and shows weak generalization ability. As a new product of the YOLO series, YOLOv8 has a complex model structure, and its size is too large to be suitable for embedded devices. YOLOX (Ge et al., 2021) has an advantage in the number of parameters, but performed mediocrely on the kiwifruit flower recognition task. In contrast, YOLOv5 introduces CSPDarknet53 as the backbone network and utilizes the PANet(Path Aggregation Network) structure to enhance feature fusion, demonstrating good accuracy and speed. The KIWI-YOLO model proposed in this paper is based on the lightweight YOLOv5n algorithm, which meets the requirements of embedded devices for model size and maintains speed and accuracy. In this study, the FDFF module is proposed to solve the problem of low recognition accuracy of small-size models. In kiwifruit cultivation, farmers usually distinguish between different types of flowers by observing their morphological characteristics. YOLOv5n has shortcomings in extracting contour information, and the $mAP_{0.5}$ in the experiment is only 87.80%. The FDFF module can extract high-frequency information and fuse spatial domain information, avoid the omission of important information, and improve the lossy process of information extraction in previous studies (Qin et al., 2021; Su et al., 2021;

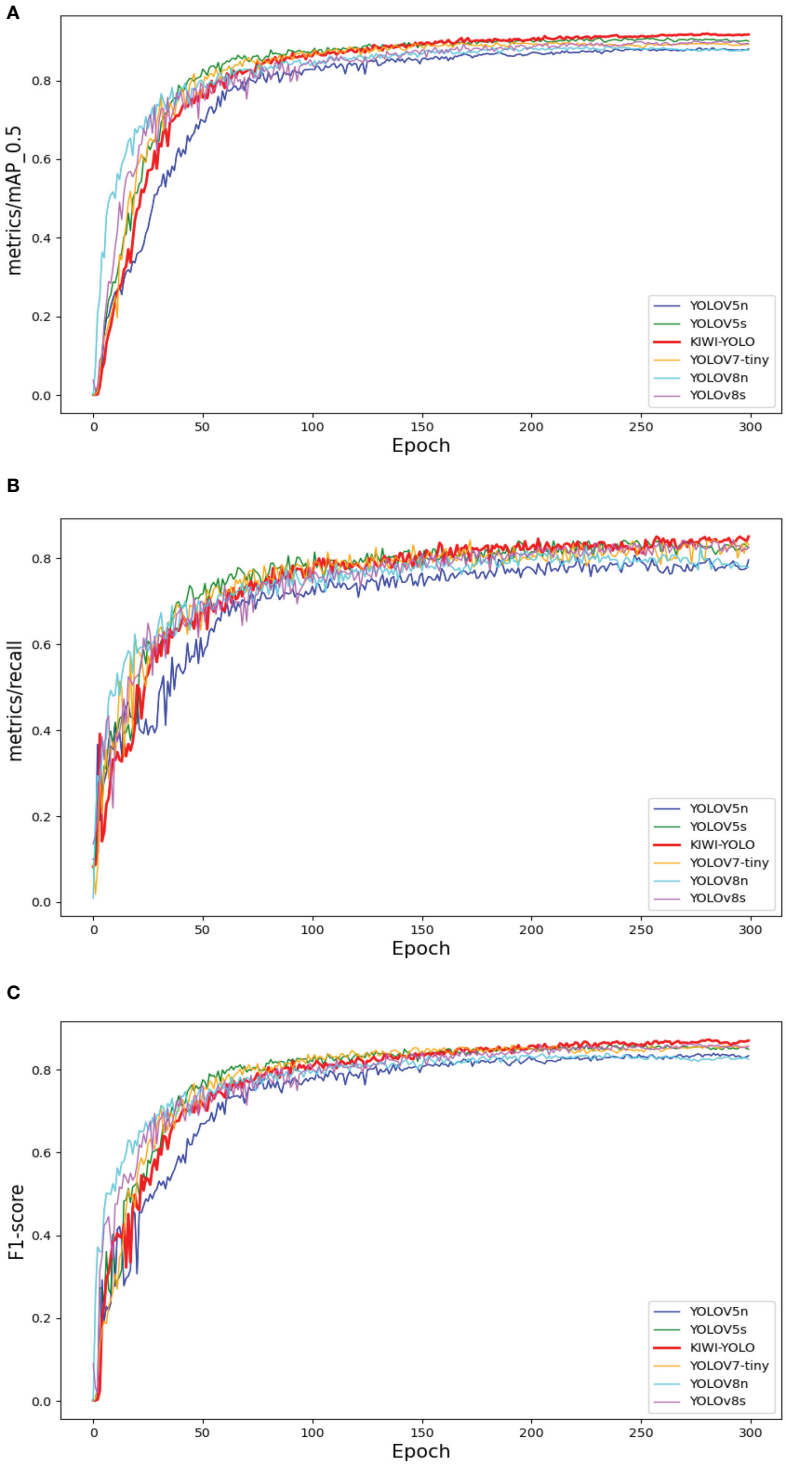


FIGURE 8
Mainstream Algorithm Performance Comparison. Panel **(A)** compares the average precision curves, **(B)** compares the recall, and **(C)** compares the F1-score.

Duan et al., 2023). After adding the FDFF module, the $mAP_{0.5}$ of the model is increased by 3%. In addition, the bud characteristics are significantly different from those of the other three types of kiwifruit flowers, while the accuracy of YOLOv5n was still low, only 85.7%. This study further proposes a C3BIF module combined with the BRA mechanism. The module enhances the recognition accuracy of the model for small targets and improves the detection accuracy of the Bud by 3.9%.

The KIWI-YOLO algorithm proposed in this study shows significant superiority over existing methods. Compared to the baseline model, it achieved a 3.8% improvement in the overall $mAP_{0.5}$ and a 3% improvement in the F1-score, which fully

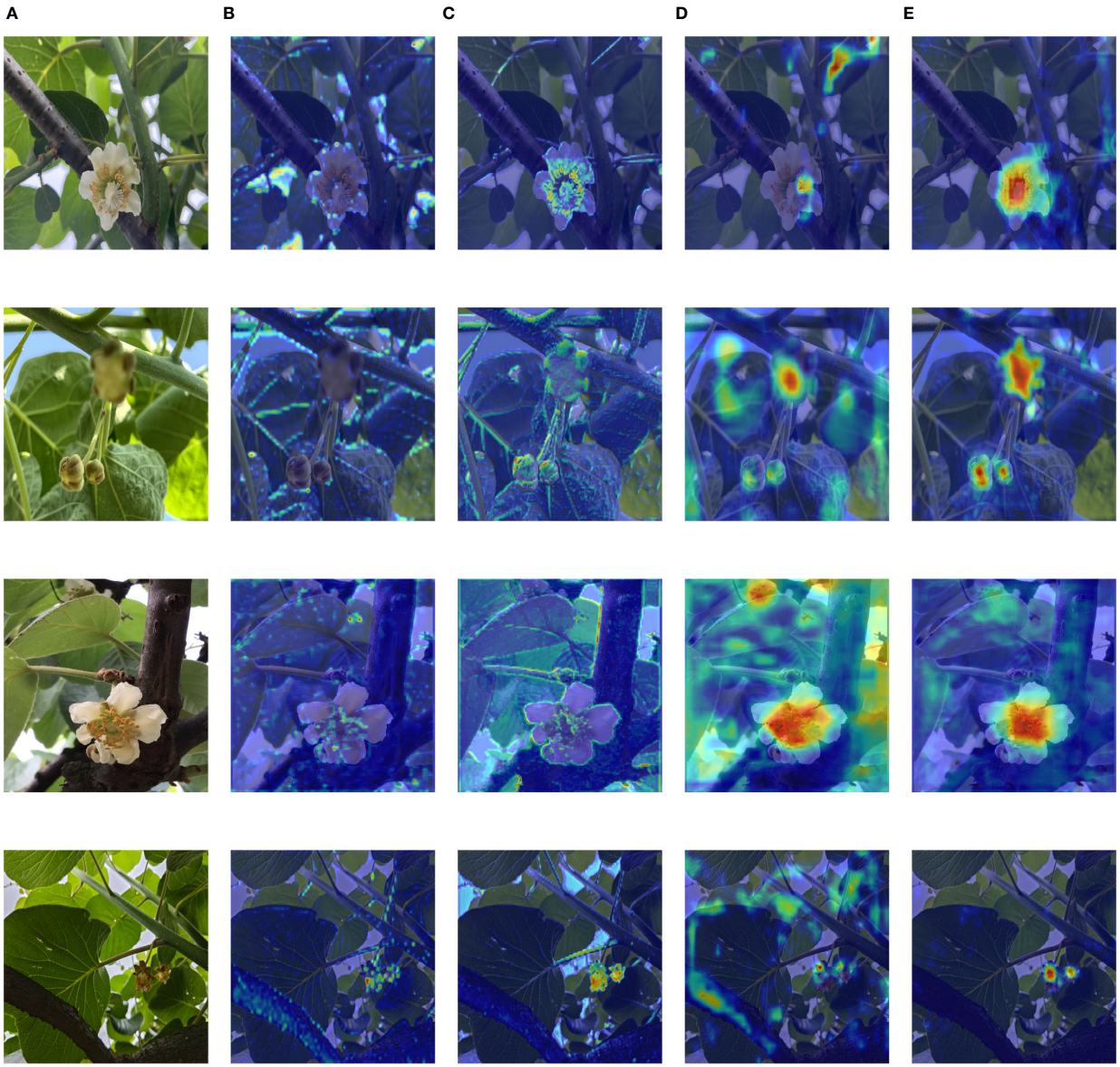


FIGURE 9
Comparison of heat maps. The four rows represent Female, Bud, Male, and Success. Panel **(A)** is the original image. Panel **(B)** is the heat map of the original P4 layer output. Panel **(C)** is the heat map output after adding FDFF. Panel **(D)** is the heat map of the original C3 output. Panel **(E)** is the heat map of the added C3BIF output.

demonstrated its effectiveness. In addition, the number of model parameters is only 1.8M, the average accuracy is as high as 91.6%, and it takes only 0.016 seconds to complete the processing of an image. This high efficiency makes the KIWI-YOLO model very suitable for deployment in automated systems. The model can accurately identify the flowers in the pollination period, thus significantly reducing invalid pollination attempts and saving manpower and resources. Its accurate detection results provide real-time data and accurate pollination timing for orchard managers, avoiding failure and fruit quality decline caused by premature or late pollination. It helps to optimize the production process and resource allocation, improve the efficiency and sustainability of agricultural production, and maintain the stability and health of the orchard ecosystem.

This study has made preliminary progress in improving the classification accuracy of kiwifruit flowers, but there are still a series of challenges. (1) The accuracy of the model in recognizing the Bud and the Success targets is not satisfactory. This is mainly because Bud targets are smaller, denser, and easily overlap. On the other hand, the Success and the Female have similar features and often appear in the same scene, which makes the model prone to misidentification. In the future, this study will continue to solve the difficult problem of small target recognition, explore the addition of a small target detection head in KIWI-YOLO, add a lightweight upsampling CARAF operator, and so on. (2) In terms of the dataset, this study collected images with different light and different shading situations to increase the richness of the data but did not conduct separate experiments for light classification and experiments

with different shading situations. At the same time, the image data failed to cover the whole environment of the actual orchard, and the generalization of the model needs to be improved. To this end, future work will focus on collecting more comprehensive data and conducting specialized classification experiments on illumination and occlusion effects. (3) Existing images contain noise and sharpening problems due to the limitation of acquisition equipment. It is planned to replace professional equipment such as robots or drones to obtain higher-quality image data. (4) Currently, more and more researchers use multi-modal data for object detection. Subsequent studies will consider collecting multi-modal data on kiwifruit and test the fusion effect of the FDFF module on modal data.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

FP: Conceptualization, Writing – original draft, Writing – review & editing. MH: Data curation, Validation, Writing – original draft, Writing – review & editing. XD: Funding acquisition, Writing – review & editing. BZ: Data curation, Validation, Writing – review & editing. PX: Data curation, Writing – review & editing. LJ: Writing – review & editing. XZ: Supervision, Writing – review & editing. DH: Data curation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the Science and Technology Planning Project of Ya'an Science and Technology Bureau under Grant No. 22ZDYFZF0006.

Acknowledgments

Thanks to Luyu Shuai, Xu Wang and Peng Chen for their advice throughout the research process.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bochkovskiy, A., Wang, C.-Y., and Liao, H. (2020). YOLOv4: Optimal speed and accuracy of object detection. *arXiv [Preprint]*. doi: 10.48550/arXiv.2004.10934

Castro, H., Siopa, C., Casais, V., Castro, M., Loureiro, J., Gaspar, H., et al. (2021). Pollination as a key management tool in crop production: Kiwifruit orchards as a study case. *Scientia Hortic.* 290, 110533. doi: 10.1016/j.scienta.2021.110533

Chen, H., Lan, Y., K Fritz, B., Clint Hoffmann, W., and Liu, S. (2021). Review of agricultural spraying technologies for plant protection using unmanned aerial vehicle (UAV). *Int. J. Agric. Biol. Eng.* 14, 38–49. doi: 10.25165/j.ijabe.20211401.5714

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv [Preprint]*. doi: 10.48550/arXiv.2010.11929

Duan, C., Hu, B., Liu, W., Ma, T., Ma, Q., and Wang, H. (2023). Infrared small target detection method based on frequency domain clutter suppression and spatial feature extraction. *IEEE Access* 11, 85549–85560. doi: 10.1109/ACCESS.2023.3303486

Gai, R., Chen, N., and Yuan, H. (2023). A detection algorithm for cherry fruits based on the improved YOLO-v4 model. *Neural Comput. Applic* 35, 13895–13906. doi: 10.1007/s00521-021-06029-z

Gao, C., He, L., Fang, W., Wu, Z., Jiang, H., Li, R., et al. (2023a). A novel pollination robot for kiwifruit flower based on preferential flowers selection and precisely target. *Comput. Electron. Agric.* 207, 107762. doi: 10.1016/j.compag.2023.107762

Gao, X., Tang, Z., Deng, Y., Hu, S., Zhao, H., and Zhou, G. (2023b). HSSNet: A end-to-end network for detecting tiny targets of apple leaf diseases in complex backgrounds. *Plants* 12, 2806. doi: 10.3390/plants12152806

Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. (2021). YOLOX: Exceeding YOLO Series in 2021. *arXiv* [Preprint]. doi: 10.48550/arXiv.2107.08430

Gökdoğan, O. (2022). Energy and Economic efficiency of kiwi fruit production in Turkey: A case study from Mersin Province. *Erwerbs-Obstbau* 64, 55–60. doi: 10.1007/s10341-021-00610-5

Hasan, A. S. M. M., Sohel, F., Diepeveen, D., Laga, H., and Jones, M. G. K. (2021). A survey of deep learning techniques for weed detection from images. *Comput. Electron. Agric.* 184, 106067. doi: 10.1016/j.compag.2021.106067

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv* [Preprint]. doi: 10.48550/arXiv.1704.04861

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (Salt Lake City, UT, USA: IEEE), 7132–7141. doi: 10.1109/CVPR.2018.00745

Li, G., Fu, L., Gao, C., Fang, W., Zhao, G., Shi, F., et al. (2022a). Multi-class detection of kiwifruit flower and its distribution identification in orchard based on YOLOv5l and Euclidean distance. *Comput. Electron. Agric.* 201, 107342. doi: 10.1016/j.compag.2022.107342

Li, G., Suo, R., Zhao, G., Gao, C., Fu, L., Shi, F., et al. (2022b). Real-time detection of kiwifruit flower and bud simultaneously in orchard using YOLOv4 for robotic pollination. *Comput. Electron. Agric.* 193, 106641. doi: 10.1016/j.compag.2021.106641

Lim, J., Ahn, H. S., Nejati, M., Bell, J., Williams, H., and MacDonald, B. A. (2020). Deep neural network based real-time kiwi fruit flower detection in an orchard environment. *arXiv* [Preprint]. doi: 10.48550/arXiv.2006.04343

Liu, M., Khan, A., Wang, Z., Liu, Y., Yang, G., Deng, Y., et al. (2019). Aptasensors for pesticide detection. *Biosensors Bioelectronics* 130, 174–184. doi: 10.1016/j.bios.2019.01.006

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). "SSD: single shot multiBox detector," in *Computer Vision – ECCV 2016* (Springer, Cham), 21–37. doi: 10.1007/978-3-319-46448-0_2

Lu, S., Chen, W., Zhang, X., and Karkee, M. (2022). Canopy-attention-YOLOv4-based immature/mature apple fruit detection on dense-foliage tree architectures for early crop load estimation. *Comput. Electron. Agric.* 193, 106696. doi: 10.1016/j.compag.2022.106696

Mithra, S., and Nagamalleswari, T. Y. J. (2023). Cucurbitaceous family flower inferencing using deep transfer learning approaches: CuCuFlower UAV imagery data. *Soft Comput.* 27, 8345–8356. doi: 10.1007/s00500-023-08186-w

Qi, J., Liu, X., Liu, K., Xu, F., Guo, H., Tian, X., et al. (2022). An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease. *Comput. Electron. Agric.* 194, 106780. doi: 10.1016/j.compag.2022.106780

Qin, Z., Zhang, P., Wu, F., and Li, X. (2021). "FcaNet: Frequency channel attention networks," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. (Montreal, QC, Canada: IEEE), 763–772. doi: 10.1109/ICCV48922.2021.00082

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Las Vegas, NV, USA: IEEE), 779–788. doi: 10.1109/CVPR.2016.91

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). "Grad-CAM: visual explanations from deep networks via gradient-based localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*. (Honolulu, HI, USA: IEEE), 618–626. doi: 10.1109/ICCV.2017.74

Su, B., Liu, J., Su, X., Luo, B., and Wang, Q. (2021). CFCANet: A complete frequency channel attention network for SAR image scene classification. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 14, 11750–11763. doi: 10.1109/JSTARS.2021.3125107

Tian, Y., Wang, S., Li, E., Yang, G., Liang, Z., and Tan, M. (2023). MD-YOLO: Multi-scale Dense YOLO for small target pest detection. *Comput. Electron. Agric.* 213, 108233. doi: 10.1016/j.compag.2023.108233

Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30, 5998–6008.

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2023). "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Vancouver, BC, Canada: IEEE), 7464–7475. doi: 10.1109/CVPR52729.2023.00721

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. (2020). "ECA-net: Efficient channel attention for deep convolutional neural networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Seattle, WA, USA: IEEE), 11531–11539. doi: 10.1109/CVPR42600.2020.01155

Wang, Y., Xun, Y., Wang, Q., and Yang, Z. (2022b). Review of smart robots for fruit and vegetable picking in agriculture. *Int. J. Agric. Biol. Eng.* 15, 33–54. doi: 10.25165/j.ijabe.20221501.7232

Wang, L., Zhao, Y., Liu, S., Li, Y., Chen, S., and Lan, Y. (2022a). Precision detection of dense plums in orchards using the improved YOLOv4 model. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.839269

Williams, H., Nejati, M., Hussein, S., Penhall, N., Lim, J. Y., Jones, M. H., et al. (2020). Autonomous pollination of individual kiwifruit flowers: Toward a robotic kiwifruit pollinator. *J. Field Robotics* 37, 246–262. doi: 10.1002/rob.21861

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). "CBAM: convolutional block attention module," in *Computer Vision – ECCV 2018*. Eds. V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss (Springer International Publishing, Cham), 3–19. doi: 10.1007/978-3-030-01234-2_1

Wu, D., Lv, S., Jiang, M., and Song, H. (2020). Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Comput. Electron. Agric.* 178, 105742. doi: 10.1016/j.compag.2020.105742

Xu, T., Qi, X., Lin, S., Zhang, Y., Ge, Y., Li, Z., et al. (2022). A neural network structure with attention mechanism and additional feature fusion layer for tomato flowering phase detection in pollination robots. *Machines* 10, 1076. doi: 10.3390/machines10111076

Yang, Z., Feng, H., Ruan, Y., and Weng, X. (2023b). Tea tree pest detection algorithm based on improved yolov7-tiny. *Agriculture* 13, 1031. doi: 10.3390/agriculture13051031

Yang, M., Lyu, H., Zhao, Y., Sun, Y., Pan, H., Sun, Q., et al. (2023a). Delivery of pollen to forsythia flower pistils autonomously and precisely using a robot arm. *Comput. Electron. Agric.* 214, 108274. doi: 10.1016/j.compag.2023.108274

Yépez-Ponce, D. F., Salcedo, J. V., Rosero-Montalvo, P. D., and Sanchis, J. (2023). Mobile robotics in smart farming: current trends and applications. *Front. Artif. Intell.* 6. doi: 10.3389/frai.2023.1213330

Yu, K., Tang, G., Chen, W., Hu, S., Li, Y., and Gong, H. (2023). MobileNet-YOLO v5s: an improved lightweight method for real-time detection of sugarcane stem nodes in complex natural environments. *IEEE Access* 11, 104070–104083. doi: 10.1109/ACCESS.2023.3317951

Zeng, T., Li, S., Song, Q., Zhong, F., and Wei, X. (2023). Lightweight tomato real-time detection method based on improved YOLO and mobile deployment. *Comput. Electron. Agric.* 205, 107625. doi: 10.1016/j.compag.2023.107625

Zhang, Y., Yin, Y., and Shao, Z. (2023). An enhanced target detection algorithm for maritime search and rescue based on aerial images. *Remote Sens.* 15, 4818. doi: 10.3390/rs15194818

Zhu, L., Wang, X., Ke, Z., Zhang, W., and Lau, R. (2023). "BiFormer: vision transformer with bi-level routing attention," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Vancouver, BC, Canada: IEEE), 10323–10333. doi: 10.1109/CVPR52729.2023.00995

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 25 July 2024
DOI 10.3389/fpls.2024.1407839

Check for updates

OPEN ACCESS

EDITED BY
Yuntao Ma,
China Agricultural University, China

REVIEWED BY
Orly Enrique Apolo-Apolo,
KU Leuven, Belgium
Yingpu Che,
Chinese Academy of Agricultural Sciences (CAAS), China

*CORRESPONDENCE
Jiong Mu
jmu@sicau.edu.cn

RECEIVED 27 March 2024
ACCEPTED 11 July 2024
PUBLISHED 25 July 2024

CITATION
Wu X, Tang R, Mu J, Niu Y, Xu Z and Chen Z (2024) A lightweight grape detection model in natural environments based on an enhanced YOLOv8 framework.
Front. Plant Sci. 15:1407839.
doi: 10.3389/fpls.2024.1407839

COPYRIGHT
© 2024 Wu, Tang, Mu, Niu, Xu and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A lightweight grape detection model in natural environments based on an enhanced YOLOv8 framework

Xinyu Wu[1,2], Rong Tang[1,2], Jiong Mu[1,2]*, Yupeng Niu[1], Zihan Xu[1] and Ziao Chen[1]

[1]College of Information Engineering, Sichuan Agricultural University, Ya'an, China, [2]Sichuan Key Laboratory of Agricultural Information Engineering, College of Information Engineering, Sichuan Agricultural University, Ya'an, China

Grapefruit and stem detection play a crucial role in automated grape harvesting. However, the dense arrangement of fruits in vineyards and the similarity in color between grape stems and branches pose challenges, often leading to missed or false detections in most existing models. Furthermore, these models' substantial parameters and computational demands result in slow detection speeds and difficulty deploying them on mobile devices. Therefore, we propose a lightweight TiGra-YOLOv8 model based on YOLOv8n. Initially, we integrated the Attentional Scale Fusion (ASF) module into the Neck, enhancing the network's ability to extract grape features in dense orchards. Subsequently, we employed Adaptive Training Sample Selection (ATSS) as the label-matching strategy to improve the quality of positive samples and address the challenge of detecting grape stems with similar colors. We then utilized the Weighted Interpolation of Sequential Evidence for Intersection over Union (Wise-IoU) loss function to overcome the limitations of CIoU, which does not consider the geometric attributes of targets, thereby enhancing detection efficiency. Finally, the model's size was reduced through channel pruning. The results indicate that the TiGra-YOLOv8 model's mAP(0.5) increased by 3.33% compared to YOLOv8n, with a 7.49% improvement in detection speed (FPS), a 52.19% reduction in parameter count, and a 51.72% decrease in computational demand, while also reducing the model size by 45.76%. The TiGra-YOLOv8 model not only improves the detection accuracy for dense and challenging targets but also reduces model parameters and speeds up detection, offering significant benefits for grape detection.

KEYWORDS

YOLOv8, grape detection, computer vision, intelligent agriculture, lightweight model

1 Introduction

Grapes are heralded as the "queen of fruits" and possess considerable economic value (Chen et al., 2023). Owing to their nutritional and financial benefits, the area under grape cultivation and the scale of production have been expanding annually (Sun et al., 2022), with the global output reaching 75 million tons (Moro et al., 2021). Grape harvesting is a critical component of the grape industry's large-scale development. Harvesting is predominantly manual, a process that requires significant labor, incurs high costs and is relatively slow (Bac et al., 2014). Estimates suggest that manually harvesting a ton of grapes costs three times more than mechanical harvesting. Hence, compared to manual picking, mechanical harvesting offers unparalleled advantages (Lin et al., 2023).

As a crucial branch of artificial intelligence, computer vision has enabled many challenging functions for traditional methods. Since their introduction, convolutional neural network models trained using deep learning methods, such as Fast-R-CNN (Girshick, 2015), Faster-R-CNN (Ren et al., 2015), YOLOv4 (Bochkovskiy et al., 2020), and SSD (Liu et al., 2016), have achieved remarkable success in computer vision. Consequently, methods based on computer vision are receiving increasing attention in technical research across various fields, particularly in remote sensing (Bai et al., 2024), transportation (Dilek and Dener, 2023), and agriculture (Tian et al., 2020). Target detection plays an essential role in harvesting robots, where the rapid and accurate identification and localization of fruits and pedicels are crucial to achieving automated harvesting (Luo et al., 2016). Although the YOLOv8 model shows improved performance over previous YOLO models in metrics like mAP, its complex structure and high computational demand render it unsuitable for deployment on endpoint devices. Therefore, achieving a lightweight structure of the YOLOv8 model without compromising its performance has become an important research topic to advance grape automated harvesting technology.

Researchers have applied traditional detection methods to detect objects in natural environments. Lin et al. (2020) proposed a novel fruit detection technique in natural settings based on a support vector machine classifier supported by color and texture features. Chaivivatrakul and Dailey (2014) introduced a plant green fruit detection technology based on texture analysis, assessing interest features in pineapples and bitter gourds. Luo et al. (2018) utilized a segmentation algorithm based on K-means clustering and practical color components to detect the cutting points of stem peduncles in overlap-ping grape clusters within unstructured vineyards. Pérez-Zavala et al. (2018) developed a method for grape berry recognition and grape bunch detection using a visible spectrum camera. However, In the face of complex natural environments, traditional target detection methods struggle to adapt to changing conditions, such as variations in lighting, occlusion and covering among detection targets, and dense distribution of objects. This situation can lead to reduced accuracy in final detection and identification, adversely affecting the recognition of grape berries, pedicels and the determination of harvesting locations.

Amid the rapid advancement and widespread adoption of deep learning, numerous methods from this domain have been applied in agricultural contexts, primarily categorized into semantic segmentation and object detection approaches (Lu and Luo, 2022). For example, within the scope of semantic segmentation, Yu et al. (2019) introduced the Mask-RCNN algorithm to generate mature fruit mask images, enhancing the efficacy of machine vision in detecting fruits for strawberry harvesting robots. Luo et al. (2015) employ improved clustering image segmentation and point-line minimum distance constraints, achieving an accuracy rate of 88.33%. Zhou et al. (2023) used an enhanced YOLACT++ model for segmenting critical structures of grapes, with the success rates of the picking point localization methods increasing by 10.95 and 81.75 percentage points. Liu et al. (2020) applied the Chan-Vese model for iterative recognition of grape clusters, attaining an average accuracy of 89.71% and a success rate of 90.91%. However, in practical applications, the achievements of target detection in natural environments based on semantic segmentation are relatively few. This arises from the extensive time needed to prepare training sets, the models' inability to perform in real-time, and their significant size, which complicates deploying them on embedded machines in contemporary agriculture.

Applying convolutional neural networks (CNNs) for object detection for this task typically falls into two categories. The first category encompasses two-stage object detection methods: R-CNN, Fast R-CNN, and Faster R-CNN. These approaches extract object regions and then perform CNN classification and regression on those regions, constituting a detection strategy based on region proposals. For instance, Gao et al. (2020) introduced a SNAP system based on Faster R-CNN for multi-category apple detection, achieving a frame rate of 0.241s. Behera et al. (2021) presented an FR-CNN algorithm that utilizes Intersection over Union (IoU) for plant fruit prediction, reaching an accuracy of 89% in fruit yield estimation. The second category involves one-stage object detection methods, such as YOLO and SSD. Ma et al. (2024) employed an enhanced YOLOv8 model for real-time detection of apples in orchard environments, achieving an average precision of 91.4%. Wu et al. (2023) developed a stem localization method for grapes from a top-down perspective using a lightweight YOLOv5m detection algorithm based on HRNet, with a detection speed of 7.7 frames per second. It is evident that, compared to the CNN series, the YOLO series achieves comparable performance and higher computational efficiency in crop fruit detection.

Although traditional deep learning methods have improved detection accuracy, their networks are complex, necessitating substantial storage space, and are thus generally unsuitable for deployment on embedded devices. YOLOv8 is one of the fundamental models in the YOLO series, offering a new state-of-the-art model. However, due to the complexity of orchard lighting and diverse backgrounds, the YOLOv8 detection algorithm still needs to improve its object recognition. Moreover, due to the relatively large size of the YOLOv8 model, deploying it on embedded devices like harvesting robots remains challenging. To address issues of large model size, slow detection speed, and low accuracy, this paper introduces the TiGra-YOLOv8 model, a

lightweight and high-precision solution designed to improve the detection efficiency of grapefruit and grape stem in the natural environment, achieving a lightweight model.

2 Materials and methods

2.1 Dataset

The grape dataset studied in this paper was collected on September 9, 2023, at the Four Seasons Vineyard in Chengdu, Sichuan Province, featuring the "Crimson" variety of grapes. We captured the images with a Canon camera, positioning the lens 0.5m to 1.5m away from the grapes. Moreover, we acquired all grape images under natural light from 12:00 to 4:00 PM. The dataset was collected in the vineyard's natural environment, having different environmental conditions.

We annotated the dataset using the professional annotation software LabelImg, marking grapefruits and grape stems with rectangular boxes (Russell et al., 2008). We labeled the fruits "grape" and the stems "grape root." The collected dataset, after a selection process, consisted of 913 images. However, when the number of training samples is insufficient, it will often lead to overfitting of the model, which will further affect the detection effect of the model. Therefore, we expanded the original data set to 2500 through data enhancement technology, including image rotation, adding noise, flipping, and changing brightness. The data enhancement methods are shown in Figure 1. After image quality screening, we finally obtained 2411 data sets as training samples, which were divided according to the ratio of 7:2:1, including 1687 training sets, 482 verification sets, and 242 test sets. The original resolution of the images used for training the model was 5184x3456 pixels. To optimize the computational efficiency and ensure compatibility with the YOLOv8 model, we resized the images to 640x640 pixels. It helps standardize the input size and balance the trade-off between computational load and model performance.

2.2 TiGra-YOLOv8

YOLOv8 is an end-to-end convolutional neural network based on deep learning, comprising three main components: Backbone, Neck, and Head. Unlike the Anchor-based approach used in earlier models of the YOLO series, YOLOv8 introduces an Anchor-free method that achieves higher detection accuracy and speed (Lou et al., 2023). In the Backbone of YOLOv8, the model extracts image features through pooling and convolution, reducing the parameters and computational load (del Pilar Martínez-Diz et al., 2020). After upsampling and feature fusion, the Neck section sends three output results to the Head layer for loss function calculation. For the matching strategy, YOLOv8 employs the TaskAlignedAssigner method (Feng et al., 2021), which matches positive and negative samples for loss calculation based on the weighted results of classification and regression scores.

This paper presents an enhanced TiGra-YOLOv8 model based on the YOLOv8n model. As illustrated in Figure 2, we have

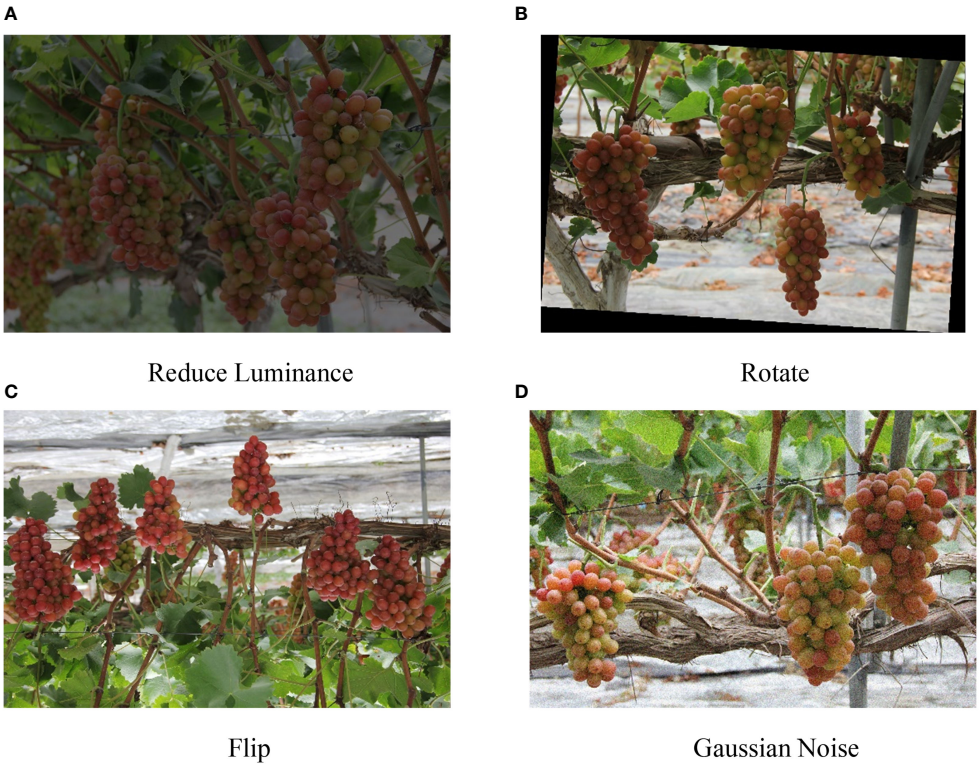


FIGURE 1
Data enhancement methods. **(A)** Reduce Luminance; **(B)** Rotate; **(C)** Flip; **(D)** Gaussian Noise.

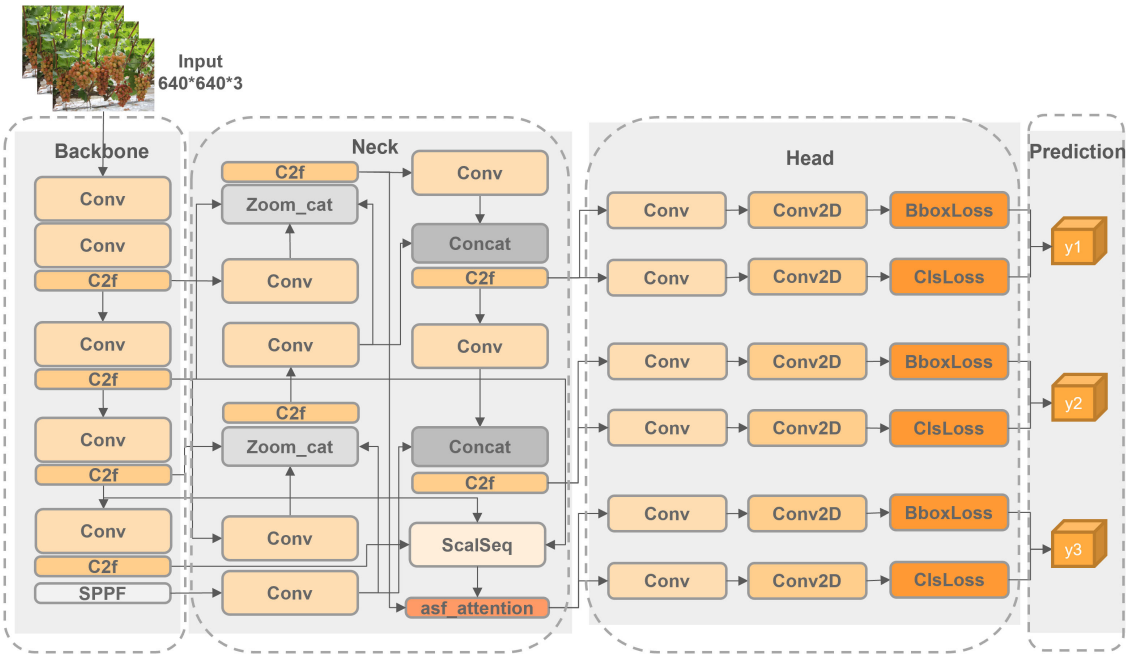


FIGURE 2
TiGra-YOLOv8 network structure.

integrated the Attentional Scale Sequence Fusion module from the weighted Bi-Directional Feature Pyramid Network ASF-YOLO into the YOLO framework to improve the Neck section of YOLOv8 (Kang et al., 2023). We have adopted the Adaptive Training Sample Selection (ATSS) method for adaptive positive and negative sample selection to refine the TaskAlignedAssigner label-matching algorithm of YOLOv8 (Zhang et al., 2020). Additionally, we chose the boundary box loss based on a dynamic non-monotonic focusing mechanism (Wise-IoU) to replace the original CIoU loss function of YOLOv8 (Tong et al., 2023). Finally, by employing the Random channel pruning method, we have reduced the model's size and parameter count while ensuring accuracy, achieving a lightweight model deployment (Liu et al., 2022).

2.3 Experimental detail

All models were trained under identical hardware conditions and with the same initial training parameters to ensure fairness in experimental comparisons. Mosaic data augmentation was turned off once in the last ten training cycles to better fine-tune model parameters. The specific initial training parameters are listed in Table 1. The model accuracy reported in this paper is based on a single training session. To ensure the reproducibility of the results, a random seed was set during the training process. Hyperparameter tuning was conducted using the validation set, and the test set's final accuracy was evaluated.

The experimental environment for this study included an RTX 3090 (24G) graphics card; Python version 3.8.16; CUDA version 11.7; Torch version 1.13.1+cu117; and TorchVision version 0.14.1+cu117.

2.4 Evaluation Index

Mean Average Precision (mAP) is the average of the Average Precision (AP) across all categories, where AP represents the average precision for each category. The formula for calculating mAP is mAP = 1/m∑AP(i), where m is the number of categories and AP(i) is the average precision of the i-th category.

$$P = \frac{TP}{TP + FP} \times 100\%$$

$$R = \frac{TP}{TP + FN} \times 100\%$$

$$AP = \int_0^1 P(R)dR \times 100\%$$

$$mAP = \frac{1}{k}\sum_{i=1}^{k} AP_i \times 100\%$$

mAP(0.5) and mAP(0.75) refer to the average precision of each category at Intersection over Union (IoU) thresholds more significant than 0.5 and 0.75, respectively. mAP(0.5:0.95) denotes the mAP across different IoU thresholds, ranging from 0.5 to 0.95, with a step size 0.05. Loc denotes localization errors, Dupe indicates duplicate detection errors, and FalsePos represents overall error metrics, including duplicate detection, incorrect classification, localization errors, and background confusion. Model Size, Parameters, and GFLOPS, respectively, refer to the storage size, number of parameters, and

TABLE 1 Initial training parameters.

Parameter	Form/Value
epochs	150
batch	16
close_mosaic	10
workers	8
optimizer	SGD
dropout	0.3
seed	42
lr0	0.01
lrf	0.01

computational cost of the model, commonly used to evaluate the size of a network model. FPS, or frames per second, measures the number of images that can be detected per second and is used to assess the detection speed of the network model.

3 Model Improvement

3.1 Neck

3.1.1 ASF network structure

Integrating the Attentional Scale Sequence Fusion (ASF) module from ASF-YOLO into the YOLO framework to enhance the Neck component of YOLOv8 enables a more effective combination of high-dimensional information from deep feature maps with detailed information from shallow feature maps. The ASF feature extraction network consists of Scale Sequence Feature Fusion (SSFF), Triple Feature Coding (TPM), and Channel and Position Attention Model (CPAM) modules. Its network structure is depicted in Figure 3.

During forward propagation, the SSFF module merges feature maps from different spatial scales, preserving scale-invariant features and enhancing the representation of multi-scale information. This process ensures that the model can effectively detect objects of varying sizes and shapes. The TFE module further processes these feature maps by adjusting their scales and combining them into a unified feature map. This helps in accurately identifying small and overlapping targets by providing a comprehensive multi-scale context. The CPAM module applies attention mechanisms to focus on relevant spatial and channel-wise information, enhancing the localization and classification of small targets like grape stems. This selective attention improves the model's ability to differentiate between similar objects and detect fine details.

During backward propagation, the enhanced feature representation and selective attention mechanisms provided by ASF-YOLO result in more stable and informative gradients during backward propagation. This leads to more efficient parameter updates and faster convergence, Stabilizing the gradient flow. The multi-scale fusion and attention mechanisms ensure that the model learns robust and discriminative features, which are essential for accurate detection. The gradients calculated for these features are more informative, leading to better weight adjustments and improved model performance.

3.1.2 SSFF module

The SSFF module effectively merges feature maps from different spatial scales (e.g., layers P3, P4, P5), capturing scale-invariant

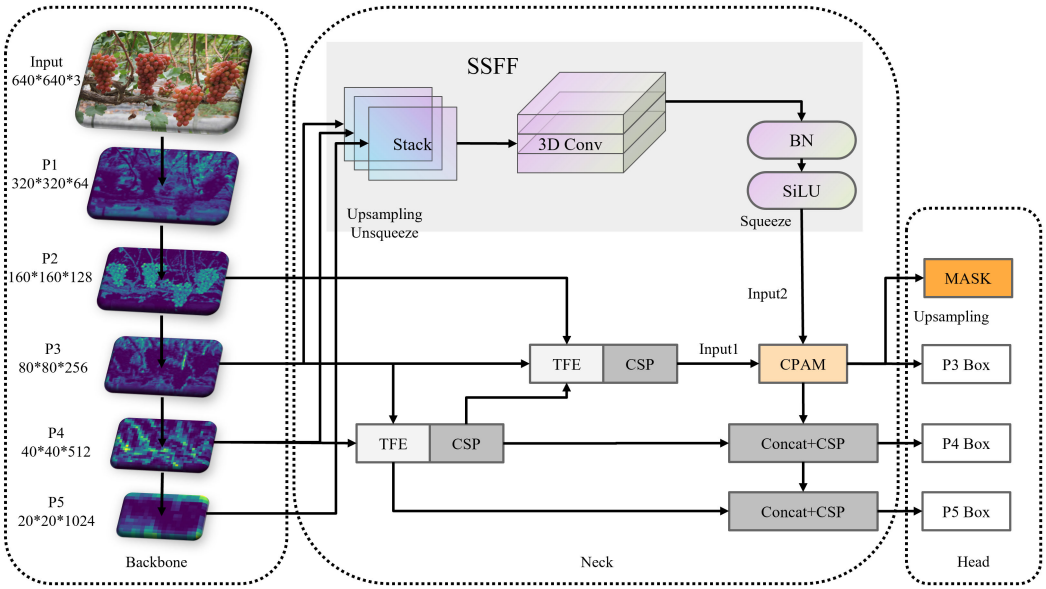


FIGURE 3
ASF module structure.

features that remain consistent despite variations in size or shape. This capability is crucial for detecting grapes of various sizes within a cluster, as it enables the model to recognize both small individual grapes and larger grape clusters.

The grape-scaled images inputted into the SSFF can be obtained by the following method:

$$F_0(w,h) = G_0(w,h) \times f(w,h)$$

$$G_\sigma(w,h) = \frac{1}{2\pi\sigma^2} e^{-(\omega^2+h^2)/2\sigma^2}$$

where $f(w,h)$denotes a 2D input image with width w and height h. $F_0(w,h)$ is generated by smoothing under a 2D Gaussian filter through convolutions. σ represents the standard deviation parameter of the 2D Gaussian filter used for convolution.

3.1.3 TFE module

The TFE module enhances the detection of small, densely overlapping objects like individual grapes within a cluster by examining and comparing image variations across different scales. By magnifying and combining feature maps from large, medium, and minor scales, the TFE module ensures that even the smallest details are captured and emphasized. The module adjusts the scales of input feature maps (large, medium, and small) to a standard scale, allowing for effective fusion and comparison. This adaptation aids the model in recognizing grapes irrespective of their size, leading to improved detection accuracy for both small and large grape clusters. The TFE module is illustrated in Figure 4.

$$F_{TFE} = Concat(F_l, F_m, F_s)$$

F_{TFE} represents the output feature map of the TFE module, with F_l, F_m, F_s enoting the large, medium, and small-sized feature maps, respectively. The information output by the TFE is integrated into each feature map branch through the PANet structure, combined with the multi-scale information output by the SSFF module into the P3 branch for subsequent attention network feature extraction.

3.1.4 CPAM module

The CPAM module employs channel attention mechanisms to focus on the most relevant channels in the feature maps. By prioritizing channels containing critical information about grape clusters and fruits, the model can more accurately identify and classify these objects. This is particularly useful for distinguishing closely packed grapes from background elements. The CPAM module also applies spatial attention mechanisms to highlight significant spatial regions within the feature maps. This enhances the model's ability to localize small targets like individual grapes and grape stems, improving detection precision and reducing false positives. By integrating inputs from the TFE and SSFF modules, the CPAM module ensures that both high-order multi-scale features and detailed spatial information are utilized. This comprehensive integration enables the model to effectively detect and differentiate grape clusters and individual grapes in complex and cluttered scenes. The structure of CPAM is shown in Figure 5.

3.2 Label-matching policy

YOLOv8 employs the TaskAlignedAssigner label-matching strategy, which selects positive samples based on a weighted score of classification and regression scores. The TiGra-YOLOv8 model this paper introduces utilizes the Adaptive Training Sample Selection (ATSS) label-matching algorithm. As illustrated in Figure 6, blue boxes represent the ground-truth boxes (gt), and red boxes represent the anchor boxes (anchor). ATSS calculates the distances x1, x2, x3 between each anchor and its center point if a feature map generates three anchors for grape stems and fruits. If x1< x2< x3, it selects the anchors corresponding to the shorter distances x1 and x2 as candidate positive samples. For L levels of feature maps, L*2 candidate positive samples are chosen. Subsequently, the IoU between each candidate positive sample and the gt is calculated, followed by the calculation of the mean and variance of the IoUs. The threshold for selecting positive samples is t = m + g, where m is the mean and g is the variance. Finally, based on the threshold t for each layer, the actual positive

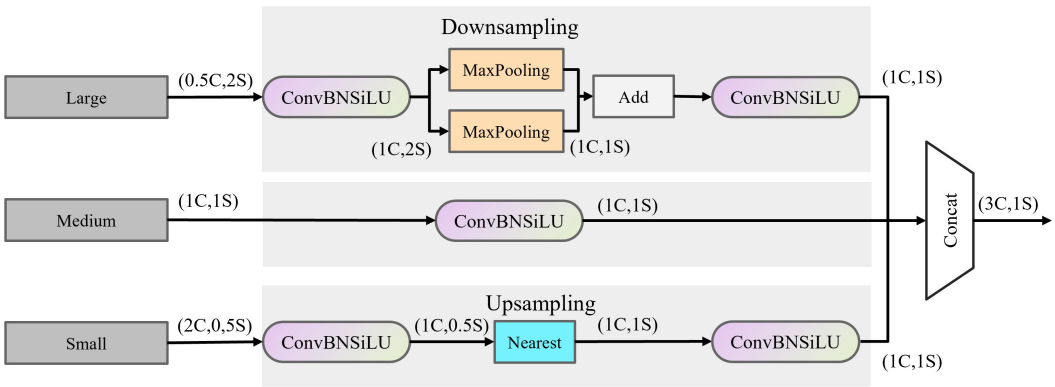


FIGURE 4
TFE module structure.

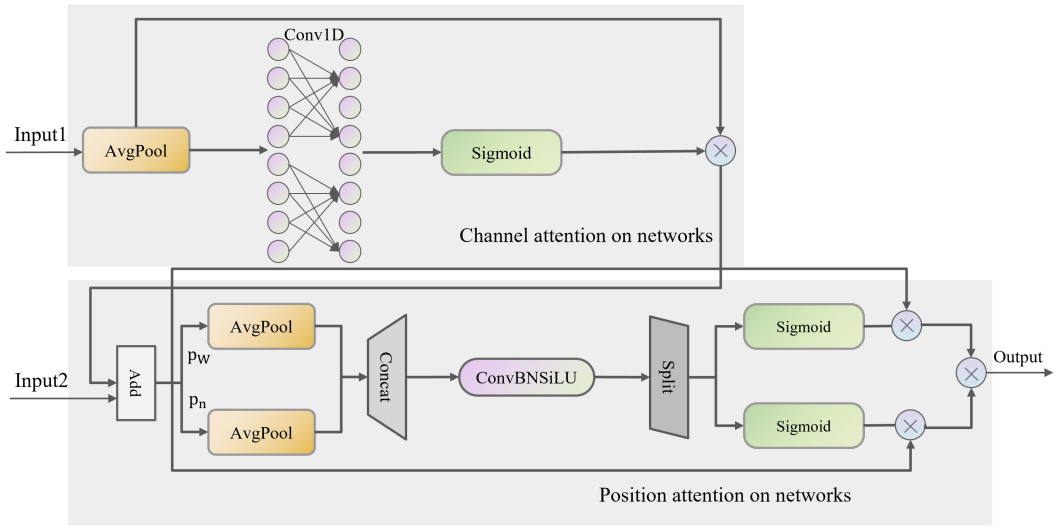


FIGURE 5
CPAM module structure.

samples to be included in the training are selected from the candidate positive samples.

During forward propagation, ATSS optimizes the selection of positive and negative samples, ensuring that the model learns features from more representative samples. These features are more indicative of the actual objects, improving the model's ability to detect and distinguish grape clusters and individual grapes. By focusing on adaptively chosen samples, the model can better localize and classify objects, leading to higher detection accuracy and fewer false positives. This is crucial for identifying small, densely packed objects such as grapefruits within clusters. In the backward propagation process, the improved selection of training samples results in more stable and informative gradients, enhancing the efficiency of parameter updates and speeding up model convergence. By ensuring a more balanced distribution of positive and negative samples across different object scales, ATSS facilitates more effective learning of multi-scale features, which is particularly beneficial for accurately detecting both small individual grapes and larger grape clusters.

3.3 IoU loss

In the YOLOv8 network, CIoU is the bounding box regression loss function. It calculates the loss by considering the overlap area between predicted boxes and ground-truth boxes, the distance between their center points, and the aspect ratio, with the formula detailed below:

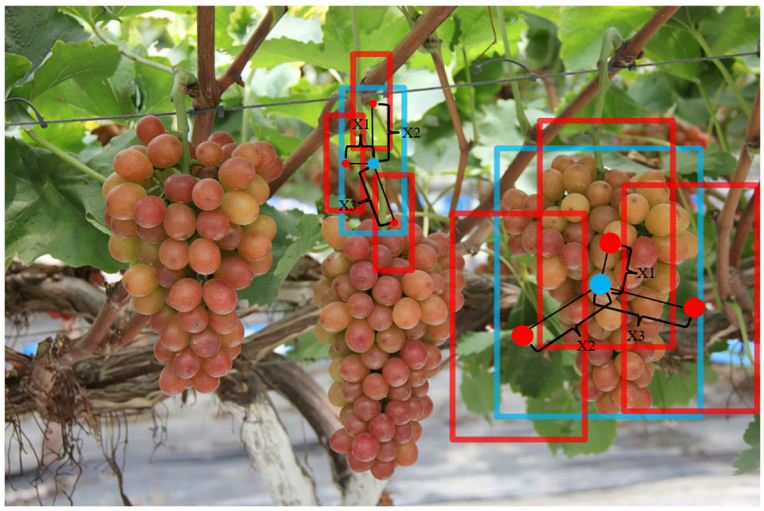


FIGURE 6
Grape stem and fruit pre-selection frame.

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + av$$

CIoU employs a monotonic focusing mechanism aimed at enhancing the fitting capability of the bounding box loss. However, when the object detection training set contains low-quality examples, indiscriminately reinforcing the regression for low-quality examples could hinder the improvement of model detection performance (Zheng et al., 2020). Consequently, this paper adopts a dynamic non-monotonic focusing mechanism loss function, WIoU, which constructs distance attention based on distance metrics, resulting in WIoUv1. The formula for this loss function is as follows:

$$\mathcal{L}_{WIoUv1} \equiv R_{WIoU}\mathcal{L}_{IoU}$$

$$R_{WIoU} = exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(w_g^2 + H_g^2)^*}\right)$$

WIoUv3 defines an outlier degree to describe the quality of anchor boxes, assigning smaller gradient gains to anchors with a more considerable outlier degree. This effectively prevents harmful gradients from low-quality examples, allowing the loss function to focus more on samples of average quality and improve the overall model performance.

$$\beta = \frac{\mathcal{L}_{IoU}^*}{\mathcal{L}_{IoU}} \in [0, +\infty]$$

$$\mathcal{L}_{WIoUv3} = \gamma\mathcal{L}_{WIoUv1}$$

$$\gamma = \frac{\beta}{\delta a^{\beta}}$$

During training, WiseIoU enhances the model's ability to learn from complex and varied object shapes by providing more informative feedback on the quality of predictions. This leads to better localization and classification performance, as the model is trained to optimize for a metric that closely aligns with the practical requirements of accurate object detection in vineyards. The improved feedback mechanism helps the model to fine-tune its predictions more effectively, reducing the number of false positives and false negatives.

In the evaluation phase, WiseIoU offers a more reliable and discriminative measure of the model's detection performance. Traditional IoU can sometimes overestimate the accuracy of detection when dealing with minor or partially overlapping objects. WiseIoU mitigates this issue by providing a more detailed and context-aware assessment, ensuring that the model's performance metrics more accurately reflect its real-world detection capabilities. Moreover, WiseIoU contributes to the robustness of the detection model by improving its generalization across different environmental conditions and image qualities. By accounting for various object shapes and contextual nuances, the model trained with WiseIoU is better equipped to handle variations in lighting, occlusion, and background clutter commonly encountered in vineyard images. This leads to more consistent and reliable detection results, enhancing the overall effectiveness of the model in practical agricultural settings.

3.4 Channel pruning

This study employs the Random channel pruning(RCP) method, offering a novel approach to estimating the contribution of neurons (filters) to the final loss. During training, the contribution to the final loss is analyzed based on average gradients and weight values, and iteratively, neurons with lower scores are removed. The model takes a trained network as input during the pruning process and prunes it at a reduced learning rate during iterative fine-tuning. Following fine-tuning, the model's accuracy is restored.

The RCP algorithm contributes to the robustness and adaptability of the detection model, encouraging the model to learn redundant and complementary features across different channels. This redundancy ensures that the model can still perform accurately even if specific channels are removed. For the detection of grape clusters and fruits, this means that the model can maintain high detection accuracy despite variations in grape size, shape, occlusion, and lighting conditions, which are commonly encountered in vineyard images. RCP also aids in mitigating overfitting. Reducing the number of channels and, thus, the model's capacity helps prevent the model from memorizing the training data. Instead, it promotes the learning of generalizable features that are applicable to unseen test data. This characteristic is particularly beneficial for detecting grape clusters and fruits, where the model must generalize well across different vineyard locations and conditions. By focusing on essential features and discarding redundant ones, RCP ensures that the model remains effective and accurate in detecting grape clusters and fruits in a wide range of scenarios.

4 Results

4.1 Comparison of neck

In the Neck section, this paper implements six distinct enhancements to the YOLOv8n model: YOLOv8n-CARAFE, YOLOv8n-EfficientRepBipan, YOLOv8n-GDFPN, YOLOv8n-GoldYolo, YOLOv8n-HSPAN, and YOLOv8n-ASF. Table 2 displays the comparative results of these network models in terms of mAP, FalsePos, and ModulSize.

Data analysis from Table 2 reveals that YOLOv8n-ASF achieves an mAP(0.5) and mAP(0.5:0.95) of 1.78% and 5.92% higher than the Baseline, respectively. Simultaneously, FalsePos and ModulSize are reduced by 27.28% and 1.69% compared to the Baseline. Although YOLOv8n-EfficientRepBipan and YOLOv8n-GDFPN also show a significant improvement in mAP compared to the Baseline, its FalsePos and ModulSize are higher, indicating that while YOLOv8n-GoldYolo may increase accuracy in grape detection, it results in more detection errors and has a larger storage volume, making it less suitable for deployment in embedded devices like automated grape harvesting robots. Among the improved network

TABLE 2 Neck end network model improvement comparison.

Model	mAP(0.5)	mAP(0.5:0.95)	FalsePos	ModulSize (M)
YOLOv8n(Baseline)	0.900	0.557	6.89	5.9
YOLOV8n-CARAFE	0.909	0.593	5.83	6.2
YOLOv8n- EfficientRepBipan	0.915	0.569	5.16	5.6
YOLOv8n-GDFPN	0.914	0.576	5.83	6.5
YOLOv8n-GoldYolo	0.910	0.580	5.38	11.8
YOLOv8n-HSPAN	0.910	0.583	5.99	4.2
YOLOV8n-ASF	**0.916**	**0.590**	**5.01**	**6.0**

The bold values in the table are the best methods in the experiment and their corresponding experimental data.

models, YOLOv8n-ASF demonstrates the best comprehensive improvement in grape detection.

4.2 IoU loss function comparison experiment

To investigate the impact of loss functions on the performance of the improved model, we set YOLOv8n as the Baseline and a model with enhancements to the Neck section and label matching strategy, YOLOv8n+ASF+ATSS, as Model A. This paper compares the mAP values and FalsePos of five commonly used IoU loss functions: IoU, EIoU, MDPIoU, and Wise-IoU. The comparative results of these loss functions are shown in the table.

Table 3 shows that the YOLOv8n model using the Wise-IoU loss function achieves higher detection accuracy. Compared to the original YOLOv8n model using the CIoU loss function, mAP(0.5) and mAP(0.5:0.95) increased by 3.11% and 7.36%, respectively, with FalsePos decreasing by 35.84%. Relative to Model A, which improved both the Neck section and label matching strategy, mAP(0.5) and mAP(0.5:0.95) increased by 1.19% and 1.35%, respectively, with FalsePos decreasing by 18.89%. These results indicate that using the Wise-IoU loss function can stabilize the model's bounding box regression and improve prediction accuracy.

TABLE 3 Comparison of model performance after improved loss function.

Model	mAP(0.5)	mAP (0.5:0.95)	FalsePos
YOLOv8n(Baseline)	0.900	0.557	6.89
YOLOv8n+ASF+ATSS(A)	0.917	0.590	5.45
A+IoU	0.919	0.589	5.13
A+EIoU	0.926	0.602	4.96
A+MPDIoU	0.926	0.592	4.30
A+ShapeIoU	0.922	0.588	5.32
A+Wise-IoU	**0.928**	**0.598**	**4.42**

The bold values in the table are the best methods in the experiment and their corresponding experimental data.

4.3 Ablation experiment

Ablation experiments were conducted to validate the effectiveness of the Attentional Scale Sequence Fusion (ASF) module, Adaptive Training Sample Selection (ATSS) label matching strategy, and the enhanced Wise-IoU loss function within the TiGra-YOLO network model for grape detection. The results, showcasing mAP values across various IoU thresholds and comparative metrics for different error types, are presented in Table 4, where "×" denotes the absence of the corresponding enhancement module and its presence otherwise.

In Table 4, the mAP values for TiGra-YOLOv8 across different IoU ranges were 0.928, 0.619, and 0.598, respectively. These values represent an increase of 3.11%, 13.16%, and 7.36% over the original YOLOv8n (Baseline). Moreover, localization errors (Loc), duplicate detection errors (Dupe), and overall error metrics (FalsePos) decreased by 19.68%, 62.66%, and 35.84% compared to the original YOLOv8n model. This indicates that the TiGra-YOLOv8 model significantly enhances the accuracy of grapefruit and stem detection. This improvement can be attributed to the ASF module, which increases detection precision for small targets like grape stems by filtering out irrelevant background noise and retaining valuable information for object detection.

Furthermore, the ATSS matching strategy effectively filters positive and negative samples. It is an adaptive selection method that divides training samples based on the statistical features (variance and mean) of grapes and their stems. Finally, the Wise-IoU, employing a dynamic non-monotonic focusing mechanism for bounding box loss, optimizes model performance based on the overlap between predicted and actual grape stems, assessing accuracy and providing a gradient gain allocation strategy focused on anchors of average quality to enhance detector performance.

4.4 Channel pruning experiment

This study conducted experiments on six commonly used pruning methods, setting YOLOv8n as Baseline1 and YOLOv8n +ASF+ATSS+WiseIoU as Baseline2. The pruning rate (speed_up) was set to 2.0, with global_pruning set to False, meaning that the number of pruned channels per layer was roughly consistent.

TABLE 4 Comparison of ablation experimental indexes.

ASF-YOLO	ATSS	Wise-IoU	mAP(0.5)	mAP(0.75)	mAP (0.5:0.95)	Loc	Dupe	FalsePos
×	×	×	0.900	0.547	0.557	5.69	0.75	6.89
√	×	×	0.916	0.588	0.590	4.52	0.62	5.01
√	√	×	0.917	0.605	0.590	4.89	0.46	5.45
√	√	√	**0.928**	**0.619**	**0.598**	**4.57**	**0.28**	**4.42**

× means that this module was not used in this round of experiment, √ means that this module was used in this round of experiment.
The bold values in the table are the best methods in the experiment and their corresponding experimental data.

Performance metrics such as model size, parameter count, computational cost, FPS, and mAP were compared, as shown in Table 5.

Comparing the six pruning models with the Baselines revealed that Random achieved the best overall performance metrics. Compared to Baseline1, it improved mAP by 3.33% and FPS by 7.49%. Compared to Baseline2, mAP increased by 0.21% and FPS by 32.72%, while model size, parameter count, and computational cost were reduced by 50% compared to both Baseline models.

To explore the impact of the pruning rate on model performance, the study chose the Random pruning method and experimented with pruning rates of 1.5, 1.7, and 2.0. A pruning rate of 1.5 implies that the computational cost of the pruned model is 2/3 of the original model, meaning that 1/3 of the original channel connections were removed from the network. Similarly, pruning rates of 1.7 and 2.0 indicate that the computational cost of the pruned model is 1/1.7 and 1/2 of the original, respectively. After pruning, the network underwent fine-tuning to compensate for lost connections, restoring accuracy and improving overall network performance. The comparative results are presented in Table 6.

Analysis of Table 6 shows that the Random pruning method achieved the highest mAP and FPS at a pruning rate of 2.0, with the smallest model size, parameter count, and computational cost. The value of mAP(0.5) increased compared to the pre-pruning model and both Baselines, indicating that this pruning method optimally enhances the comprehensive performance of the model at a pruning rate of 2.0.

4.5 Network model comparison experiment

This experiment contrasts the detection heatmaps of YOLOv8 and TiGra-YOLOv8, as shown in Figure 7.

From the heat maps of A and B in Figure 7, it can be seen that compared to the original YOLOv8 model's heatmap, the improved model's heatmap exhibits more distinct edges and contours, with a more concentrated focus on the centers of the target objects.This indicates that the TiGra-YOLOv8 model proposed in this paper can better focus on the features of detection targets, thereby enhancing detection accuracy.

We also compared the improved model with existing models, such as YOLOv5n, YOLOv7-tiny, YOLOv8n, and YOLOv9-c, with the results presented in Table 7.

According to the table, compared to YOLOv5n, the mAP(0.5) increased by 2.87%, while Parameters and GFLOPS decreased by 19.47% and 6.67%, respectively. Compared to YOLOv7-tiny, the mAP(0.5) increased by 3.91%, with Parameters and GFLOPS decreasing by 87.56% and 67.94%, respectively. Compared to YOLOv8n (Baseline), the mAP(0.5) increased by 3.33%, with Parameters and GFLOPS decreasing by 52.19% and 51.72%, respectively. Compared to the currently available YOLOv9-c

TABLE 5 Comparison of indicators after pruning.

Prune Method (2.0x)	ModulSize (M)	Parameter	GFLOPS	FPS	mAP(0.5)
YOLOv8n (Baseline1)	5.9	3.20M	8.7	1066.3	0.900
(Baseline2)	6.0	3.05M	8.5	863.6	0.928
L1	3.2	1.53M	4.2	1144.1	0.928
Lamp				1144.4	0.915
Group_taylor				1144.9	0.922
Group_norm				1142.3	0.920
Group_hessian				1140.8	0.925
Random				**1146.2**	**0.930**

The bold values in the table are the best methods in the experiment and their corresponding experimental data.

TABLE 6 Comparison of different pruning rate indicators.

speed_up	ModulSize (M)	Parameters	GFLOPS	FPS	mAP(0.5)
YOLOv8n(Baseline1)	5.9	3.20M	8.7	1066.3	0.900
Baseline2	6.0	3.05M	8.5	863.6	0.928
1.5x	4.1	2.02M	5.6	1028.4	0.929
1.7x	3.6	1.79M	5.0	1063.5	0.931
2.0x	**3.2**	**1.53M**	**4.2**	**1146.2**	**0.930**

The bold values in the table are the best methods in the experiment and their corresponding experimental data.

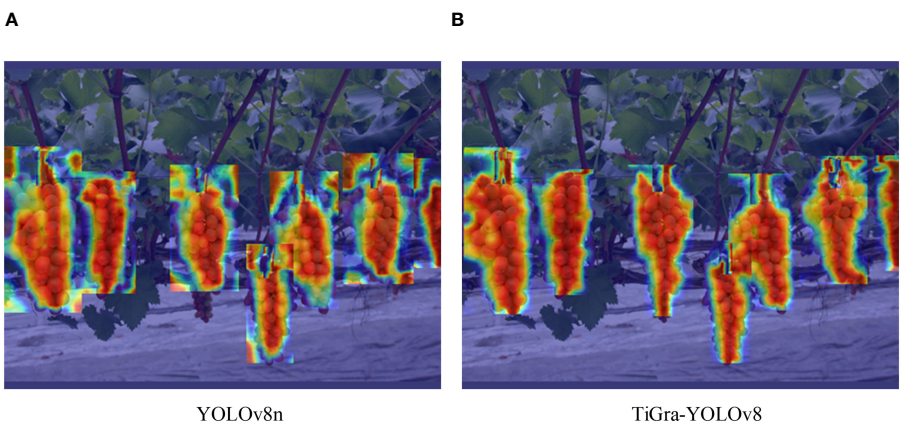


FIGURE 7
Comparison of detection heat maps. **(A)** YOLOv8; **(B)** TiGra-YOLOv8.

TABLE 7 Comparison of indicators of different models.

Model	mAP(0.5)	Parameters	GFLOPS
YOLOv5n	0.904	1.90M	4.5
YOLOv7-tiny	0.895	12.3M	13.1
YOLOv8n	0.900	3.20M	8.7
YOLOv9-c	0.892	25.30M	102.1
TiGra-YOLOv8(ours)	**0.930**	**1.53M**	**4.2**

The bold values in the table are the best methods in the experiment and their corresponding experimental data.

model, the mAP(0.5) increased by 4.26%, with Parameters and GFLOPS decreasing by 93.95% and 95.88%, respectively.

The detection accuracy, parameter count, and computational load for grape recognition in natural environments have improved compared to both previous and the most recently proposed YOLO series models.

5 Discussion

This study proposes a lightweight detection method based on YOLOv8. It attains a detection accuracy of 93% while reducing the model size by half. Specifically, the integration of ASF-YOLO into the neck of the TiGra-YOLOv8 model plays a crucial role in enhancing feature aggregation and improving the model's capability to capture intricate details within images, thereby contributing significantly to the overall improvement in precision. The adoption of the ATSS matching strategy provides a dynamic approach to selecting positive and negative samples during the training process. It ensures that the model learns from a representative set of training samples, leading to more robust learning and better generalization. The utilization of the Wise-IoU loss function considers both the overlap and the distance between predicted and ground truth boxes. This dual consideration ensures that the model predicts the location of objects with greater accuracy. The implementation of the Random pruning algorithm is instrumental in reducing the model's size by eliminating redundant and less significant parameters while preserving critical features. In fact, the reduced complexity can lead to more focused learning, as evidenced by the 4.1.4 observed increase in precision post-pruning.

The significant improvements in detection accuracy and reduction in model size achieved by the TiGra-YOLOv8 model have important implications for its potential application in agricultural settings. The enhanced precision of the model, which allows for more accurate identification of grapefruits and stems, can improve the efficiency of grape harvesting processes and reduce waste. The observed reduction in model size, achieved through the implementation of the Random pruning algorithm, indicates a

promising direction towards making the model more computationally efficient. This reduction in complexity suggests the potential for the model to be adapted for deployment on devices with limited computational resources. However, additional work is required to evaluate the performance and feasibility of the TiGra-YOLOv8 model on such low-resource hardware.

The results of this study demonstrate the advantages of our approach when compared to those reported in the existing literature. Traditional machine learning algorithms, commonly used in earlier studies, need help to adapt to varying environmental conditions and typically exhibit lower accuracy. Additionally, instance segmentation models, while precise, are often large and computationally intensive, posing challenges for deployment on resource-constrained devices. The prediction accuracy of the TiGra-YOLOv8 model on this dataset is better than that of classical object detection models such as YOLOv5 and YOLOv8, which indicates that the model has particular potential in dealing with such tasks.

6 Conclusion and prospect

In this study, aimed at recognizing grapefruits and their stems in natural environments, we constructed a dataset from self-captured images of grapefruits and stems, considering the complex backgrounds, density, and occlusions typical of orchard settings. We designed a lightweight object detection model, TiGra-YOLOv8, incorporating the ASF module into the YOLOv8 network structure. The model also features modifications to the IoU loss function and the positive and negative sample matching strategy, enhancing detection accuracy. Furthermore, model size was reduced through channel pruning. This lightweight approach is significant for deploying the model on mobile devices. In summary, the TiGra-YOLOv8 model achieved a detection accuracy of 93%, with a model size of 3.2M, a parameter count of 1.53M, a computational cost of 4.2 GFLOPS, and an FPS of 1146.2. Compared to YOLOv5 and other models, TiGra-YOLOv8 boasts higher detection accuracy and lower model parameters.

Despite the positive findings of this study, there are certain limitations. Firstly, it should be noted that the dataset used in this study was derived solely from a single crop species, thus necessitating future testing and validation on more diverse datasets. Secondly, when deploying and applying the model in practice, practical factors such as device compatibility and real-time performance need to be taken into consideration. Additionally, although model pruning techniques have successfully reduced model complexity, further optimization is still necessary to accommodate a broader range of application scenarios. In future research, we will explore mobile deployment of the model and deploy the TiGra-YOLOv8 model on small computing devices. Additionally, we plan to collect more diverse grape datasets and train a more generalized TiGra-YOLOv8 model.

In summary, the lightweight detection method based on YOLOv8 proposed in this study demonstrated exceptional performance in detecting grape fruit and grape stalk, thereby offering a novel technical approach for agricultural automation and intelligence. Future studies will further investigate the model's generalization ability and practical application potential.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

XW: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. RT: Conceptualization, Data curation, Formal analysis, Software, Validation, Visualization, Writing – review & editing. JM: Project administration, Writing – review & editing. YN: Investigation, Writing – review & editing. ZX: Writing – review & editing. ZC: Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study has been funded by the Science and Technology Department of Sichuan Province (Contract number: 2022NSFSC0172 and 22ZDYF0095).

Acknowledgments

Thanks to Four Seasons Vineyard for helping with data acquisition and Professor Jiong Mu from Sichuan Agricultural University for her valuable comments.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bac, C. W., Hemming, J., and Van Henten, E. J. (2014). Stem localization of sweet-pepper plants using the support wire as a visual cue. *Comput. Electron. Agric.* 105, 111–120. doi: 10.1016/j.compag.2014.04.011

Bai, R., Lu, J., Zhang, Z., Wang, M., and Wang, Q. (2024). AeroDetectNet: A lightweight, high-precision network for enhanced detection of small objects in aerial remote sensing imagery. *Measurement Sci. Technol.* 35. doi: 10.1088/1361-6501/ad41fe

Behera, S. K., Rath, A. K., and Sethy, P. K. (2021). Fruits yield estimation using Faster R-CNN with MIoU. *Multimedia Tools Appl.* 80, 19043–19056. doi: 10.1007/s11042-021-10704-7

Bochkovskiy, A., Wang, C. Y., and Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*. doi: 10.48550/arXiv.2004.10934

Chaivivatrakul, S., and Dailey, M. N. (2014). Texture-based fruit detection. *Precis. Agric.* 15, 662–683. doi: 10.1007/s11119-014-9361-x

Chen, J., Ma, A., Huang, L., Su, Y., Li, W., Zhang, H., et al. (2023). GA-YOLO: A lightweight YOLO model for dense and occluded grape target detection. *Horticulturae* 9, 443. doi: 10.3390/horticulturae9040443

del Pilar Martínez-Diz, M., Eichmeier, A., Spetik, M., Bujanda., R., Díaz-Fernández., Á, Díaz-Losada., E., et al. (2020). Grapevine pruning time affects natural wound colonization by wood-invading fungi. *Fungal Ecol.* 48, 100994. doi: 10.1016/j.funeco.2020.100994

Dilek, E., and Dener, M. (2023). Computer vision applications in intelligent transportation systems: a survey. *Sensors* 23, 2938. doi: 10.3390/s23062938

Feng, C., Zhong, Y., Gao, Y., Scott, M. R., and Huang, W. (2021). "Tood: Task-aligned one-stage object detection," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada: IEEE Computer Society, 3490–3499. doi: 10.1109/ICCV48922.2021.00349

Gao, F., Fu, L., Zhang, X., Majeed, Y., and Zhang, Q. (2020). Multi-class fruit-on-plant detection for apple in SNAP system using Faster R-CNN. *Comput. Electron. Agric.* 176, 105634. doi: 10.1016/j.compag.2020.105634

Girshick, R. (2015). "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, IEEE Computer Society, pp. 1440–1448. doi: 10.1109/ICCV.2015.169

Kang, M., Ting, C., Ting, F., and Phan, R. (2024). ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation. *Image and Vision Computing* 147. doi: 10.1016/j.imavis.2024.105057

Lin, G., Tang, Y., Zou, X., Cheng, J., and Xiong, J. (2020). Fruit detection in natural environment using partial shape matching and probabilistic Hough transform. *Precis. Agric.* 21, 160–177. doi: 10.1007/s11119-019-09662-w

Lin, J., Wang, H., Zou, X., Zhang, P., Li, C., Zhou, Y., et al. (2023). Obstacle avoidance path planning and simulation of mobile picking robot based on DPPO. *J. System Simulation* 35, 1692–1704.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). "SSD: Single shot multibox detector," in *European conference on computer vision.Berlin, German*. Springer, Cham. 9905, 22–37. doi: 10.1007/978-3-319-46448-0_2

Liu, S., Chen, T., Chen, X., Shen, L., Mocanu, D. C., Wang, Z., et al. (2022). The unreasonable effectiveness of random pruning: Return of the most naive baseline for sparse training. *arXiv preprint arXiv:2202.02643*. doi: 10.48550/arXiv.2202.02643

Liu, P., Zhu, Y., Zhang, T., and Hou, J. (2020). Algorithm for recognition and image segmentation of overlapping grape cluster in natural environment. *Trans. Chin. Soc. Agric. Eng. (Transactions CSAE)* 36, 161–169. doi: 10.11975/j.issn.1002-6819.2020.06.019

Lou, H., Duan, X., Guo, J., Liu, J., Liu, H., Gu, J., et al. (2023). DC-YOLOv8: small-size object detection algorithm based on camera sensor. *Electronics* 12, 2323. doi: 10.3390/electronics12102323

Lu, H., and Luo, M. (2022). Survey on new progresses of deep learning based computer vision. *J. Data Acquis. Process* 37, 247–278. doi: 10.16337/j.1004-9037.2022.02.001

Luo, L., Tang, Y., Lu, Q., Chen, X., Zhang, P., Zou, X., et al. (2018). A vision methodology for harvesting robot to detect cutting points on peduncles of double overlapping grape clusters in a vineyard. *Comput. industry* 99, 130–139. doi: 10.1016/j.compind.2018.03.017

Luo, L., Tang, Y., Zou, X., Ye, M., Feng, W., Li, G., et al. (2016). Vision-based extraction of spatial information in grape clusters for harvesting robots. *Biosyst. Eng.* 151, 90–104. doi: 10.1016/j.biosystemseng.2016.08.026

Luo, L., Zou, X., Xiong, J., Zhang, Y., Peng, H., and Lin, G. (2015). Automatic positioning for picking point of grape picking robot in natural environment. *Trans. Chin. Soc. Agric. Eng. (Transactions CSAE)* 31, 14–21. doi: 10.3969/j.issn.1002-6819.2015.02.003

Ma, B., Hua, Z., Wen, Y., Deng, H., Zhao, Y., Pu, L., et al. (2024). Using an improved lightweight YOLOv8 model for real-time detection of multi-stage apple fruit in complex orchard environments. *Artif. Intell. Agric.* 11, 70–82. doi: 10.1016/j.aiia.2024.02.001

Moro, K. I. B., Bender, A. B. B., da Silva, L. P., and Penna, N. G. (2021). Green extraction methods and microencapsulation technologies of phenolic compounds from grape pomace: A review. *Food Bioprocess Technol.* 14, 1407–1431. doi: 10.1007/s11947-021-02665-4

Pérez-Zavala, R., Torres-Torriti, M., Cheein, F. A., and Troni, G. (2018). A pattern recognition strategy for visual grape bunch detection in vineyards. *Comput. Electron. Agric.* 151, 136–149. doi: 10.1016/j.compag.2018.05.019

Ren, S., He, K., Girshick, R., and Sun, J. (2017). "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence, United States, IEEE Computer Soc*, vol. 39 (6), pp. 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vision* 77, 157–173. doi: 10.1007/s11263-007-0090-8

Sun, J., Gong, D., Yao, K., Lu, B., Dai, C., and Wu, X. (2022). Real-time semantic segmentation method for field grapes based on channel feature pyramid. *Trans. Chin. Soc. Agric. Eng. (Transactions CSAE)* 38, 150–157. doi: 10.11975/j.issn.1002-6819.2022.17.016

Tian, H., Wang, T., Liu, Y., Qiao, X., and Li, Y. (2020). Computer vision technology in agricultural automation—A review. *Inf. Process. Agric.* 7, 1–19. doi: 10.1016/j.inpa.2019.09.006

Tong, Z., Chen, Y., Xu, Z., and Yu, R. (2023). Wise-iou: Bounding box regression loss with dynamic focusing mechanism. *arXiv preprint arXiv:2301.10051*. doi: 10.48550/arXiv.2301.10051

Wu, Z., Xia, F., Zhou, S., and Xu, D. (2023). A method for identifying grape stems using keypoints. *Comput. Electron. Agric.* 209, 107825. doi: 10.1016/j.compag.2023.107825

Yu, Y., Zhang, K., Yang, L., and Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Comput. Electron. Agric.* 163, 104846. doi: 10.1016/j.compag.2019.06.001

Zhang, S., Chi, C., Yao, Y., Lei, Z., and Li, S. Z.>. (2020). "Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, IEEE Computer Society pp. 9756–9765. doi: 10.1109/CVPR42600.2020.00978

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., and Ren, D. (2020). "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. New York, USA: AAAI, 12993–13000. doi: 10.1609/aaai.v34i07.6999

Zhou, X., Wu, F., Zou, X., Meng, H., Zhang, Y., Luo, X., et al. (2023). Method for locating picking points of grape clusters using multi-object recognition. *Trans. Chin. Soc. Agric. Eng. (Transactions CSAE)* 39, 166–177. doi: 10.11975/j.issn.1002-6819.202309105

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 01 August 2024
DOI 10.3389/fpls.2024.1442225

Check for updates

OPEN ACCESS

EDITED BY
Wenyu Zhang,
Jiangsu Academy of Agricultural Sciences Wuxi Branch, China

REVIEWED BY
Wenxiu Wang,
Hebei Agricultural University, China
Hongzhe Jiang,
Nanjing Forestry University, China

*CORRESPONDENCE
Hongwei Sun
sunhongwei@hdu.edu.cn
Kaihua Wu
wukaihua@hdu.edu.cn

RECEIVED 01 June 2024
ACCEPTED 15 July 2024
PUBLISHED 01 August 2024

CITATION
Luo B, Sun H, Zhang L, Chen F and Wu K (2024) Advances in the tea plants phenotyping using hyperspectral imaging technology.
Front. Plant Sci. 15:1442225.
doi: 10.3389/fpls.2024.1442225

COPYRIGHT
© 2024 Luo, Sun, Zhang, Chen and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advances in the tea plants phenotyping using hyperspectral imaging technology

Baidong Luo[1], Hongwei Sun[1*], Leilei Zhang[2], Fengnong Chen[1] and Kaihua Wu[1*]

[1]College of Artificial Intelligence, Hangzhou Dianzi University, Hangzhou, China, [2]Key Laboratory of Specialty Agri-Products Quality and Hazard Controlling Technology of Zhejiang Province, College of Life Sciences, China Jiliang University, Hangzhou, China

Rapid detection of plant phenotypic traits is crucial for plant breeding and cultivation. Traditional measurement methods are carried out by rich-experienced agronomists, which are time-consuming and labor-intensive. However, with the increasing demand for rapid and high-throughput testing in tea plants traits, digital breeding and smart cultivation of tea plants rely heavily on precise plant phenotypic trait measurement techniques, among which hyperspectral imaging (HSI) technology stands out for its ability to provide real-time and rich-information. In this paper, we provide a comprehensive overview of the principles of hyperspectral imaging technology, the processing methods of cubic data, and relevant algorithms in tea plant phenomics, reviewing the progress of applying hyperspectral imaging technology to obtain information on tea plant phenotypes, growth conditions, and quality indicators under environmental stress. Lastly, we discuss the challenges faced by HSI technology in the detection of tea plant phenotypic traits from different perspectives, propose possible solutions, and envision the potential development prospects of HSI technology in the digital breeding and smart cultivation of tea plants. This review aims to provide theoretical and technical support for the application of HSI technology in detecting tea plant phenotypic information, further promoting the trend of developing high quality and high yield tea leaves.

KEYWORDS

hyperspectral imaging, tea plants, plant phenomics, high-throughput, rapid detection, environmental stress

1 Introduction

As a popular beverage crop, tea plants are extensively cultivated in regions including Asia and Africa. For instance, over seventy million labors in China are engaged in tea-related industries in 2020, and the income of them is closely related to the yield and quality of tea product. In modern agricultural science, digital breeding and precision cultivation

technologies emerged to obtain specific phenotypic traits of tea plants, and thus to promote tea production (Lu et al., 2020; Wang et al., 2023). The digital breeding mainly focuses on interpreting the genetic foundation of the tea plant phenotypic traits, while the precision cultivation technology is to unveil the regulatory network between the tea plant phenotyping and environment factors. Therefore, it can also be seen that the phenotype of tea plants is affected by the synergistic regulation of genetic material and environment. Unlike cereal crops and other horticultural plants, tea plants are typical foliar plants with unique phenotypic indices of interest. For instance, leaf area in tea plants correlates directly with yield to some extent, and the morphology of tea buds serves as a criterion in the grade evaluation of Chinese green tea. The content of polyphenols and theanine in leaves significantly influences the suitable processing methods for tea leaves and plays a critical role in the flavor and taste profile of the tea. In sum, the phenotype information of tea plants encompasses aspects such as environmental stress responses, growth status, and quality traits, characterized by rich connotations and target-oriented traits.

Compared to genetic breeding and cultivation techniques, traditional phenotypic detection in plants researches is usually performed manually, which is less efficient, highly dependent on experienced professionals, time-consuming, and labor-intensive. Especially in the determination of biochemical components in plants, wet chemistry method leads to "Phenotyping bottleneck" (Watt et al., 2020; Tao et al., 2022). In order to meet the need of high-throughput rapid detection of plant phenotyping, hyperspectral imaging technology has been paid more attention by researchers. Hyperspectral imaging (HSI) technique can be used to invert plant traits based on the changes of plant optical properties, and can collect multi-modal data of spectra and space information, which is especially suitable for simultaneous monitoring of some plant traits. Combined with machine learning and artificial intelligence technologies, HSI technology enables automated and precise detection of plant phenotypes, which has the potential to enhance the efficiency of detection, reduces artificial error, and provide more objective and accurate presentation of plant traits (Atefi et al., 2021).

In recent years, as a powerful manner for detecting plant information, hyperspectral imaging technology has been extensively applied in tea plant phenotyping researches, involving environmental resistant, physiological, and biochemical traits. However, many challenges are faced by HSI technology to further promote the application in digital breeding and precision cultivation research. For example, it is necessary to deeply understand the relationship between tea plant phenotypic traits and hyperspectral information, and comprehensively analyze its feasibility and limitations for phenotyping. Hence, we elucidate the principles of hyperspectral imaging technology and summarize typical workflows and related algorithmic techniques for hyperspectral data analysis. Based on these previous studies, we categorize the application of hyperspectral imaging technology in tea plant phenotyping into environmental stress responses, growth status, and quality traits. Simultaneously, we discuss the current challenges faced when using HSI technology in the detection of tea plant phenotypic information and potential future trends. Our aim is to comprehensively summarize the current state of tea plant phenotypic information detection researches using hyperspectral imaging technology and provide in-depth commentary and suggestions on the current predicaments in the analysis of tea plant phenotyping.

2 Principles of hyperspectral imaging technology for tea plant phenotypic detection

Hyperspectral imaging technology combines spectroscopy and computer vision, which is able to measure the optical absorption, scattering, and reflectance properties of plant tissues and organs across a wide variety of wavelengths, to perform quantitative analyses and qualitative assessments for key agronomic traits such as the growth status, yield, and quality of tea plants. In the visible to near-infrared spectral wavelength (400-2500nm), this technology captures the molecular vibration information of plant biochemical components with hydrogen-containing groups, like chlorophyll and other photosynthetic pigments (Slaton et al., 2001; Tang et al., 2024). In addition, hyperspectral imaging technology also has the ability to monitor the spatial distribution of plant phenotypic traits at different scales, from cells, organs, individuals to populations, as shown in Figure 1 (Mahlein, 2016; Kuswidiyanto et al., 2022). Therefore, hyperspectral data is always stored in the format of a multi-dimensional cube, as its spectral dimension records chemical component content in plant leaf, while the spatial dimension details the distribution of various compositional contents of canopy, or the morphological structures of tea plant and epidemic range of pests and diseases. For instance, by utilizing visual representations of moisture distribution in tea leaves, one can observe the water content at each pixel level in a more intuitive and comprehensive manner, thereby offering an innovative approach for assessing irrigation needs of tea plants (Sun et al., 2019).

In the plant phenomics research field, the principles of plant phenotyping using hyperspectral imaging technology can be divided into direct and indirect manners (Zhang N. et al., 2020). In the study of detecting the content of water and polyphenols in leaves by hyperspectral imaging technology, since water and polyphenols have strong absorption peaks in the near-infrared spectral band, a mathematical model between the absorbance of tea leaves and the content of water and polyphenols can be directly constructed, which belongs to direct detection methods. In plant pest detection studies, plant leaves often appear different forms of disease spots, necrotic or wilted areas after being infected by germs, and the content and activity of pigments would be reduced, resulting in an increase in the reflectivity of the visible region, while the red edge (670-730nm) moves to the shortwave direction (Bai et al., 2020). In these process of plant infestation by pests and diseases, the spectral response is related to the symptoms on plant pigment, water, leaf area and other biochemical substances caused by diseases and pests, which belongs to indirect detection methods. Compared to direct detection, the precision of indirect detection is commonly susceptible to variations influenced by plant species, age, and even the cultivation environment, resulted in necessary to recalibrate the detection model in a period of time. However, this

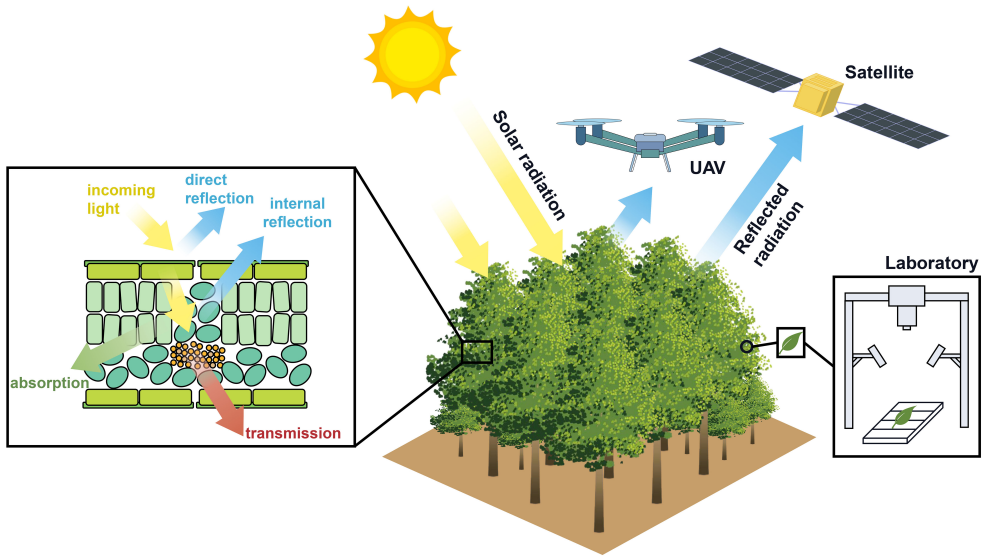


FIGURE 1
Multi-scale hyperspectral data measurement scheme.

method can reflect the relationship between different traits of plants, as well as the influencing factors and evolution rules.

3 The workflow of HSI application in tea plant phenotypic analysis

No uniform standard workflow for hyperspectral data processing in tea plants has been established. However, drawing upon a summary of previous researches in detecting tea plant phenotypic information using hyperspectral imaging technology, the entire data processing workflow can generally be divided into four steps: data acquisition, data preprocessing, data analysis, and data application (Figure 2) (Li et al., 2013; Paulus and Mahlein, 2020).

3.1 Hyperspectral data acquisition

Hyperspectral imaging systems are usually deployed on diverse platforms such as satellites, airplanes, unmanned aerial vehicles

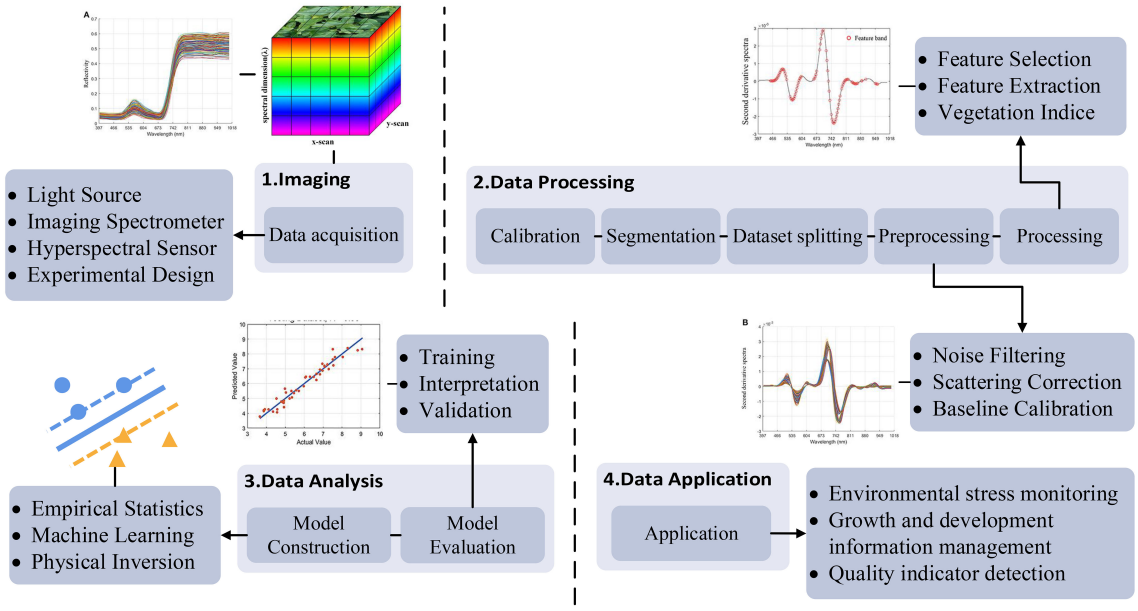


FIGURE 2
Workflow of hyperspectral data processing.

(UAVs) and ground-based fixed equipment, which is typically composed of components like light sources, optical lens, imaging spectrometers, photoelectric detector, and computer data processing systems (Lu et al., 2020). In the process of capturing optical characteristic information of tea plants, samples are illuminated by either natural or artificial light sources (such as halogen lamps), and the optical lens gathers the reflected light from the samples. This reflected light is split into different wavelengths by a diffraction grating element within the spectrometer, and photoelectric detector subsequently transform these light intensity into electrical signals. Finally, computer data processing systems acquire and process these electrical signals, transform them into digital signals, and form a spectrum of target plant (Mishra et al., 2017).

There are mainly four hyperspectral data collection methods: point scanning, line scanning, area scanning, and snapshot (Figure 3). Point scanning (whisk-broom) acquires spectral information pixel by pixel, providing high hyperspectral resolution, its slow speed requires a highly stable platform to avoid image blurring and distortion. Line scanning (push-broom), currently the most commonly used scanning method, captures one row of pixels at a time, balancing spatial and spectral resolution, but care should be taken to minimize motion artifacts during the acquisition process. Area scanning (staring) uses bandpass filters to capture image data for individual bands, scanning band by band to construct a hyperspectral data cube, providing high spatial resolution images, but the scene must be kept stationary during acquisition to ensure the accuracy of the localization and spectral information. Snapshot technology can simultaneously capture spatial and spectral information of the entire scene. A single snapshot represents a perspective projection of the data cube, from which its three-dimensional structure can be reconstructed. However, as this technique is at an early stage of development, the captured images often exhibit comparatively low spatial and spectral resolution (Hagen and Kudenov, 2013). Based on their respective detection principles, each mode has its own set of advantages and weaknesses for practical application. With an in-depth understanding of the basic principles of each imaging modes, we are able to flexibly choose the appropriate acquisition manner in various acquisition environments to maximize the imaging requirements and improve the data acquisition efficiency. In this case, different data collection methods also make it difficult to standardize data, so it is necessary to build a hyperspectral phenotyping data sharing platform, which not only facilitates the global exchange of raw data and processing methods, so as to maintain the consistency of data (Xu et al., 2020; Sun et al., 2022). Despite continuous advancements in hyperspectral imaging technology, the high resolution it offers comes with significant costs that may detract from its cost-effectiveness compared to other spectral sensing techniques. Nevertheless, as optical technologies continue to evolve, these economic barriers are

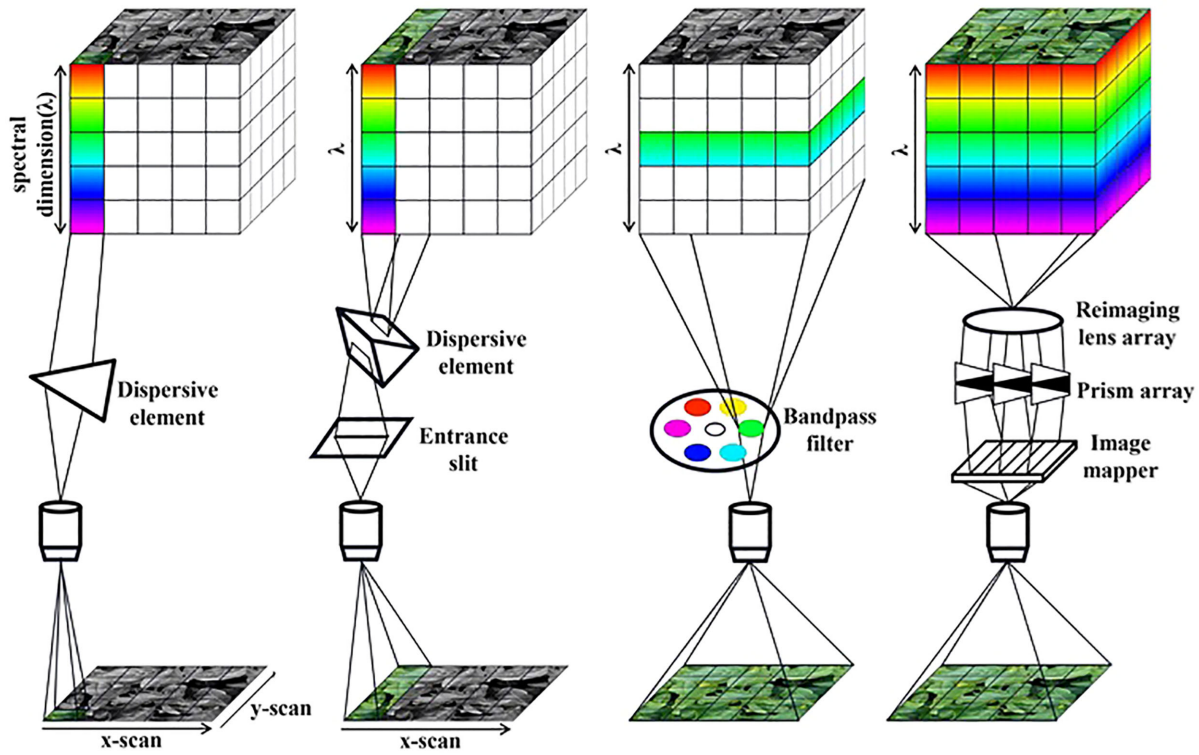


FIGURE 3
Four modes of hyperspectral data acquisition.

expected to be progressively mitigated, thereby fostering the mainstream adoption of hyperspectral imaging in precision agriculture, environmental monitoring, and beyond.

3.2 Hyperspectral data preprocessing

Before the spectral domain analysis of hyperspectral data, it is necessary to segment the image data in the spatial domain to extract the spectral and spatial characteristics of the target. Image segmentation aims to effectively differentiate various substances or objects within an image, dividing them into mutually independent regions, with each region consisting of a collection of pixels that share certain similar characteristics and attributes. For the detection of plant phenotypic information, such operation is to generate masks to accurately demarcate regions of interest (ROI), thus eliminating irrelevant or even possible interference information (ElMasry et al., 2009). Common strategies for image segmentation include threshold segmentation, morphological processing, edge segmentation and spectral image segmentation (Tripathi et al., 2012; Wu and Sun, 2013). Typically, before implementing hyperspectral image segmentation, a single band sub-image with the greatest contrast against the background is selected and segmented to serve as a mask for excluding background pixels (Wang et al., 2020d). In the case of threshold segmentation, an appropriate threshold is set to convert a grayscale image into a binary image mask, to extract hyperspectral data of target object based on its position in the image.

When measuring plant spectral data using a hyperspectral imaging system, interference noise caused by variations from the collection instrument, light source, and environmental conditions is often encountered (Mishra et al., 2020; Çetin and Yıldız, 2022). Depending on the characteristics, noise typically manifests as continuous disturbances or transient spikes in the data. The Savitzky-Golay (S-G) filter algorithm (Steinier et al., 1972), widely used for smoothing, effectively removes the noise while preserving the overall trend of the data by fitting a polynomial to the input signal, while the limited effects in removing low-frequency noise from spectral data. During practical measurement, the appearance of multi-source noise in spectral data may be accompanied by baseline drift (Zhang and Wang, 2023). Derivative algorithms can realize the correction of baseline effects on the basis of removing background interference and reduce the impact of baseline drift. Nevertheless, directly applying a high-order derivative process to spectral data with a low original signal-to-noise ratio might amplify the noise present in the spectrum (Yan et al., 2000; Chu et al., 2004). Therefore, for spectra with low signal-to-noise ratios, the derivative algorithm needs to be used with caution, and noise smoothing or other preprocessing steps may need to be applied before proceeding to minimize potential noise effects. Moreover, the complex tissue structure and surface microfeatures of tea plants can lead to non-uniform scattering of light at different locations, resulting in various scattering effects. Standard Normal Variate Transformation (SNV) (Barnes et al., 1989) and Multiplicative Scatter Correction (MSC) (Wu et al., 2019) are commonly used to correct scattering effects. The MSC algorithm adjusts scattering effects in spectra by comparing the ratio of sample spectra to reference spectra (Geladi et al., 1985), while the SNV algorithm performs mean subtraction and standard deviation division at each wavelength to transform the spectral data into a standard normal distribution. Both algorithms can eliminate scattering effects and spectral differences caused by baseline drift, reducing noise interference encountered during

TABLE 1 Common preprocessing algorithms used in hyperspectral imaging technology for tea plant phenotypic information detection.

Common preprocessing algorithms		Functions	References
Algorithms	Abbreviations		
Normalization	-	Elimination of redundant information, resulting in better interpretability and comparability of data	(Chu et al., 2004)
Standardization	-		
Mean Centering	-		
Orthogonal signal correction	OSC	Filtering out irrelevant information and retaining the main information in the original spectrum	(Wold et al., 1998)
Savitzky-Golay smoothing	S-G	Preserving the essential information and the morphology of the signal while reducing random high-frequency noise in the spectral data	(Steinier et al., 1972)
Moving average smoothing	-		
Standard normal variate transformation	SNV	Elimination of non-uniform scattering due to heterogeneity of tea plant surfaces	(Barnes et al., 1989)
Multiplicative scatter correction	MSC		(Wu et al., 2019)
Derivative	1st Der,2nd Der	Eliminating baseline effects and removing background interference for improved resolution and sensitivity	(Chu et al., 2004)
De-trending	DT	Eliminating baseline drift, baseline tilt and curvature	(Wold et al., 1998)

derivative algorithm processing. To showcase the characteristics of commonly used preprocessing algorithms in hyperspectral imaging techniques for tea plant phenotypic analysis, Table 1 provides a detailed description of the operating mechanism of each preprocessing technique. To further enhance the effectiveness of spectral data preprocessing, a combination of various preprocessing techniques or structural improvements and parameter optimizations of classical algorithms is commonly employed. Additionally, machine learning models are increasingly applied to spectral preprocessing. For instance, autoencoders (AE) utilize self-supervised learning methods to map noisy inputs to latent variable space, reconstructing outputs that closely approximate the original noise-free data, thus achieving automatic denoising (Zhang C. et al., 2020). The application of these methods not only improves the accuracy and efficiency of spectral data modeling but also lays a solid foundation for conducting more complex analyses and applications.

3.3 Hyperspectral data processing

The huge amount of hyperspectral data often leads to the complexity of the calculation, which makes it difficult to find really useful information. It is imperative to select the most informative features from a multitude of bands and remove redundant or unnecessary data (Cheng et al., 2022; Omia et al., 2023). Therefore, prior to establishing relational model between plant phenotypic trait and hyperspectral data, it is often necessary to perform dimensionality reduction operations such as feature selection or feature extraction to address curse of dimensionality of hyperspectral data. Feature selection is a method of filtering out the most representative subset of features from the original feature set. The aim is to preserve information most relevant to the research objective and eliminate other unrelated features to reduce the impact of redundancy and noise (Adao et al., 2017). In the process of analyzing tea plant phenotypic information, Competitive Adaptive Reweighted Sampling (CARS) and Uninformative Variable Elimination (UVE) are two commonly used feature selection methods. The CARS algorithm draws on the "survival of the fittest" principle and calculates the competitiveness score of features based on their inter-correlation within a subset. Li et al. used the CARS algorithm to select 19 representative bands from a dataset of 700 bands, and the corresponding CARS-PLS model reduces the RMSECV by 0.0433 in terms of detection effectiveness compared to the traditional PLS model (Li et al., 2009). The UVE algorithm, on the other hand, selects features with high information gain or chi-square statistic between the selected target variable and the features as the initial selection set, and progressively removes the uninformative variables that do not have a significant impact on the classification performance (Centner et al., 1996). Feature extraction is the process of mapping original features onto a new feature space through a certain transformation method and then selecting key features that can describe essential characteristics of the original data. The Successive Projections Algorithm (SPA) (Milanez et al., 2017) is a widely used feature extraction method that sequentially selects projection directions that maximally retain key characteristics of the original data, mapping high-dimensional data into a lower-dimensional subspace to achieve data dimensionality reduction. In the study of post-harvest fresh tea leaf quality indicators, the SPA-MLR model constructed via SPA significantly outperformed other models in optimizing model performance, demonstrating that the algorithm improves prediction accuracy while simplifying the modeling process (Wang et al., 2020b). Above all, these dimensionality reduction algorithms are data-driven, and it is difficult to explain the physical meaning behind them. And in fact, hyperspectral imaging equipment is facing the dilemma of high cost. In response to this challenge, researchers have ventured into the deployment of multispectral imaging technology, which is more economically viable (Tran, 2001; Xu et al., 2023). This technology typically involves the acquisition of a finite number (usually not exceeding ten) of discrete broadband (50-200nm) images for target objects. By amalgamating hyperspectral data with theoretical analysis to pinpoint wavelength variables associated with specific plant phenotype indices, multispectral technology emerges as an efficacious alternative, especially in designing portable instruments targeted at specific detection objectives (Wang et al., 2023). Future developments in portable multispectral instruments not only have the potential to lower the application costs of the technology but also to broaden the applicability of hyperspectral technology in agricultural production practices, thereby enhancing its viability and economic benefits in real-world production environments (Yang et al., 2021; Saric et al., 2022).

Moreover, in the study of satellite remote sensing and UAV observation, the construction of vegetation indices is also a common method for processing hyperspectral data. Typically formulated through combinations of the differences or ratios of two or more spectral band variables, vegetation indices serve to represent one or multiple parameters indicative of vegetation status (Xue and Su, 2017; He et al., 2023). Typical multispectral vegetation indices are primarily composed of red and near-infrared spectral bands, as the red band reflects chlorophyll content of vegetation, while the near-infrared band is closely related to the structural characteristics of plant leaves. Such indices are primarily employed in performing vegetation monitoring tasks and are capable of mitigating the influence of background interference to some extent. The advent of hyperspectral technology broadens the observable spectral range and offers more refined spectral bands. This enables the hyperspectral vegetation indices to be used for more detailed and precise monitoring of general plant phenotypes on the basis of the original, and to be able to estimate and invert more relevant phenotypic features (Verrelst et al., 2015; He et al., 2023). For instance, the utilization of hyperspectral vegetation indices allows differentiation between vegetation individuals and populations under various stresses, such as water stress and nutrient deficiency. The high-precision spectral data thus provided carry

TABLE 2 Commonly used hyperspectral vegetation indices in the study of tea plant phenotypic information detection.

Vegetation Indices		Characteristics and Functions	Definition	Reference
Structure (LAI, Green biomass, etc.)				
Greenness Index	GI	Estimate biochemical constituents and LAI at leaf and canopy levels	R_{554}/R_{677}	(Zarco-Tejada et al., 2005)
Improved Soil Adjusted Vegetation Index	MSAVI	A more sensitive indicator of vegetation amount than SAVI at canopy level	$0.5[2R_{800}+1-((2R_{800}+1)^2-8(R_{800}-R_{670}))^{1/2}]$	(Qi et al., 1994)
Narrow Band Normalized Difference Vegetation Index	NBNDVI	Responds to change in the amount of green biomass and more efficiently in vegetation with low to moderate density	$(R_{874}-R_{1225})/(R_{874}+R_{1225})$	(Rouse et al., 1974)
Hyperspectral Perpendicular VI	PVI_{hyp}	More efficiently quantify the low amount of vegetation by minimizing soil background influence on vegetation spectrum	$(R_{1148}-aR_{807}-b)/(1+a^2)^{1/2}$, where a = 1.17, b = 3.37	(Schlerf et al., 2005)
Normalization or Standard of the LAIDI	sLAIDI	Sensitive to LAI variation at canopy level with a saturation point >8	$S(R_{1050}-R_{1250})/(R_{1050}+R_{1250})$, where S = 5	(Delalieux et al., 2008)
Pigments (Chls, Cars)				
Leaf Chlorophyll Index	LCI	Estimate Chl content in higher plants, sensitive to variation in reflectance caused by Chl absorption	$(R_{850}-R_{710})/(R_{850}+R_{680})$	(Datt, 1999)
Modified Chlorophyll Absorption in Reflectance Index	MCARI	Respond to Chl variation and estimate Chl absorption	$[(R_{701}-R_{671})-0.2(R_{701}+R_{549})]\times(R_{701}/R_{671})$	(Daughtry et al., 2000)
Ratio Analysis of Reflectance Spectra	RARS	Estimate carotenoid pigment contents in foliage	R_{760}/R_{500}	(Chappelle et al., 1992)
Triangle chlorophyll vegetation index	TCI	Quantify chlorophyll content of tallgrass at leaf and canopy levels	$[(R_{800}+1.5\times R_{550})-R_{675}]/(R_{800}-R_{700})$	(Gao, 2006)
Vogelmann red edge index	VOG	Sensitive to the combined effects of foliage chlorophyll concentration, canopy leaf area, and water content	R_{740}/R_{720}	(Vogelmann et al., 1993)
Water				
Leaf Water VI 1	LWVI-1	Estimate leaf water content, an NDWI variant	$(R_{1094}-R_{893})/(R_{1094}+R_{893})$	(Galvao et al., 2005)
ND Water Index	NDWI	Improving the accuracy in retrieving the vegetation water content at both leaf and canopy levels	$(R_{860}-R_{1240})/(R_{860}+R_{1240})$	(Gao, 1996)
Shortwave Infrared Water Stress Index	SIWSI	Estimate leaf or canopy water stress, especially in the semiarid environment	$(R_{860}-R_{1640})/(R_{860}+R_{1640})$	(Fensholt and Sandholt, 2003)
Water Index	WI	Quantify relative water content at leaf level	R_{900}/R_{970}	(Peñuelas et al., 1997)
Stress				
Red-Edge Vegetation Stress Index	RVSI	Assess vegetation community stress at canopy level	$[(R_{712}+R_{752})]/2-R_{732}$	(Merton and Huntington, 1999)

unparalleled scientific value for grasping the health status of vegetation, growth dynamics, and even adaptive studies in response to environmental changes (Fu et al., 2021). Table 2 lists some commonly used hyperspectral vegetation indices in the research for detecting phenotypic information in tea plants.

3.4 Construction of phenotypic information detection models

The most critical step in the workflow of tea plants phenotyping using hyperspectral technology is the construction of math models between plant traits and plant optical characteristics. As illustrated in Figure 4, a model is established using machine learning techniques, with measured spectral characteristics, spatial features, or fused features as independent variables, and actual measured phenotypic parameters of tea plants as dependent variables (Ang and Seng, 2021).

During model construction, we will involve two types of univariate statistical models and multivariate statistical models. The former focuses on the statistical analysis of a single variable to reveal characteristics of the data. In phenotypic studies of tea plants, univariate statistical techniques typically involve the creation of spectral feature variables such as vegetation indices variables and spectral position variables, etc., which reflect phenotypic characteristics of the tea plants through statistical processing of specific bands (Sonobe et al., 2018b; Cao et al., 2022b). According to

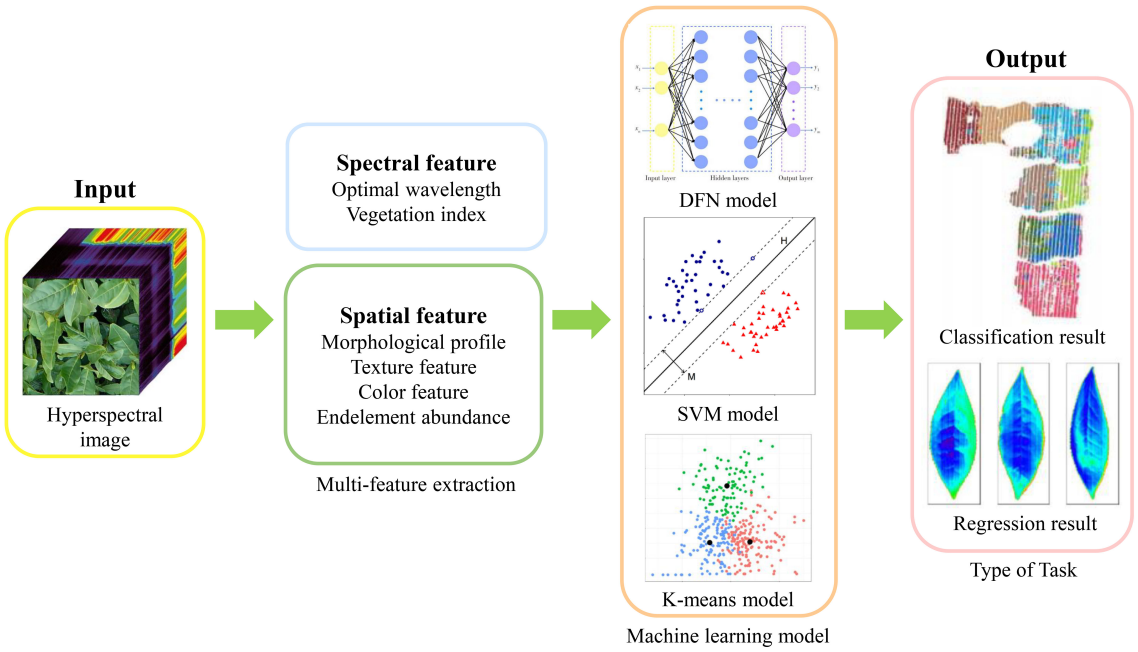


FIGURE 4
Workflow for constructing the phenotypic information detection model.

Wu et al. the constructed vegetation index variable VI_4 ($VI_4 = SD_r/SD_y$) not only had a correlation coefficient as high as 0.799 with the nitrogen content of tea tree leaves, but also the index model constructed using VI_4 had the lowest RMSE of 0.1029, which showed an excellent fitting effect (Wu et al., 2018). On the other hand, statistical analysis methods rely on the entire spectral range or selected feature bands, aiming to comprehensively recognize patterns hidden amongst multiple variables, thereby facilitating an in-depth understanding of the overall data characteristics. Regression algorithms, one class of statistical analysis methods, are often used in plant phenotyping task, to detect continuous plant traits such as plant height, chlorophyll content. The most commonly used multivariate statistical models in current research are the Multiple Linear Regression (MLR) model (Huang et al., 2020) and the Partial Least Squares Regression (PLSR) model (Cao et al., 2022a). The PLSR model, in particular, is designed with the correlation among independent variables in mind and has the capacity to transform variables in the original spectral data that exhibit multicollinearity into a small number of independent latent variables. Therefore, PLSR models typically exhibit superior predictive efficacy, thus it is regarded as a fundamental method for hyperspectral data analysis. Besides regression task, classification and clustering are also common phenotypic information detection models (Plaza et al., 2009; Layne et al., 2020). The former aims to predict discrete category labels or to distinguish image pixels, such as tea plant breeds identification, environmental stress diagnosis, and disease and pest regions detection. The latter can be used to find the internal structure of a data set. Tea plant samples can be divided into autonomous groups by applying clustering techniques [e.g., K-means (Shan et al., 2018)], which allows for higher similarity between data within the same category than between data belonging to other categories (Davis et al., 2020). Clustering algorithms can also be employed in image segmentation, as it can proficiently discern and identify objects and shapes within images (Zhang and Xu, 2018; Nikbakhsh et al., 2020).

Deep learning, as a new branch of machine learning, focuses on the architecture and function of neural networks, particularly deep neural networks. In recent years, it has demonstrated outstanding performance in applications such as agricultural pest and disease detection (Golhani et al., 2018; Xiao et al., 2022), and stress monitoring (Singh et al., 2018). Compared to conventional machine learning approaches, deep learning models autonomously discover latent patterns and features within data, obviating the need for manual feature engineering and selection. Deep neural networks consist of multiple hidden layers, each transforming input data, enabling deeper networks to capture more intricate features (LeCun et al., 2015; Zhao et al., 2022b). This capability makes deep learning especially suitable for handling high-dimensional and large-scale complex datasets. Currently, some deep learning models have been applied to phenotypic studies of tea plants. For instance, models based on convolutional neural networks (CNN) exhibit superior overall performance in quantifying low temperature stress in tea leaves (Mao et al., 2023a), compared to traditional machine learning methods. As deep learning technologies advance, the introduction of generative models and the proliferation of large-scale models in the processing of agricultural hyperspectral data deserve particular attention. Through generative models, effective data augmentation and synthesis, data dimensionality reduction, anomaly detection, and spectral unmixing are facilitated, thereby opening new avenues in hyperspectral data processing (Bhatt and Joshi, 2020; Hu et al., 2022; Guerri et al., 2024). These advancements not only enhance the performance of existing methods but also propel hyperspectral data processing towards greater intelligence and efficiency.

4 Application of hyperspectral imaging technology in tea plant phenotyping

With an overview of the researches published recently, we reviewed the application of hyperspectral imaging technology in tea plant phenotyping from three aspects: environmental stress diagnosis, growth state monitor, and quality indicators estimation. Hyperspectral imaging technology could provide high-throughput comprehensive technical support and data source for tea plant phenotyping. In Table 3, we have summarized the application of hyperspectral imaging technology to the phenotypic traits of tea plants.

4.1 Environmental stress diagnosis

The phenotype of plants is influenced by a combination of genetic and environmental factors. Hyperspectral imaging technology has been widely used in studying the phenotype of tea plants under environmental stress, including both biotic and abiotic stress conditions (Sun et al., 2021; Sanaeifar et al., 2023). Physiological changes triggered by these conditions can significantly affect the growth process of tea plants. Accurate monitoring and assessment of the degree of environmental stress on tea plants through hyperspectral imaging technology can not only take timely and reasonable measures to alleviate the negative

TABLE 3 Examples of the application of hyperspectral imaging technology in the detection of tea plant phenotypic information.

Phenotypic Trait	Domain	Data preprocessing	Data processing	Data analysis	Optimal performance	Refs
Quality indicator detection						
Polyphenols	Laboratory	1st Der, Smoothing, Normalization	SPA	MLR, LS-SVM	Rp^2(EGC) = 0.918 Rp^2 (EGCG) = 0.918 Rp^2 (EC) = 0.918 Rp^2 (ECG) = 0.918	(Dutta et al., 2015; Zhang, 2017; Tu et al., 2018)
	Field	WD, CR, SNV, 1st Der, 2nd Der	MNF, PCA, ICA	MLC, MDC, ANN, SVM, PLSR, BPANN	R_{CV} = 0.9322 $RMSE_{CV}$ = 0.9036	
		SR, NR, FDR	SDA, PCA	SMLR, PLSR	R^2 = 0.81 RMSE = 1.39	
Amino acids	Field	WD, CR, SNV,1st Der,2nd Der	MNF, PCA, ICA	MLC, MDC, ANN, SVM, PLSR, BPANN	R_{CV} = 0.9322 $RMSE_{CV}$ = 0.9036	(Tu et al., 2018)
Environmental stress monitoring						
Heavy metal stress	Laboratory	BC, DT, SNV, MSC	SPA, CARS, CARS-SPA	MLR, LS-SVM, PLS	R^2(Chl) = 0.691 R^2 (ASA) = 0.596 R^2 (GSH) = 0.646 R^2(SP) = 0.624	(Jin, 2019)
Drought	Laboratory	SG, 1st Der, 2nd Der, MSC	SPA, CARS, UVE	RF, SVM, PLSR	Rcal = 0.97 Rp = 0.95 RPD = 4.28	(Chen et al., 2021, 2022)
		SG, 1st Der, 2nd Der, MSC, SNV	SPA, CARS, UVE	RF, SVM, PLSR	R^2 = 0.77 RMSE = 0.073 MAPE = 0.16	
Low temperatures	Laboratory	SG, 1st Der, MSC	SPA, CARS, UVE	RF, SVM, PLS, BP, CNN, LSTM	R^2 = 0.890 $RMSE_P$ = 0.325 RPD = 2.904	(Mao et al., 2023a, 2023)
		SG, 1st Der, MSC	SPA, UVE	SVM, PLS, CNN	Rp^2 = 0.812 $RMSE_P$ = 0.008	

(Continued)

TABLE 3 Continued

Phenotypic Trait	Domain	Data preprocessing	Data processing	Data analysis	Optimal performance	Refs
Environmental stress monitoring						
Tea green leafhopper	Laboratory	-	CWA, Spectral derivative features, Continuum removal features	K-means, SVM, FLDA, RF, SVM	OA = 90.69%	(Cui et al., 2021; Zhao et al., 2022a)
	Field	-	SPA, Spectral derivative features, Continuum removal features, Vegetation indices	KNN, RF, Fisher	OA >= 98%	
Anthracnose	Laboratory	SG, 2nd Der, SNV	SPA, PLS-DA	SVM, ELM	$RMSE_P$ = 0.2874 Accuracy = 95.77%	(Yuan et al., 2019; Bing et al., 2021; Cui et al., 2021; Zhao et al., 2022a)
		-	Feature variables, Vegetation indices	ISODATA	OA = 98% Kappa = 0.96	
		-	CWA, Spectral derivative features, Continuum removal features	K-means, SVM, FLDA, RF, SVM	OA = 90.69%	
	Field	-	SPA, Spectral derivative features, Continuum removal features, Vegetation indices	KNN, RF, Fisher	OA >= 98%	
White star Disease	Laboratory	SG, 2nd Der, SNV	SPA, PLS-DA	SVM, ELM	$RMSE_P$ = 0.2874 Accuracy = 95.77%	(Bing et al., 2021)
Growth and development information management						
Water	Laboratory	-	RF, SPA	PLSR, LS-SVR	Rp2 = 0.956 $RMSE_P$ = 0.027	(Sun et al., 2019; Wei et al., 2019; Wang et al., 2020b)
		SG, OSC, MSC, DT	SPA, CARS, SPA-SR, CARS-SR	MLR	Rp2 = 0.8631 $RMSE_P$ = 0.0163	
		-	SPA, CARS	MLR, PLSR	Rp = 0.8543, 0.9357,0.8188	
Chlorophyll	Laboratory	-	-	Vegetation indices	R = 0.8323 RMSE = 8.601	(Zhao et al., 2011; Sonobe et al., 2018a; Wang et al., 2019; Yamashita et al., 2020)
		1st Der, CR, SNV, MSC, DT	-	RF, SVM, Cubist, SGB, KELM	RPD(N) = 0.0163 RPD(Chl) = 0.0163	
		1st Der, SNV, MSC	2nd Der, RC	PLSR	Rp = 0.9322 $RMSE_P$ = 0.9036	
		-	GA	RF, DBN, ELM, SVM	RMSE = 0.36 ± 0.08	
Carotenoid	Laboratory	1st Der, SNV, MSC	2nd Der, RC	PLSR	Rp = 0.9322 $RMSE_P$ = 0.9036	(Sonobe et al., 2018a; Wang et al., 2019)
		-	GA	RF, DBN, ELM, SVM	RMSE = 0.36 ± 0.08	
Nitrogen	Laboratory	SNV	-	PLS-DA, LS-SVM, PLSR	CCR = 92% Rp = 0.924 $RMSE_P$ = 0.209	(Wang et al., 2018; Wu et al., 2018; Yamashita et al., 2020; Wang et al., 2020b, 2020)
		-	PCA, GLGCM	SVM, ELM	CCR = 100%	
		-	SPA, CARS	MLR, PLSR	Rp = 0.8543, 0.9357,0.8188	
			-			

(Continued)

TABLE 3 Continued

Phenotypic Trait	Domain	Data preprocessing	Data processing	Data analysis	Optimal performance	Refs
Growth and development information management						
		1st Der, CR, SNV, MSC, DT		RF, SVM, Cubist, SGB, KELM	RPD(N) = 0.0163 RPD(Chl) = 0.0163	
	Field	-	Feature variables	Regression model	R^2(N) = 0.8303 RMSE(N) = 0.1029 R^2 (LAI) = 0.9 RMSE(LAI) = 0.0876	
Phosphorus, Potassium	Laboratory	SNV, MSC, 1st Der, 2nd Der, Normalization	SPA, RC	PLSR, MLR	Rp = 0.9168 $RMSE_p$ = 0.4941	(Wang et al., 2020a)
Leaf Area Index	Field	-	Feature variables	Regression model	R^2(N) = 0.8303 RMSE(N) = 0.1029 R^2 (LAI) = 0.9 RMSE(LAI) = 0.0876	(Wu et al., 2018)
Growing status	Field	-	Spectral band, Spectral derivative features, Continuum removal features, Vegetation indices	FLDA, RF, GA-SVM	OA = 98% Kappa = 0.97	(Zhang et al., 2019)

impact of stress on tea plant growth, but also help the selection and breeding of resistant tea plant varieties, which in turn promotes the enhancement of tea yield and quality.

4.1.1 Biotic stress

Biotic stress refers to the attack and interference of biological factors such as pathogenic microorganisms, pests, weeds, or other plant competition to which the tea plant is subjected during the growth process (Jeyaraj et al., 2020; Zahir et al., 2022). For example, hyperspectral technology was applied to detect anthracnose, a common foliar disease on tea plants. As shown in Figure 5, Yuan et al. screened the optimal hyperspectral feature set suitable for anthracnose scab detection by observing the red-shift phenomenon due to anthracnose in the wavelength range of 450-950 nm, and further used autocorrelation analysis to accurately and rapidly identify the anthracnose scab region on tea leaves. They combined unsupervised classification and two-dimensional threshold adaptation to construct an analytical framework for scab detection, which achieved an overall accuracy of 94% in recognizing the disease at the pixel level. Compared with traditional pixel-based classification methods, this method can effectively identify diseased tea leaves and analyze the degree of infection, indicating that hyperspectral technology can help improve tea plant disease detection and field management (Yuan et al., 2019). In the discrimination of leaf diseases

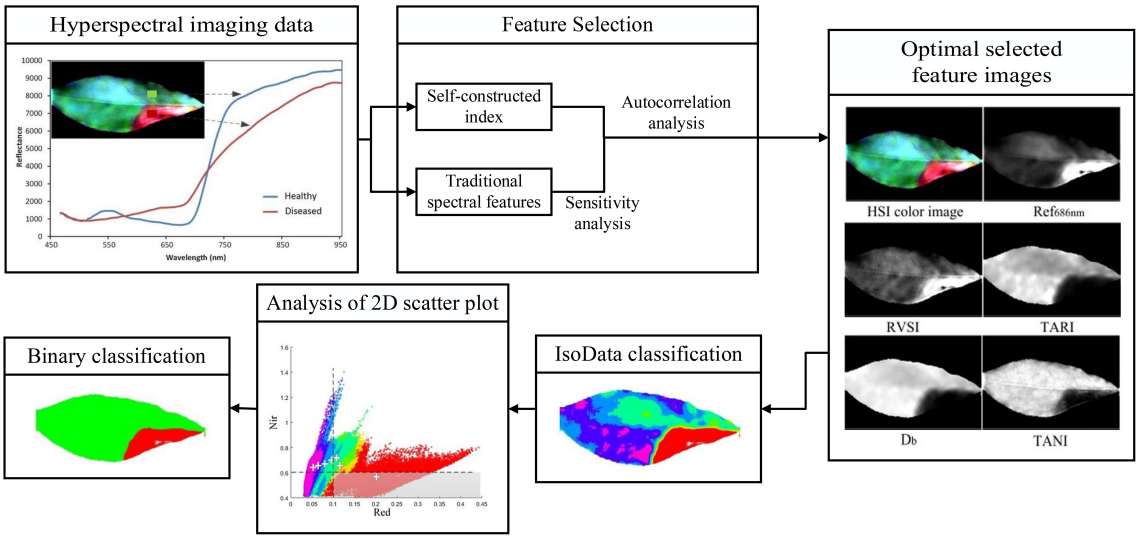


FIGURE 5
Schematic of the workflow for detection of anthracnose in tea plants based on hyperspectral imaging.

with multiple similar symptoms, Lu et al. used hyperspectral technology (420-946nm) for identifying and distinguishing between white star disease and anthracnose that have similar image features. The highly similar morphological characteristics in mean spectra of the whole leaf infected by the two diseases resulted in relatively low data separability, while the mean spectra of the diseased spot regions were significantly different. Therefore, the mean spectra of diseased spot areas extracted after threshold segmentation and mask processing were combined with different machine learning models for classification, and the results of the study showed that the prediction accuracy of the ELM model based on the neural network structure was higher than that of the SVM model using different kernel functions, and its classification accuracy reached 95.77% (Bing et al., 2021). For these two similar diseases, hyperspectral technology can accurately recognize and detect the disease severity at the early stage of the onset of tea plant diseases. Compared to the computer vision technology, the above studies show that hyperspectral technology has certain advantages in the early detection of tea plant diseases, however its application is still mainly based on the external characteristics of tea plants affected by diseases, and it lacks the ability to qualitatively detect tea plants at the early stage of infection by pathogens. Therefore, further researches should take the implementation of predictive analysis in the early infect stage of tea plant disease into consideration, so as to realize the effective prevention and control of tea plant diseases (Lowe et al., 2017).

To minimize the loss caused by pests, pests early detection plays a key role in preventing the spread of pests. In the actual tea garden environment, the combined presence of pests and diseases is a common phenomenon, and it is often difficult to discriminate between plants diseases and pests due to the similarity of their damage symptoms. Cui et al. proposed a method based on canopy hyperspectral data (450-950nm) for detecting and distinguishing between three tea plant stresses with similar external characteristics (anthracnose, tea leafhopper and sunburn). Through spectral sensitivity analysis, the best spectral features in the three processes of removing plant background, identifying plant damaged areas and distinguishing tea plant stresses were determined, and then K-means clustering and KNN were used to construct a model for detecting damaged areas of the tea plant, and finally Fisher linear discriminant method was used to discriminate the stress types in the damaged areas. The validation results of the model showed that the detection accuracy of the damaged areas could reach 95% and the accuracy for stress discrimination was 98% (Cui et al., 2021). As can be seen from the above studies, the research method based on hyperspectral imaging data mainly establishes the relationship between spectral features and the degree of disease or pest types through various statistical discrimination or data mining algorithms, which is similar to the analysis method of the process of identifying the disease of tea plant, and thus reaches an accurate analysis of the infected areas under different types of stresses, revealing the potential application in large-scale tea plantation disease management, particularly in real-time monitoring and accurate diagnosis. By reducing pesticide usage and optimizing pest control strategies, hyperspectral imaging technology can significantly mitigate the negative environmental impact of agricultural production while enhancing both the yield and quality of tea leaves.

4.1.2 Abiotic stress

During the initial period when tea plants are in abiotic stress, tea plants often do not exhibit obvious external characteristic changes. However, significant external trait differences may emerge only when the stress reaches a certain severity level, at which point the stress may have caused irreversible damage to the tea plants. Therefore, early identification and evaluation of characteristic changes in tea plants under abiotic stress are vital for devising timely measures to mitigate or reverse the stress effects (Gerhards et al., 2019; Krishnatreya et al., 2021). As early signs of abiotic stress in tea plants are usually reflected by changes in relevant biochemical parameters within the tea plant tissues during the stress response process, these changes can be captured as corresponding spectral differences. This makes it possible to qualitatively analyze the adaptation level of tea plants to current abiotic stresses. Nowadays, the application of hyperspectral technology to detect abiotic stress levels is being carried out and in-depth research. The following sections will discuss the three considered abiotic stress factors for tea plants: drought, low temperature, and heavy metal.

4.1.2.1 Drought stress

With the global climate warming leading to an increased frequency of extreme weather events and the lack of agricultural water causing ineffective irrigation of tea plants, tea plants are facing more frequent drought stress, negatively impacting their growth, development, and yield. To combat the adverse effects caused by drought stress, tea plants undergo a series of physiological defense mechanisms, including stomatal regulation, osmotic adjustment, and regulation of the antioxidative enzyme system (Liu and Chen, 2014; Tony et al., 2016). Hyperspectral technology was used to understand the changes in physiological and biochemical parameters within these mechanisms to assess the severity of drought stress. Chen et al. obtained five drought-related physiological and biochemical index parameters, namely malondialdehyde (MDA), electrolyte leakage (EL), maximal efficiency of photosystem II (Fv/Fm), soluble saccharide (SS), and drought damage degree (DDD), during the drought stress treatments with different levels in tea seedlings. By trying various data processing algorithms and modeling methods, they predicted the drought stress severity for different tea seedlings. Among these, UVE-SVM detection model for the drought damage degree (DDD) achieved a high prediction accuracy (Rcal=0.97, Rp=0.95, RPD=4.28), providing a comprehensive and objective assessment of the tea plants' drought resistance (Chen et al., 2021).

In another study by the same research team, hyperspectral imaging technology (400-1000nm) was used to monitor drought stress in 10 different tea germplasm resources to determine their drought tolerance, proving the feasibility and effectiveness of hyperspectral technology in screening for drought-resistant tea plant germplasm. 3 indicators (malondialdehyde, soluble sugar, and total phenol) directly used in drought evaluation positively contributed to constructing a comprehensive index, DTC (Drought

Tolerance Coefficient), for assessing the drought tolerance of tea plants. The results indicated that the model constructed with the MSC-2D-UVE-SVM method showed the best performance in screening for drought-resistant tea germplasm resources (R^2=0.77, RMSE=0.073, MAPE=0.16) (Chen et al., 2022). Based on the above research, tea plant drought stress detection using hyperspectral technology, from the initial stages of genetic selection to the entire cycle of plant life activities, have acquired real-time precise monitoring technologies for traits exhibited by tea plants under drought stress. However, the tea plants' growing environments and their growth stages causes variability in their drought stress responses to some extent. Therefore, further research into tea plants with complex phenotypic trait changes must be conducted, and these studies will bring greater prospects for improvement in tea plant breeding optimization and precision cultivation. In general, current research on environmental stress detection in tea plants predominantly focuses on single types of stress, overlooking the reality that tea plants often encounter multiple environmental stresses simultaneously in their natural growth processes (Safavi-Rizi et al., 2021). For instance, tea plants might concurrently endure the dual stresses of drought and high temperatures during the same period, a scenario in which their phenotypic responses could markedly differ from those observed under singular conditions of drought or heat stress alone (Tang et al., 2023). This focus could severely limit the applicability of research findings, leading to an incomplete understanding of the overall tea plant phenotype and an inability to reveal the complex interplay and comprehensive responses of environmental factors that influence the growth and physiology of tea plants (Cao et al., 2015).

4.1.2.2 Low temperature stress

Low temperature environment is one of the important constraints on the growth and development rate, tea yield and natural geographic distribution of tea plants. Under low temperature stress, the cell membrane structure and function of the tea plant will be damaged, triggering a series of abnormal physiological and biochemical reactions, leading to cellular metabolic disorders, and thus causing damage to the tea plant. Relevant studies have shown that tea plant produces protective enzymes such as catalase (CAT), peroxidase (POD), and superoxide dismutase (SOD) to scavenge excessive active oxygen free radicals accumulated under low temperature stress conditions (Wang et al., 2021). Therefore, changes in these protective enzymes can be tracked using hyperspectral techniques to measure the extent of cold stress.

In addition, some physiological indicators such as malondialdehyde, soluble proteins, and soluble sugars involved in the cold resistance response can also be used as dependent factors for assessing the degree of low temperature stress suffered by tea plants. Mao et al. investigated the spectral changes of tea leaves after exposure to low temperature stress at 8h and 12h intervals under normal (25°C), chilling (4°C), and freezing (-4°C) conditions. Various variable selection methods and machine learning models were used to analyze the contents of chlorophyll (SPAD value), soluble sugar (SS), and malondialdehyde (MDA), as well as the activities of catalase (CAT), peroxidase (POD), superoxide dismutase (SOD), and the Low Temperature Response Index (LTRI). The UVE-CNN model demonstrated high predictive accuracy corresponding to LTRI, with Rp^2= 0.8631 and RMSEP = 0.325 (Mao et al., 2023a). This study identified the potential advantages of applying hyperspectral techniques to detect low temperature stress in tea plants, which could provide accurate data support for preventing damage caused by frost damage in tea plants. The multi-parameters evaluation model was generally superior to the single physicochemical index model in analyzing abiotic stress in tea tree. This may be due to the multi-parameters evaluation model in tea plant can be as internal standards of each other and their response to low temperature stress were considered. However, the selection of biochemical components in tea plant, and the mechanism between them and low temperature stress may need further investigation.

4.1.2.3 Heavy metal stress

Heavy metals in tea plant mainly originate from anthropogenic factors, such as industrial emissions and excessive use of agricultural fertilizers, which cause pollution of air, water and soil. These heavy metals mainly enter the tea plant through two pathways, root absorption and foliar infiltration, leading to impaired nutrient absorption and metabolic function, and hindering the normal growth and development of the tea plant (Zhang J. et al., 2018; Sanaeifar et al., 2022). In addition, heavy metals may accumulate in tea leaves, posing a serious threat to human health. Therefore, timely monitoring and management of heavy metal pollution is crucial for tea plantations. Like other abiotic stresses, some stress-related physiological indexes of tea plants under heavy metal stress also changed with the progress of stress.

Jin et al. proposed a method using hyperspectral imaging (420-946 nm) combined with chemometrics to quantify the changes of four relevant indicators, chlorophyll (Chl), ascorbic acid (ASA), glutathione (GSH)and soluble protein (SP), in the foliage and roots of tea plant at different times and under different lead stresses. The quantitative analysis model for each indicator was established by optimal pretreatment and selection of characteristic wavelengths. All the models based on MLR were simple and efficient while showing good predictive effects, and the model prediction coefficient of determination Rp^2 could reach about 0.6~0.9 (Jin, 2019). Similarly, Sanaeifar et al. successfully applied a PLS-RBFNN model based on visible and near-infrared (Vis-NIR) spectroscopy to precisely predict the quality indices of tea leaves under foliar stress at different intervals and varying treatment levels (Sanaeifar et al., 2022). Both methodologies offer an effective approach for the rapid analysis and accurate detection of the physiological status of tea plant under heavy metal stress. However, the problem of heavy metal contamination from air and soil that tea plant may face is not limited to a single heavy metal, and there are differences in the effects of different heavy metals on tea trees. Therefore, the detection of various heavy metal stresses has the potential to become a valuable research direction in the field of hyperspectral technology-based detection of tea plant phenotypic information (Zhang et al., 2021).

4.2 Growth and developmental information monitor

Information on the growth and development of the tea plant mainly includes water, photosynthetic pigments, nutrient elements, and morphological and structural parameters of the tea plant. This information can be used to gain an in-depth understanding of the real-time physiological state of the tea tree and the complexities of the long-term growth process, which in turn provides a scientific basis for assessing and improving tea yields (Li et al., 2022). The application of hyperspectral technology allows for more accurate quantitative analysis and visual representation of the phenotypic parameters to be measured in a non-destructive manner.

Water is the most basic life element of the tea plant, which is directly involved in the growth and development of the tea plant, physiological metabolism and the construction of morphological structure. Meanwhile, water also plays an important regulating role in the formation of tea quality components. Therefore, real-time and accurate detection of tea plant water status is of great significance in tea plant cultivation and tea garden irrigation. Estimating the moisture content of tea plant leaves has been an important area of research in hyperspectral technology (Zhang X. et al., 2018). Researchers have used near-infrared hyperspectral imaging in conjunction with a variety of combinatorial algorithms for quantitative modeling and visualization of tea leaf moisture distribution (Sun et al., 2019). In addition, Wei et al. considered that the hyperspectral data (380-1030nm) of the front and back sides of the tea leaves would be different during the tea production process, and designed a logistic classification regressor with a 100% correct classification rate to identify the front and back sides of the tea leaves, and then used the characteristic spectral bands corresponding to the front and back sides of the tea leaves to build a Least Squares Support Vector Machine (LS-SVR) model to generate tea leaf water content distribution maps (Wei et al., 2019). These published studies have mainly focused on the accurate analysis of tea plant water content at the leaf scale. In order to map the overall moisture status of tea trees more comprehensively, future studies should aim to extend the assessment to the canopy level of tea plants. Research in the field of monitoring the water status of other crops, such as the studies by Lü et al., has demonstrated the potential of canopy hyperspectral imaging technology to quantify and predict the water status of Arabidopsis thaliana (Lü et al., 2020). Similarly, Wocher et al. have shown a significant relationship between canopy reflectance and moisture content in winter wheat and maize (Wocher et al., 2018). These studies underscore the effective application of spectral data at the canopy scale across different crops, providing crucial foundations and references for extending this approach to monitoring the water status of tea plants.

The photosynthetic pigments in the tea plant are mainly composed of chlorophyll and carotenoid, which are important components of photosynthesis. The content of photosynthetic pigments directly reflects the photosynthetic capacity of tea plants and the degree of environmental adaptation. At present, hyperspectral technology has become an effective tool for accurately estimating the content of photosynthetic pigments and their spatial distribution in tea tree leaves. For example, Zhao et al. successfully calculated the chlorophyll concentration at the scale of image elements by the MSAVI2 (Modified secondary soil adjustment vegetation index) prediction model (Zhao et al., 2011). Meanwhile, the joint measurement of chlorophyll and carotenoid content has the potential to detect environmental stresses. Sonobe et al. used the KELM (Kernel-based Extreme Learning Machine) model to establish the relationship between hyperspectral data (400-900nm) and the content of the three photosynthetic pigments (chlorophylls a, b, and carotenoids), demonstrating their accuracy, with root mean square errors of 1.95 ± 0.36, 1.08 ± 0.11 and 0.68 ± 0.10 μg/cm^2 for each photosynthetic pigment corresponding model, respectively (Sonobe et al., 2018a). In addition, nitrogen in tea plant leaves mainly exists in photosynthetic pigments, so the nitrogen nutritional status of tea trees can be effectively monitored by grasping the dynamic changes of photosynthetic pigment content in real time. Using hyperspectral imaging technology, through the SNV-PLSR model constructed on the basis of the entire wavelength range of 400-1000 nm and the 2-Der-PLSR model constructed on the basis of characteristic wavelengths, Wang et al. highly accurate prediction of chlorophyll a, chlorophyll b, total chlorophyll and carotenoids contents in tea leaves under different nitrogen application levels was successfully realized (Wang et al., 2019). This series of studies vividly demonstrates the high precision and efficiency of hyperspectral technology in measuring photosynthetic pigments in tea plants. The dynamic changes in plant photosynthetic pigments are closely linked to their physiological state. Subsequent research can translate the spatiotemporal dynamics of photosynthetic pigments into quantitative assessments of key indicators for tea plant growth and development, thereby expanding its potential applications in actual tea production.

Nitrogen, phosphorus and potassium are essential nutrients for the growth, development and yield formation of tea plants, and they usually supply in the form of fertilizers to tea plants. Efficient, non-destructive, and accurate monitoring of the nutrient status of tea trees can help to diagnose the growth status of tea plants, improve the efficiency of fertilizer application, and reduce the damage of over-fertilization to soil nutrients and the negative impact on the environment, so as to realize the precision management of tea plant cultivation. Wang et al. utilized hyperspectral technology combined with a multivariate classification algorithm to discriminate the nitrogen application level of tea trees. The study firstly extracted the five optimal wavelengths by principal component analysis from the hyperspectral data containing 553 wavelengths in the spectral range of 400-1000 nm, and then extracted texture features from the images of the optimal wavelengths by using the gray level gradient covariance matrix (GLGCM). Finally, the SVM model using fused data gave the best performance with highest correct classification rate of 100% for prediction set (Wang et al., 2018). In a related study, the research team combined hyperspectral image technology (400-1000nm) with chemometrics methods to accurately discriminate the nitrogen level and status of tea plants by using the SNV-LS-SVM model, and its correct classification rate reached 82% and 92%, respectively. Meanwhile, the nitrogen content of tea

tree leaves was successfully quantified by the SNV-PLSR model with a correlation coefficient Rp of 0.924 (Wang et al., 2020c). In addition, the research team extended its research results to the determination of phosphorus and potassium content in tea leaves (Figure 6), and successfully achieved high-precision prediction results of phosphorus and potassium content in tea samples of different varieties with correlation coefficients Rp of 0.9423 and 0.9168, respectively, by the SPA-MLR prediction model (Wang et al., 2020a). This series of studies have demonstrated the promising application of hyperspectral technology in dynamic nutrient tracking and field fertilization management of tea trees, but there are many challenges and limitations to be faced in translating the results of the studies into practical applications. One of the common problems is the weak transferability of the constructed prediction models, which mainly stems from the influence of different tea plant breeds, geographical locations and environmental factors on the prediction of tea tree growth and development information. Therefore, diagnostic models with broader applicability by relying on hyperspectral data of different tea tree species under various environmental conditions should be built in the future.

Morphological and structural parameters of tea tree are also an important topic in the study of tea plant growth information detection, which mainly involves leaf area, biomass and crown width and other related parameters. Various types of vegetation indices derived from hyperspectral data can be used to monitor the changes of morphological and structural parameters, effectively assess the growth status of tea plant, and reflect the production potential of tea plant. To address the problems of time-consuming and destructive in the traditional manual detection methods of tea tree LAI (leaf area index), Wu et al. analyzed the correlation between LAI and hyperspectral feature variables, screened out the three feature variables with the highest correlation (amplitude of yellow edge, green peak reflectance and red valley position), and set up different inversion models through the three variables, among which the logarithmic fitting model with green peak reflectance R_g as the independent variable was modeled better with validation sample R^2 and RMSE values of 0.764 and 0.0876, respectively. This study verified the feasibility of applying hyperspectral technology for real-time, rapid and non-destructive measurement of tea tree leaf area index method (Wu et al., 2018). As the morphological and structural characteristics of tea tree are more intuitive in observation, conventional spectral detection techniques can already provide more accurate results. Hyperspectral techniques have not generated a strong demand in the study of tea tree morphological structural parameters due to their high economic cost and complexity of analyzing and processing the data. However, through an in-depth analysis of hyperspectral data, its rich information across spectral bands and high-resolution spectral features can compensate for the limitations of traditional methods in terms of vegetation spatial resolution and data sensitivity, enabling the construction of vegetation indices capable of accurately detecting morphological structural parameters in remote sensing.

4.3 Quality indicator detection

The quality of tea is one of the key factors determining its market value and processing methods. The content of organic chemicals in tea plant leaf, such as tea polyphenols, amino acids and caffeine, is an important indicator of tea quality and flavor (Peng et al., 2016; Cao et al., 2020). Traditionally, the content of these organic chemicals has been assessed by destructive sampling followed by wet chemical

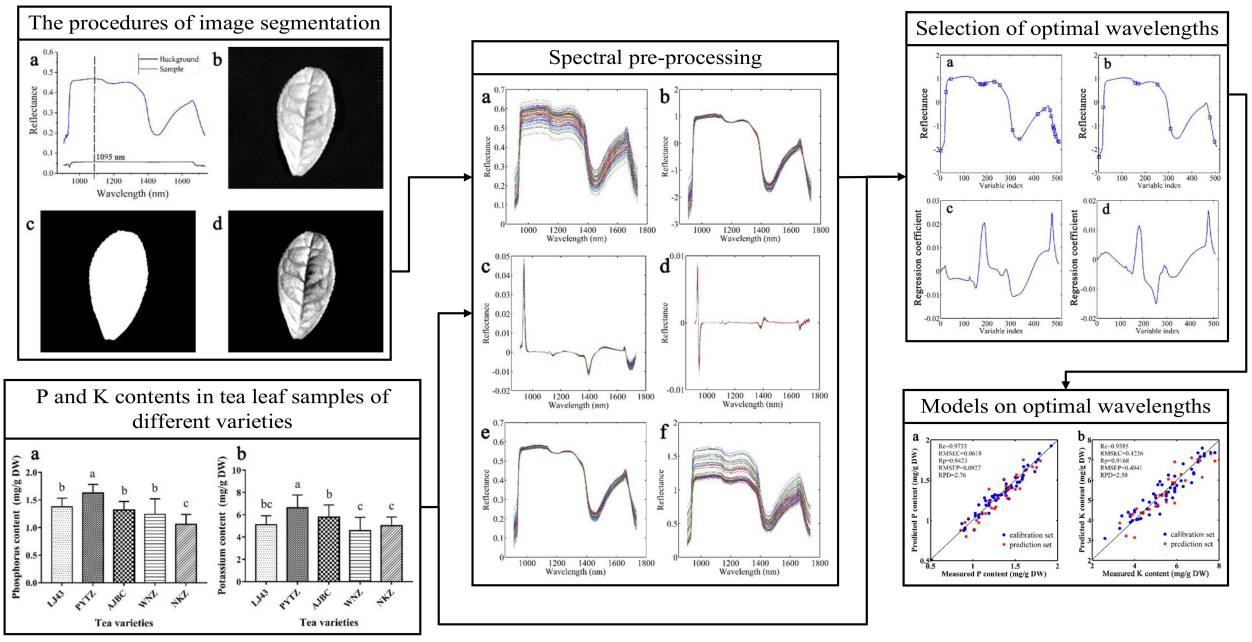


FIGURE 6
Illustration of the image processing and data analysis process used for quantifying the phosphorus and potassium content in different tea plant cultivars.

methods, which are time-consuming and labor-intensive and difficult to adequately represent spatial variability in whole tea plant. Dutta et al. analyzed *in situ*-measured hyperspectral data in the range of 347-2506 nm by applying different multivariate analytical methods to screen the best predictive models, and developed a partial least squares regression (FDR-PLSR) model for first-order derivative spectra to achieve accurate estimation of tea polyphenol content at the leaf level, with R^2 and RMSE values of 0.81 and 1.39 for the model, respectively (Dutta et al., 2015). Hyperspectral techniques also offer the possibility of estimating and monitoring the quality of tea on a spatial scale. Tu et al. accurately classified tea tree varieties in large-scale tea plantations based on the spectral characteristics of tea tree canopy data acquired by a 450-998 nm hyperspectral camera mounted on a UAV. The hyperspectral data were also processed through partial least squares regression (PLSR) using standard normal variable transformation (SNV) preprocessing method to predict tea polyphenol and amino acid contents, although the model showed a relatively significant prediction accuracy (R_{CV} = 0.66, $RMSE_{CV}$ = 13.27) in terms of the ratio of tea polyphenols to amino acids (the main indicator of tea sensory quality), which proves that there is an association between the spectral data and quality indicators. To apply this correlation to the prediction of actual tea tree quality indicators based on the UAV platform, more in-depth research and exploration are still needed to improve the accuracy of the model (Tu et al., 2018). On the basis of existing researches, hyperspectral technology can be used to achieve independent analysis and measurement of multiple subcomponent contents of the main components of tea, in order to accurately feedback the current quality of tea and provide favorable guidance for the optimization of the subsequent production and processing processes. On the other hand, current research primarily relies on spectral data obtained in a single instance, without considering the temporal dynamics of tea plant phenotypic characteristics throughout phenological cycles (Mahlein et al., 2018). Phenotypic expression in tea plants may vary over time, necessitating continuous observation to accurately capture these dynamics. It is recommended that future studies expand the range of tea plant sample collection to encompass phenotypic differences in different environments. By conducting long-term observational studies to establish comprehensive tea plant phenotypic databases, the statistical significance and reliability of research can be enhanced, facilitating the development of more refined and generalizable models (Wijewardane et al., 2023).

5 Summary and outlook

In this paper, we comprehensively review the research progress of hyperspectral imaging technology in the field of tea plant phenotyping information detection, the main applications of which involve environmental stress diagnosis, growth and development information monitoring, and quality indicator detection of tea plant. Hyperspectral technology in the detection of tea plant phenotypic information is still in its nascent stage, with notable gaps in research. Despite the considerable enrichment of our understanding of plant phenotypes facilitated by this technique, the issue of potentially missed critical phenotypic information due to restricted acquisition at specific wavelengths or resolutions remains pertinent. It thus becomes markedly important to introduce more advanced hyperspectral technologies to enhance the accuracy and comprehensiveness of data collection. In the data processing and analysis phase, hyperspectral data necessitates navigation through a complex and meticulous data processing workflow. Errors occurring at any stage within this workflow harbor the potential to precipitate data inconsistencies. Consequently, the development of more efficacious data processing and analysis algorithms is instrumental in extracting more precise and reliable information from hyperspectral data. In addition, in order to ensure the comparability of data between different studies, it is of great significance to further establish standardized hyperspectral data processing and analysis methods. In application, current research on tea plant phenotypic detection also exhibits certain limitations, particularly the absence of a comprehensive method that spans multiple observational scales, involving leaf organs, plants individual, and populations. Integrating observations from multiple scales is crucial for a comprehensive understanding of the phenotypic information of tea plants. However, a unified methodological framework to integrate findings across different scales has yet to be established. In addition, future research should focus on the cross-stress response of tea plant, in order to deconstruct the association between phenotypic characteristics of tea tree in a more comprehensive perspective, and thus promote the deep optimization of tea plant breeding strategies and cultivation management.

Author contributions

BL: Writing – review & editing. HS: Project administration, Writing – original draft. LZ: Writing – review & editing. FC: Conceptualization, Methodology, Writing – review & editing. KW: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The Project was supported by Zhejiang Provincial Natural Science Foundation of China (No. LQ21C130007), “Pioneer” and “Leading Goose” R&D Program of Zhejiang (No. 2022C04002) and Wenzhou Major Science and Technology Innovation Project (No. ZNF2023007).

Acknowledgments

The authors thank Zhejiang Provincial Natural Science Foundation.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adao, T., Hruska, J., Pádua, L., Bessa, J., Peres, E., Morais, R., et al. (2017). Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sens.* 9, 1110. doi: 10.3390/rs9111110

Ang, K. L. M., and Seng, J. K. P. (2021). Big data and machine learning with hyperspectral information in agriculture. *IEEE Access.* 9, 36699–36718. doi: 10.1109/Access.6287639

Atefi, A., Ge, Y. F., Pitla, S., and Schnable, J. (2021). Robotic technologies for high-throughput plant phenotyping: contemporary reviews and future perspectives. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.611940

Bai, X. B., Yu, J. S., Fu, Z. T., Zhang, L. X., and Li, X. X. (2020). Application of spectral imaging technology for detecting crop disease information: A review. *Spectrosc. Spect. Anal.* 40, 350–355. doi: 10.3964/j.issn.1000-0593(2020)02-0350-06

Barnes, R. J., Dhanoa, M. S., and Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 43, 772–777. doi: 10.1366/0003702894202201

Bhatt, J. S., and Joshi, M. V. (2020). "Deep learning in hyperspectral unmixing: A review," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium* (IEEE, Waikoloa, HI, USA), 2189–2192. doi: 10.1109/IGARSS39084.2020

Bing, L., Sun, J., Yang, N., Wu, X. H., and Zhou, X. (2021). Identification of tea white star disease and anthrax based on hyperspectral image information. *J. Food Process Eng.* 44, e13584. doi: 10.1111/jfpe.13584

Cao, D., Lin-Long, M. A., Jin, X. F., and Gong, Z. M. (2015). Research advance of resistance to abiotic stresses of tea. *Hunan. Agric. Sci.* 10, 152–154. doi: 10.16498/j.cnki.hnnykx.2015.10.043

Cao, Q., Yang, G. J., Duan, D. D., Chen, L. Y., Wang, F., Xu, B., et al. (2022a). Combining multispectral and hyperspectral data to estimate nitrogen status of tea plants (Camellia sinensis (L.) O. Kuntze) under field conditions. *Comput. Electron. Agric.* 198, 107084. doi: 10.1016/j.compag.2022.107084

Cao, Q., Yang, G. J., Wang, F., Chen, L. Y., Xu, B., Zhao, C. J., et al. (2022b). Discrimination of tea plant variety using *in-situ* multispectral imaging system and multi-feature analysis. *Comput. Electron. Agric.* 202, 107360. doi: 10.1016/j.compag.2022.107360

Cao, Y., Li, H. R., Sun, J., Zhou, X., Yao, K. S., and Nirere, A. (2020). Nondestructive determination of the total mold colony count in green tea by hyperspectral imaging technology. *J. Food Process Eng.* 43, e13570. doi: 10.1111/jfpe.13570

Centner, V., Massart, D. L., de Noord, O. E., de Jong, S., Vandeginste, B. M., and Sterna, C. (1996). Elimination of uninformative variables for multivariate calibration. *Anal. Chem.* 68, 3851–3858. doi: 10.1021/ac960321m

Çetin, V., and Yıldız, O. (2022). A comprehensive review on data preprocessing techniques in data analysis. *Pamukkale Üniv Mühendis Bilim Derg.* 28, 299–312. doi: 10.5505/pajes.2021.62687

Chappelle, E. W., Kim, M. S., and McMurtrey Iii, J. E. (1992). Ratio analysis of reflectance spectra (RARS): an algorithm for the remote estimation of the concentrations of chlorophyll a, chlorophyll b, and carotenoids in soybean leaves. *Remote Sens. Environ.* 39, 239–247. doi: 10.1016/0034-4257(92)90089-3

Chen, S. Z., Gao, Y., Fan, K., Shi, Y. J., Luo, D. N., Shen, J. Z., et al. (2021). Prediction of drought-induced components and evaluation of drought damage of tea plants based on hyperspectral imaging. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.695102

Chen, S. Z., Shen, J. Z., Fan, K., Qian, W. J., Gu, H. L., Li, Y. C., et al. (2022). Hyperspectral machine-learning model for screening tea germplasm resources with drought tolerance. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1048442

Cheng, J. H., Sun, J., Yao, K. S., Xu, M., and Cao, Y. (2022). A variable selection method based on mutual information and variance inflation factor. *Spectrochim. Acta A.* 268, 120652. doi: 10.1016/j.saa.2021.120652

Chu, X. L., Yuan, H. F., and Lu, W. Z. (2004). Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique. *Prog. Chem.* 16, 528–542. doi: 10.3321/j.issn:1005-281X.2004.04.008

Cui, L. H., Yan, L. J., Zhao, X. H., Yuan, L., Jin, J., and Zhang, J. C. (2021). Detection and discrimination of tea plant stresses based on hyperspectral imaging technique at a canopy level. *Phyton. Int. J. Exp. Bot.* 90, 621–634. doi: 10.32604/phyton.2021.015511

Datt, B. (1999). A new reflectance index for remote sensing of chlorophyll content in higher plants: tests using Eucalyptus leaves. *J. Plant Physiol.* 154, 30–36. doi: 10.1016/S0176-1617(99)80314-9

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., De Colstoun, E. B., and McMurtrey Iii, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens. Environ.* 74, 229–239. doi: 10.1016/S0034-4257(00)00113-9

Davis, R. L., Greene, J. K., Dou, F. G., Jo, Y. K., and Chappell, T. M. (2020). A practical application of unsupervised machine learning for analyzing plant image data collected using unmanned aircraft systems. *Agronomy* 10, 633. doi: 10.3390/agronomy10050633

Delalieux, S., Somers, B., Hereijgers, S., Verstraeten, W. W., Keulemans, W., and Coppin, P. (2008). A near-infrared narrow-waveband ratio to determine Leaf Area Index in orchards. *Remote Sens. Environ.* 112, 3762–3772. doi: 10.1016/j.rse.2008.05.003

Dutta, D., Das, P. K., Bhunia, U. K., Singh, U., Singh, S., Sharma, J. R., et al. (2015). Retrieval of tea polyphenol at leaf level using spectral transformation and multi-variate statistical approach. *Int. J. Appl. Earth Obs. Geoinf.* 36, 22–29. doi: 10.1016/j.jag.2014.11.001

ElMasry, G., Wang, N., and Vigneault, C. (2009). Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks. *Postharvest Biol. Technol.* 52, 1–8. doi: 10.1016/j.postharvbio.2008.11.008

Fensholt, R., and Sandholt, I. (2003). Derivation of a shortwave infrared water stress index from MODIS near- and shortwave infrared data in a semiarid environment. *Remote Sens. Environ.* 87, 111–121. doi: 10.1016/j.rse.2003.07.002

Fu, Y. Y., Yang, G. J., Song, X. Y., Li, Z. H., Xu, X. G., Feng, H. K., et al. (2021). Improved estimation of winter wheat aboveground biomass using multiscale textures extracted from UAV-based digital images and hyperspectral feature analysis. *Remote Sens.* 13, 581. doi: 10.3390/rs13040581

Galvao, L. S., Formaggio, A. R., and Tisot, D. A. (2005). Discrimination of sugarcane varieties in southeastern Brazil with EO-1 hyperion data. *Remote Sens. Environ.* 94, 523–534. doi: 10.1016/j.rse.2004.11.012

Gao, B.-C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58, 257–266. doi: 10.1016/S0034-4257(96)00067-3

Gao, J. (2006). *Canopy Chlorophyll Estimation With Hyperspectral Remote Sensing* (Manhattan: Kansas State University).

Geladi, P., MacDougall, D., and Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc.* 39, 491–500. doi: 10.1366/0003702854248656

Gerhards, M., Schlerf, M., Mallick, K., and Udelhoven, T. (2019). Challenges and future perspectives of multi-/hyperspectral thermal infrared remote sensing for crop water-stress detection: A review. *Remote Sens.* 11, 1240. doi: 10.3390/rs11101240

Golhani, K., Balasundram, S. K., Vadamalai, G., and Pradhan, B. (2018). A review of neural networks in plant disease detection using hyperspectral data. *Inf. Process. Agric.* 5, 354–371. doi: 10.1016/j.inpa.2018.05.002

Guerri, M. F., Distante, C., Spagnolo, P., Bougourzi, F., and Taleb-Ahmed, A. (2024). Deep learning techniques for hyperspectral image analysis in agriculture: A review. *ISPRS Open J. Photogramm. Remote Sens.* 12, 100062. doi: 10.1016/j.ophoto.2024.100062

Hagen, N., and Kudenov, M. W. (2013). Review of snapshot spectral imaging technologies. *Opt. Eng.* 52, 90901. doi: 10.1117/1.OE.52.9.090901

He, Z. T., Wu, K. H., Wang, F. M., Jin, L. S., Zhang, R. X., Tian, S. P., et al. (2023). Fresh yield estimation of spring tea via spectral differences in UAV hyperspectral images from unpicked and picked canopies. *Remote Sens.* 15, 1100. doi: 10.3390/rs15041100

Hu, X., Xie, C., Fan, Z., Duan, Q., Zhang, D., Jiang, L., et al. (2022). Hyperspectral anomaly detection using deep learning: A review. *Remote Sens.* 14, 1973. doi: 10.3390/rs14091973

Huang, Y. F., Dong, W. T., Sanaeifar, A., Wang, X. M., Luo, W., Zhan, B. S., et al. (2020). Development of simple identification models for four main catechins and caffeine in fresh green tea leaf based on visible and near-infrared spectroscopy. *Comput. Electron. Agric.* 173, 105388. doi: 10.1016/j.compag.2020.105388

Jeyaraj, A., Elango, T., Li, X. H., and Guo, G. Y. (2020). Utilization of microRNAs and their regulatory functions for improving biotic stress tolerance in tea plant [*Camellia sinensis*(L.) O. Kuntze]. *RNA Biol.* 17, 1365–1382. doi: 10.1080/15476286.2020.1774987

Jin, J. (2019). Detection of Heavy Metal Stress in Camellia sinensis and Physiological Indices Changes Based on Spectral Technology. Zhejiang University, Hangzhou Zhejiang.

Krishnatreya, D. B., Agarwala, N., Gill, S. S., and Bandyopadhyay, T. (2021). Understanding the role of miRNAs for improvement of tea quality and stress tolerance. *J. Biotechnol.* 328, 34–46. doi: 10.1016/j.jbiotec.2020.12.019

Kuswidiyanto, L. W., Noh, H. H., and Han, X. Z. (2022). Plant disease diagnosis using deep learning based on aerial hyperspectral images: A review. *Remote Sens.* 14, 6031. doi: 10.3390/rs14236031

Layne, E., Dort, E. N., Hamelin, R., Li, Y., and Blanchette, M. (2020). Supervised learning on phylogenetically distributed data. *Bioinformatics* 36, I895–I902. doi: 10.1093/bioinformatics/btaa842

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Li, H., Wang, Y., Fan, K., Mao, Y. L., Shen, Y. Z., and Ding, Z. T. (2022). Evaluation of important phenotypic parameters of tea plantations using multi-source remote sensing data. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.898962

Li, H. D., Liang, Y. Z., Xu, Q. S., and Cao, D. S. (2009). Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal. Chim. Acta* 648, 77–84. doi: 10.1016/j.aca.2009.06.046

Li, Q. L., He, X. F., Wang, Y. T., Liu, H. Y., Xu, D. R., and Guo, F. M. (2013). Review of spectral imaging technology in biomedical engineering: achievements and challenges. *J. Biomed. Opt.* 18, 100901. doi: 10.1117/1.JBO.18.10.100901

Liu, S. C., and Chen, L. (2014). Research advances on the drought-resistance mechanism and strategy of tea plant. *J. Tea Sci.* 34, 111–121. doi: 10.3969/j.issn.1000-369X

Lowe, A., Harrison, N., and French, A. P. (2017). Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* 13, 80. doi: 10.1186/s13007-017-0233-z

Lu, B., Dao, P. D., Liu, J. G., He, Y. H., and Shang, J. L. (2020). Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sens.* 12, 2659. doi: 10.3390/rs12162659

Lü, M. Q., Song, Y. J., Weng, H. Y., Sun, D. W., Dong, X. Y., Fang, H., et al. (2020). Effect of near infrared hyperspectral imaging scanning speed on prediction of water content in Arabidopsis. *Spectrosc. Spect. Anal.* 40, 3508–3514. doi: 10.3964/j.issn.1000-0593(2020)11-3508-07

Mahlein, A. K. (2016). Plant disease detection by imaging sensors - parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis.* 100, 241–251. doi: 10.1094/PDIS-03-15-0340-FE

Mahlein, A. K., Kuska, M. T., Behmann, J., Polder, G., and Walter, A. (2018). Hyperspectral sensors and imaging technologies in phytopathology: state of the art. *Annu. Rev. Phytopathol.* 56, 535–558. doi: 10.1146/annurev-phyto-080417-050100

Mao, Y. L., Li, H., Wang, Y., Fan, K., Shen, J. Z., Zhang, J., et al. (2023a). Low temperature response index for monitoring freezing injury of tea plant. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1096490

Mao, Y. L., Li, H., Wang, Y., Fan, K., Sun, L. T., Wang, H., et al. (2023b). Quantitative judgment of freezing injury of tea leaves based on hyperspectral imaging. *Spectrosc. Spect. Anal.* 43, 2266–2271. doi: 10.3964/j.issn.1000-0593(2023)07-2266-06

Merton, R., and Huntington, J. (1999). "Early simulation results of the ARIES-1 satellite sensor for multi-temporal vegetation research derived from AVIRIS," in *Proceedings of the eighth annual JPL airborne earth science workshop* (Jet Propulsion Lab., California Inst. of Tech, Pasadena, CA, USA), 9–11.

Milanez, K., Nóbrega, T. C. A., Nascimento, D. S., Galvao, R. K. H., and Pontes, M. J. C. (2017). Selection of robust variables for transfer of classification models employing the successive projections algorithm. *Anal. Chim. Acta* 984, 76–85. doi: 10.1016/j.aca.2017.07.037

Mishra, P., Asaari, M. S. M., Herrero-Langreo, A., Lohumi, S., Diezma, B., and Scheunders, P. (2017). Close range hyperspectral imaging of plants: A review. *Biosyst. Eng.* 164, 49–67. doi: 10.1016/j.biosystemseng.2017.09.009

Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., and Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *Trends Anal. Chem.* 132, 116045. doi: 10.1016/j.trac.2020.116045

Nikbakhsh, N., Baleghi, Y., and Agahi, H. (2020). A novel approach for unsupervised image segmentation fusion of plant leaves based on G-mutual information. *Mach. Vision Appl.* 32, 5. doi: 10.1007/s00138-020-01130-0

Omia, E., Bae, H., Park, E., Kim, M. S., Baek, I., Kabenge, I., et al. (2023). Remote sensing in field crop monitoring: A comprehensive review of sensor systems, data analyses and recent advances. *Remote Sens.* 15, 354. doi: 10.3390/rs15020354

Paulus, S., and Mahlein, A. K. (2020). Technical workflows for hyperspectral plant image assessment and processing on the greenhouse and laboratory scale. *GigasScience* 9, giaa090. doi: 10.1093/gigascience/giaa090

Peng, J. Y., Song, X. L., Liu, F., Bao, Y. D., and He, Y. (2016). Fast detection of Camellia sinensis growth process and tea quality informations with spectral technology: A review. *Spectrosc. Spect. Anal.* 36, 775–782. doi: 10.3964/j.issn.1000-0593(2016)03-0775-08

Peñuelas, J., Pinol, J., Ogaya, R., and Filella, I. (1997). Estimation of plant water concentration by the reflectance water index WI (R900/R970). *Int. J. Remote Sens.* 18, 2869–2875. doi: 10.1080/014311697217396

Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., et al. (2009). Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.* 113, S110–S122. doi: 10.1016/j.rse.2007.07.028

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., and Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sens. Environ.* 48, 119–126. doi: 10.1016/0034-4257(94)90134-1

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ.* 351, 309.

Safavi-Rizi, V., Uellendahl, K., Ohrlein, B., Safavi-Rizi, H., and Stohr, C. (2021). Cross-stress tolerance: Mild nitrogen (N) deficiency effects on drought stress response of tomato (Solanum lycopersicum L.). *Plant. Environ. Interact.* 2, 217–228. doi: 10.1002/pei3.10060

Sanaeifar, A., Yang, C., Guardia, M. D., Zhang, W. K., Li, X. L., and He, Y. (2023). Proximal hyperspectral sensing of abiotic stresses in plants. *Sci. Total Environ.* 861, 160652. doi: 10.1016/j.scitotenv.2022.160652

Sanaeifar, A., Zhang, W. K., Chen, H. T., Zhang, D. Y., Li, X. L., and He, Y. (2022). Study on effects of airborne Pb pollution on quality indicators and accumulation in tea plants using Vis-NIR spectroscopy coupled with radial basis function neural network. *Ecotox. Environ. Safe.* 229, 113056. doi: 10.1016/j.ecoenv.2021.113056

Saric, R., Nguyen, V. D., Burge, T., Berkowitz, O., Trtílek, M., Whelan, J., et al. (2022). Applications of hyperspectral imaging in plant phenotyping. *Trends Plant Sci.* 27, 301–315. doi: 10.1016/j.tplants.2021.12.003

Schlerf, M., Atzberger, C., and Hill, J. (2005). Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sens. Environ.* 95, 177–194. doi: 10.1016/j.rse.2004.12.016

Shan, J. J., Wang, X., Zhou, H., Han, S. Q., Al Riza, D. F., and Kondo, N. (2018). Variable selection based on clustering analysis for improvement of polyphenols prediction in green tea using synchronous fluorescence spectra. *Method. Appl. Fluoresc.* 6, 025006. doi: 10.1088/2050-6120/aaae0a

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., and Singh, A. (2018). Deep learning for plant stress phenotyping: trends and future perspectives. *Trends Plant Sci.* 23, 883–898. doi: 10.1016/j.tplants.2018.07.004

Slaton, M. R., Raymond Hunt, E. Jr., and Smith, W. K. (2001). Estimating near-infrared leaf reflectance from leaf structural characteristics. *Am. J. Bot.* 88, 278–284. doi: 10.2307/2657019

Sonobe, R., Miura, Y., Sano, T., and Horie, H. (2018a). Monitoring photosynthetic pigments of shade-grown tea from hyperspectral reflectance. *Can. J. Remote Sens.* 44, 104–112. doi: 10.1080/07038992.2018.1461555

Sonobe, R., Sano, T., and Horie, H. (2018b). Using spectral reflectance to estimate leaf chlorophyll content of tea with shading treatments. *Biosyst. Eng.* 175, 168–182. doi: 10.1016/j.biosystemseng.2018.09.018

Steinier, J., Termonia, Y., and Deltour, J. (1972). Smoothing and differentiation of data by simplified least square procedure. *Anal. Chem.* 44, 1906–1909. doi: 10.1021/ac60319a045

Sun, D. W., Robbins, K., Morales, N., Shu, Q. Y., and Cen, H. Y. (2022). Advances in optical phenotyping of cereal crops. *Trends Plant Sci.* 27, 191–208. doi: 10.1016/j.tplants.2021.07.015

Sun, J., Zhou, X., Hu, Y. G., Wu, X. H., Zhang, X. D., and Wang, P. (2019). Visualizing distribution of moisture content in tea leaves using optimization algorithms and NIR hyperspectral imaging. *Comput. Electron. Agric.* 160, 153–159. doi: 10.1016/j.compag.2019.03.004

Sun, Y. J., Zhou, J., and Guo, J. S. (2021). Advances in the knowledge of adaptive mechanisms mediating abiotic stress responses in Camellia sinensis. *Front. Biosci. Landmrk.* 26, 1714–1722. doi: 10.52586/5063

Tang, T., Luo, Q., Yang, L., Gao, C. L., Ling, C. J., and Wu, W. B. (2024). Research review on quality detection of fresh tea leaves based on spectral technology. *Foods* 13, 25. doi: 10.3390/foods13010025

Tang, Z.-Y., Du, Y., Yang, H.-B., Li, X.-H., Yu, Y.-B., and Wang, W.-D. (2023). Changes of endogenous hormone contents and expression analysis of related genes in leaves of tea plants under heat and drought stresses. *J. Tea Sci.* 43, 489–500. doi: 10.3969/j.issn.1000-369X.2023.04.005

Tao, H. Y., Xu, S., Tian, Y. C., Li, Z. F., Ge, Y., Zhang, J. P., et al. (2022). Proximal and remote sensing in plant phenomics: 20 years of progress, challenges, and perspectives. *Plant Commun.* 3, 100344. doi: 10.1016/j.xplc.2022.100344

Tony, M., Samson, K., Charles, M., Paul, M., Richard, M., Mark, W., et al. (2016). Transcriptome-based identification of water-deficit stress responsive genes in the tea plant, Camellia sinensis. *J. Plant Biotechnol.* 43, 302–310. doi: 10.5010/JPB.2016.43.3.302

Tran, C. D. (2001). Development and analytical applications of multispectral imaging techniques: an overview. *Fresenius J. Anal. Chem.* 369, 313–319. doi: 10.1007/s002160000653

Tripathi, S., Kumar, K., Singh, B. K., and Singh, R. P. (2012). Image segmentation: A review. *Int. J. Comput. Sci. Manage. Res.* 1, 838–843. doi: 10.5772/15833

Tu, Y. X., Bian, M., Wan, Y. K., and Fei, T. (2018). Tea cultivar classification and biochemical parameter estimation from hyperspectral imagery obtained by UAV. *Peerj* 6, e4858. doi: 10.7717/peerj.4858

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J., et al. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - A review. *ISPRS J. Photogramm. Remote Sens.* 108, 273–290. doi: 10.1016/j.isprsjprs.2015.05.005

Vogelmann, J. E., Rock, B. N., and Moss, D. M. (1993). Red edge spectral measurements from sugar maple leaves. *Int. J. Remote Sens.* 14, 1563–1575. doi: 10.1080/01431169308953986

Wang, J. H., Li, X., Wang, W. C., Wang, F., Liu, Q. C., and Yan, L. (2023). Research on rapid and low-cost spectral device for the estimation of the quality attributes of tea tree leaves. *Sensors-Basel* 23, 571. doi: 10.3390/s23020571

Wang, Y., Li, Y., Wang, J. H., Xiang, Z., Xi, P. Y., and Zhao, D. G. (2021). Physiological Changes and Differential Gene Expression of Tea Plants (Camellia sinensis (L.) Kuntze var.*niaowangensis* QH Chen) Under Cold Stress. *DNA Cell Biol.* 40, 906–920. doi: 10.1089/dna.2021.0147

Wang, Y. J., Hu, X., Hou, Z. W., Ning, J. M., and Zhang, Z. Z. (2018). Discrimination of nitrogen fertilizer levels of tea plant (*Camellia sinensis*) based on hyperspectral imaging. *J. Sci. Food Agric.* 98, 4659–4664. doi: 10.1002/jsfa.8996

Wang, Y. J., Hu, X., Jin, G., Hou, Z. W., Ning, J. M., and Zhang, Z. Z. (2019). Rapid prediction of chlorophylls and carotenoids content in tea leaves under different levels of nitrogen application based on hyperspectral imaging. *J. Sci. Food Agric.* 99, 1997–2004. doi: 10.1002/jsfa.9399

Wang, Y. J., Jin, G., Li, L. Q., Liu, Y., Kalkhajeh, Y. K., Ning, J. M., et al. (2020a). NIR hyperspectral imaging coupled with chemometrics for nondestructive assessment of phosphorus and potassium contents in tea leaves. *Infrared Phys. Technol.* 108, 103365. doi: 10.1016/j.infrared.2020.103365

Wang, Y. J., Li, T. H., Jin, G., Wei, Y. M., Li, L. Q., Kalkhajeh, Y. K., et al. (2020c). Qualitative and quantitative diagnosis of nitrogen nutrition of tea plants under field condition using hyperspectral imaging coupled with chemometrics. *J. Sci. Food Agric.* 100, 161–167. doi: 10.1002/jsfa.10009

Wang, Y. J., Li, L. Q., Shen, S. S., Liu, Y., Ning, J. M., and Zhang, Z. Z. (2020b). Rapid detection of quality index of postharvest fresh tea leaves using hyperspectral imaging. *J. Sci. Food Agric.* 100, 3803–3811. doi: 10.1002/jsfa.10393

Wang, Y. Z., Lv, H., Deng, R., and Zhuang, S. B. (2020d). A comprehensive survey of optical remote sensing image segmentation methods. *Can. J. Remote Sens.* 46, 501–531. doi: 10.1080/07038992.2020.1805729

Watt, M., Fiorani, F., Usadel, B., Rascher, U., Muller, O., and Schurr, U. (2020). Phenotyping: new windows into the plant for breeders. *Ann. Rev. Plant Biol.* 71, 689–712. doi: 10.1146/annurev-arplant-042916-041124

Wei, Y. Z., Wu, F. Y., Xu, J., Sha, J. J., Zhao, Z. F., He, Y., et al. (2019). Visual detection of the moisture content of tea leaves with hyperspectral imaging technology. *J. Food Eng.* 248, 89–96. doi: 10.1016/j.jfoodeng.2019.01.004

Wijewardane, N. K., Zhang, H. C., Yang, J. L., Schnable, J. C., Schachtman, D. P., and Ge, Y. F. (2023). A leaf-level spectral library to support high-throughput plant phenotyping: predictive accuracy and model transfer. *J. Exp. Bot.* 74, 4050–4062. doi: 10.1093/jxb/erad129

Wocher, M., Berger, K., Danner, M., Mauser, W., and Hank, T. (2018). Physically-based retrieval of canopy equivalent water thickness using hyperspectral data. *Remote Sens.* 10, 1924. doi: 10.3390/rs10121924

Wold, S., Antti, H., Lindgren, F., and Öhman, J. (1998). Orthogonal signal correction of near-infrared spectra. *Chemometr. Intell. Lab.* 44, 175–185. doi: 10.1016/S0169-7439(98)00109-9

Wu, D., and Sun, D. W. (2013). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review - Part I: Fundamentals. *Innovative Food Sci. Emerg. Technol.* 19, 1–14. doi: 10.1016/j.ifset.2013.04.014

Wu, W. B., Li, J. Y., Zhang, Z. B., Ling, C. J., Lin, X. K., and Chang, X. L. (2018). Estimation model of LAI and nitrogen content in tea tree based on hyperspectral image. *Trans. Chin. Soc Agric. Eng.* 34, 195–201. doi: 10.11975/j.issn.1002-6819.2018.03.026

Wu, Y. F., Peng, S. L., Xie, Q., Han, Q. J., Zhang, G. W., and Sun, H. G. (2019). An improved weighted multiplicative scatter correction algorithm with the use of variable selection: Application to near-infrared spectra. *Chemometr. Intell. Lab.* 185, 114–121. doi: 10.1016/j.chemolab.2019.01.005

Xiao, Z. T., Yin, K., Geng, L., Wu, J., Zhang, F., and Liu, Y. B. (2022). Pest identification via hyperspectral image and deep learning. *Signal Image Video P.* 16, 873–880. doi: 10.1007/s11760-021-02029-7

Xu, J. C., Qu, F. F., Shen, B. H., Huang, Z. X., Li, X. L., Weng, H. Y., et al. (2023). Rapid detection of tea polyphenols in fresh tea leaves based on fusion of visible/short-wave and long-wave near infrared spectroscopy and its device development. *Appl. Sci. Basel.* 13, 1739. doi: 10.3390/app13031739

Xu, L. X., Chen, J. W., Ding, G. H., Lu, W., Ding, Y. F., Zhu, Y., et al. (2020). Indoor phenotyping platforms and associated trait measurement: Progress and prospects. *Smart Agric.* 2, 23. doi: 10.12133/j.smartag.2020.2.1.202003-SA002

Xue, J. R., and Su, B. F. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *J. Sens.* 2017, 1353691. doi: 10.1155/2017/1353691

Yamashita, H., Sonobe, R., Hirono, Y., Morita, A., and Ikka, T. (2020). Dissection of hyperspectral reflectance to estimate nitrogen and chlorophyll contents in tea leaves based on machine learning algorithms. *Sci. Rep.* 10, 17360. doi: 10.1038/s41598-020-73745-2

Yan, Y., Zhao, L., Li, J., Zhang, L., and Min, S. (2000). Information technology of modern NIR spectral analysis. *Spectrosc. Spect. Anal.* 20, 777–780. doi: 10.3321/j.issn:1000-0593.2000.06.009

Yang, Z. Y., Albrow-Owen, T., Cai, W. W., and Hasan, T. (2021). Miniaturization of optical spectrometers. *Science* 371, eabe0722. doi: 10.1126/science.abe0722

Yuan, L., Yan, P., Han, W. Y., Huang, Y. B., Wang, B., Zhang, J. C., et al. (2019). Detection of anthracnose in tea plants based on hyperspectral imaging. *Comput. Electron. Agric.* 167, 105039. doi: 10.1016/j.compag.2019.105039

Zahir, S., Omar, A. F., Jamlos, M. F., Azmi, M. A. M., and Muncan, J. (2022). A review of visible and near-infrared (Vis-NIR) spectroscopy application in plant stress detection. *Sens. Actuator A Phys.* 338, 113468. doi: 10.1016/j.sna.2022.113468

Zarco-Tejada, P. J., Berjón, A., López-Lozano, R., Miller, J. R., Martín, P., Cachorro, V., et al. (2005). Assessing vineyard condition with hyperspectral indices:: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sens. Environ.* 99, 271–287. doi: 10.1016/j.rse.2005.09.002

Zhang, C., Zhou, L., Zhao, Y., Zhu, S., Liu, F., and He, Y. (2020). Noise reduction in the spectral domain of hyperspectral images using denoising autoencoder methods. *Chemometr. Intell. Lab.* 203, 104063. doi: 10.1016/j.chemolab.2020.104063

Zhang, J., Yang, R. D., Chen, R., Peng, Y. S., Wen, X. F., and Gao, L. (2018). Accumulation of heavy metals in tea leaves and potential health risk assessment: A case study from Puan County, Guizhou Province, China. *Int. J. Environ. Res. Public Health* 15, 133. doi: 10.3390/ijerph15010133

Zhang, J. C., He, Y. H., Yuan, L., Liu, P., Zhou, X. F., and Huang, Y. B. (2019). Machine learning-based spectral library for crop classification and status monitoring. *Agronomy* 9, 496. doi: 10.3390/agronomy9090496

Zhang, N., Yang, G. J., Pan, Y. C., Yang, X. D., Chen, L. P., and Zhao, C. J. (2020). A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades. *Remote Sens.* 12, 3188. doi: 10.3390/rs12193188

Zhang, P., and Xu, L. H. (2018). Unsupervised segmentation of greenhouse plant images based on statistical method. *Sci. Rep.* 8, 4465. doi: 10.1038/s41598-018-22568-3

Zhang, S. Y., Fei, T., You, X., Wan, Y. K., Wang, Y. J., and Bian, M. (2021). Two hyperspectral indices for detecting cadmium and lead contamination from arice canopy spectrum. *Land Degrad. Dev.* 32, 66–78. doi: 10.1002/ldr.3721

Zhang, X., Xia, W., and Tang, S. (2018). Rapid determination of moisture content in fresh tea leaves based on visible and near infrared spectroscopy. *J. Southwest Minzu Univ. (Nat. Sci. Ed.).* 44, 352–357. doi: 10.11920/xnmdzk.2018.04.004

Zhang, Y. (2017). Detection of the Catechin Monomer in Green Tea Based on Spectroscop. Zhejiang University, Hangzhou (Zhejiang.

Zhang, Y. Y., and Wang, Y. Z. (2023). Machine learning applications for multi-source data of edible crops: A review of current trends and future prospects. *Food Chem. X.* 19, 100860. doi: 10.1016/j.fochx.2023.100860

Zhao, J. W., Wang, K. L., Ouyang, Q., and Chen, Q. S. (2011). Measurement of chlorophyll content and distribution in tea plant's leaf using hyperspectral imaging technique. *Spectrosc. Spect. Anal.* 31, 512–515. doi: 10.3964/j.issn.1000-0593(2011)02-0512-04

Zhao, X. H., Zhang, J. C., Huang, Y. B., Tian, Y. Y., and Yuan, L. (2022a). Detection and discrimination of disease and insect stress of tea plants using hyperspectral imaging combined with wavelet analysis. *Comput. Electron. Agric.* 193, 106717. doi: 10.1016/j.compag.2022.106717

Zhao, X. H., Zhang, J. C., Tang, A. L., Yu, Y. F., Yan, L. J., Chen, D. M., et al. (2022b). The stress detection and segmentation strategy in tea plant at canopy level. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.949054

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 01 August 2024
DOI 10.3389/fpls.2024.1436120

Check for updates

OPEN ACCESS

EDITED BY
Liujun Xiao,
Nanjing Agricultural University, China

REVIEWED BY
Yahui Guo,
Central China Normal University, China
Zhenwang Li,
Yangzhou University, China

*CORRESPONDENCE
Changying Li
cli2@ufl.edu
Kyle Johnsen
kjohnsen@uga.edu

RECEIVED 21 May 2024
ACCEPTED 04 July 2024
PUBLISHED 01 August 2024

CITATION
Rodriguez-Sanchez J, Snider JL, Johnsen K and Li C (2024) Cotton morphological traits tracking through spatiotemporal registration of terrestrial laser scanning time-series data. *Front. Plant Sci.* 15:1436120. doi: 10.3389/fpls.2024.1436120

COPYRIGHT
© 2024 Rodriguez-Sanchez, Snider, Johnsen and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Cotton morphological traits tracking through spatiotemporal registration of terrestrial laser scanning time-series data

Javier Rodriguez-Sanchez[1], John L. Snider[2], Kyle Johnsen[1]* and Changying Li[3]*

[1]School of Electrical and Computer Engineering, University of Georgia, Athens, GA, United States, [2]Department of Crop and Soil Sciences, University of Georgia, Tifton, GA, United States, [3]Bio-Sensing, Automation and Intelligence Laboratory, Department of Agricultural and Biological Engineering, University of Florida, Gainesville, FL, United States

Understanding the complex interactions between genotype-environment dynamics is fundamental for optimizing crop improvement. However, traditional phenotyping methods limit assessments to the end of the growing season, restricting continuous crop monitoring. To address this limitation, we developed a methodology for spatiotemporal registration of time-series 3D point cloud data, enabling field phenotyping over time for accurate crop growth tracking. Leveraging multi-scan terrestrial laser scanning (TLS), we captured high-resolution 3D LiDAR data in a cotton breeding field across various stages of the growing season to generate four-dimensional (4D) crop models, seamlessly integrating spatial and temporal dimensions. Our registration procedure involved an initial pairwise terrain-based matching for rough alignment, followed by a bird's-eye view adjustment for fine registration. Point clouds collected throughout nine sessions across the growing season were successfully registered both spatially and temporally, with average registration errors of approximately 3 cm. We used the generated 4D models to monitor canopy height (CH) and volume (CV) for eleven cotton genotypes over two months. The consistent height reference established via our spatiotemporal registration process enabled precise estimations of CH (R^2 = 0.95, RMSE = 7.6 cm). Additionally, we analyzed the relationship between CV and the interception of photosynthetically active radiation ($IPAR_f$), finding that it followed a curve with exponential saturation, consistent with theoretical models, with a standard error of regression (SER) of 11%. In addition, we compared mathematical models from the Richards family of sigmoid curves for crop growth modeling, finding that the logistic model effectively captured CH and CV evolution, aiding in identifying significant genotype differences. Our novel TLS-based digital phenotyping methodology enhances precision and efficiency in field phenotyping over time, advancing plant phenomics and empowering efficient decision-making for crop improvement efforts.

KEYWORDS

terrestrial laser scanning, crop morphological traits tracking, time-series field phenotyping, Spatiotemporal registration, 4D imaging, LiDAR, remote sensing

1 Introduction

The increasing demand for agricultural products has emphasized the need for more efficient methods to accelerate crop productivity (Voss-Fels et al., 2019). Field-based plant phenotyping, key for evaluating plant traits in distinct environmental conditions, has become integral in crop breeding (Großkinsky et al., 2015). However, conventional techniques are laborious and time-consuming, often restricting assessments of traits to the end of the growing season and creating bottlenecks in breeding programs (Furbank and Tester, 2011). Recent advancements in remote and proximal sensing have increased the throughput and precision of field-based phenotyping (Singh et al., 2016). Nevertheless, persistent challenges in data processing and automation impede further progress in plant phenomics (Chawade et al., 2019; Kim, 2020). Therefore, novel strategies are imperative to streamline procedures and enable enhanced trait monitoring for crop growth tracking.

Integrating temporal data analysis into phenotyping pipelines can unveil plant development patterns and cyclic phenomena in growth. Continuous monitoring of plant growth is key for understanding plant behavior and responses to external factors (Miao et al., 2020). Recent studies have demonstrated the potential of continuous plant growth monitoring to identify specific genotypes contributing to particular traits (Xavier et al., 2017; Wang et al., 2018; Zhou et al., 2020; Li et al., 2022). However, the variability in crop growth within fields makes it challenging to establish standardized and consistent monitoring systems, hindering uniform trait tracking over time for large-scale fields (Tao et al., 2022). Exploring innovative and adaptable solutions for in-field trait assessment is key to overcoming these challenges and improving field phenotyping efficiency.

Advancements in three-dimensional (3D) imaging have expanded the application of plant modeling for the purpose of predicting crop traits over time. Detailed plant 3D models obtained from imaging data through photogrammetry methods (Paproki et al., 2012; Gelard et al., 2018; Hui et al., 2018), structured light scanning (Li et al., 2013), and active laser triangulation (Paulus et al., 2014) have been applied to track individual organs and monitor canopy growth of potted plants. However, these methods have proved labor-intensive and may lack scalability for larger field experiments. At the field scale, photogrammetry methods have also been proposed to gather 3D information from the crop over time using 2D imagery from ground (Carlone et al., 2015; Dong et al., 2017) or aerial systems (Chebrolu et al., 2018; Malambo et al., 2018). Despite their potential for large field coverage, these methods face constraints in capturing complex canopies due to RGB sensor limitations in handling occlusions, repetitive patterns, and changing light conditions.

In recent years, advanced methods using light detection and ranging (LiDAR) technologies for crop phenotyping have gained prominence. LiDAR scanners can mitigate some of the limitations associated with traditional 2D image sensors (Lin, 2015). Notably, employing a linear LiDAR scanning approach, mobile laser scanning (MLS) techniques have found application in field crop monitoring using moving ground vehicles (Jiménez-Berni et al., 2018; Sun et al., 2018) or specialized rail-based phenotyping platforms (Li et al., 2023). However, these methodologies involve analyzing crops from a fixed top-to-bottom orientation, potentially limiting the capture of lower canopy layers in dense crops with complex architectures, as certain parts of the crop may be shadowed or obscured from the sensor's perspective.

Terrestrial laser scanning (TLS) techniques based on stand-alone 3D LiDAR scanners offer reliable solutions for overcoming occlusion issues and simplifying field-based crop phenotyping. Those scanners have demonstrated effectiveness in monitoring seasonal crop changes in small-scale breeding fields (Hosoi and Omasa, 2009, 2012) as well as daily canopy alterations in both individual plants (Herrero-Huerta et al., 2018) and large field trials (Jin et al., 2021a). Their consistent performance under various lighting conditions and superior depth-sensing capabilities make them particularly well-suited for precise crop canopy measurements in field-based plant monitoring (Madec et al., 2017; Jin et al., 2021b). However, the adoption of TLS techniques for plant phenotyping may be hindered by the absence of standardized and automated data processing methodologies (Medic et al., 2023), making continuous monitoring challenging. Simplifying LiDAR data processing and extraction of information can enhance plant trait analysis over time and promote broader usage in field phenotyping.

To ensure consistent crop growth modeling and plant traits monitoring using TLS, a fundamental step is to precisely associate spatial information from data collected over time into the same geospatial context. However, the complex transformations that crops undergo during the growing season complicate the registration of point clouds collected at different time points. The most straightforward technique for TLS-based crop growth tracking involves scanning the field from a fixed location at different points in time (Eitel et al., 2016). This single-position scanning approach introduces challenges such as laser shadows obscuring parts of the crop and variations in point density, particularly for plants closer to the laser scanner, that can impact subsequent data analysis (Malambo et al., 2019). An alternative approach involves scanning the field from multiple locations (e.g., multi-scan TLS). A study successfully employed this method to track wheat height over the growing season (Guo et al., 2019). Despite its advantages in reducing occlusions, this approach relied on precise geolocation of point cloud data, requiring accurate positioning systems that increase costs and processing time for data analysis (Crommelinck and Höfle, 2016). Another proposed method involves using a motorized gantry-type phenotyping platform to mount the LiDAR scanner and collect TLS data (Jin et al., 2021a). While effective for accurate trait monitoring, this approach requires a dedicated structure and highly accurate position sensors for real-time scanner location, introducing complexities to the phenotyping process.

To further enhance the precision and consistency of TLS-based time-series field phenotyping, the alignment of point clouds collected over time into a common coordinate system—4D registration— becomes fundamental. Traditionally, registering point clouds for crop growth monitoring has relied on the use of registration targets. This approach involves installing registration targets throughout the field to assist in aligning successive point clouds and has proven effective in monitoring crop growth in large

breeding fields (Tilly et al., 2014; Friedli et al., 2016), as well as individual plants evolution under field conditions (Su et al., 2019). However, it can be labor-intensive and error-prone. Deviations in the target placement between surveys can impact the temporal alignment accuracy, and hence the estimation of traits over time. Ensuring that the registration targets are placed consistently in the same exact location in agricultural fields, where machinery needs to operate or other experiments need to be executed, can be challenging. More adaptable approaches used crop distribution to improve the alignment of remotely sensed data (Chebrolu et al., 2018; Günder et al., 2022). Although these studies were limited to aerial imagery, the concept could enhance spatiotemporal alignment of TLS data, facilitating its application for field phenotyping.

Alternative methods for 4D registration of point clouds have also been implemented to accommodate changes in plant structures, albeit with limitations for large-scale fields. Recent studies have proposed non-rigid registration methods to align point clouds at the organ-level during the growing season (Chebrolu et al., 2020; Magistri et al., 2020; Chebrolu et al., 2021). In contrast to rigid-body transformations (Besl and McKay, 1992), which assume a fixed relationship between the point clouds, non-rigid registration allows for more flexible and adaptive alignment, enabling more accurate tracking of plant growth and structural changes over time. However, these approaches require highly-detailed plant models, which can hinder their practicality for field crops where segregating individual plants might not be feasible. The development of methodologies for spatiotemporal registration of large-scale LiDAR data can significantly broaden the application of these technologies in field phenotyping, enabling efficient crop growth modeling and trait monitoring over time for enhanced crop development.

In this study, we present a novel methodology for automating the registration of field-based time-series TLS point clouds, addressing critical challenges in data registration and processing to facilitate consistent crop phenotyping and growth tracking. Our primary technical contribution lies in the development of a two-phase TLS data registration approach, which exploits terrain morphology and crop row distribution to minimize alignment errors between point clouds collected over time. This innovative method reduces dependence on fixed registration targets and streamlines the data processing pipeline, offering breeders a robust solution to acquire accurate phenotypic traits from the same physical locations at different time points. From a crop science perspective, our work introduces an efficient approach to track the evolution of crop traits throughout the season, enhancing our understanding of cotton plant development and trait variations. The specific objectives of the study were to: (1) develop a robust methodology for precise spatiotemporal registration of field-acquired point clouds; (2) validate the effectiveness of the methodology in reducing point cloud alignment errors throughout the growing season; (3) investigate the utility of registered time-series TLS data for extracting key morphological traits and facilitating precise monitoring of cotton crop growth dynamics; and (4) assess the applicability of the developed methodology in informing crop modeling efforts, with a focus on its potential to deepen our understanding of cotton plant growth and genotype-specific trait variations.

2 Materials and methods

2.1 Experimental field

The field research site used in this study was located at the University of Georgia's Iron Horse research farm in Greene County, GA, USA (33°43'01.6"N 83°18'1.8"W). The field contained 11 different genotypes, arranged in a randomized complete block design with 8 replications. These genotypes included conventional Upland cotton varieties from public breeding programs (Acala Maxxa, DES 56, Tamcot Sphinx, UA 48), exotic genotypes (T0018MDN, T0246BC3MDN, and MDN0101 (GH191)), commercial Upland cultivars (DG 3615, DP 1646, and ST5020), and a Pima cotton cultivar (DP 314). A more detailed description of the included cultivars can be found in (Kaur et al., 2023).

The field's layout consisted of single-row plots, each measuring 3.05 meters in length, with an inter-row spacing of 1.83 meters (Figure 1A). A 1.52-meter wide bare soil alley separated each range of plots. The field was organized into 88 plots, distributed across 8 rows, with 11 plots per row, resulting in approximate dimensions of 52 meters in length and 14 meters in width. A total of 15 seeds were sowed in each plot on June 18, 2021 (planting date), and the final plant density in each plot varied based on germination and survival rates.

2.2 Field data collection

A FARO Focus 3D S70 laser scanner (FARO Technologies, Florida, US) was used for terrestrial laser scanning (TLS). To reduce occlusions and collect as much information from the crop as possible, a multi-scan methodology was adopted to scan the field. Between nine and thirteen scan locations, depending on the growth status of the crop, were used to ensure an adequate coverage of the crop during the different growth stages. The LiDAR was mounted on an elevating tripod (Figure 1B), whose height was adjusted *in situ* during each data collection session in accordance with crop height. At the beginning of the season, the height of the scanner was configured to 1.25 m. When the plants reached the canopy closure stage, the scanner height was adjusted to 1.8 m. The quality parameter for the LiDAR scanner was set to 2x, while the angular resolution was set to 1/2 (angular step of 0.18°) for both the vertical and horizontal angles, which is equivalent to a point distance of 3.05 mm over 10 m. The collected point clouds were colorized using information gathered from the color camera integrated in the LiDAR scanner. Additionally, global positioning system (GPS) and inclinometer information from the scanner's internal sensors were also saved during the scan. For each scan location, raw point cloud data were stored in an SD card as FARO Laser Scan (.FLS) files for further processing.

Nine data collection sessions were conducted to monitor canopy development from crop establishment to canopy closure. The LiDAR scanner was used to survey the cotton field approximately at 35, 42, 49, 56, 62, 70, 77, 84, and 98 days after planting (DAP). During each session, five spherical targets with a diameter of 200 mm (Koppa Target Spheres, California, USA) were strategically placed throughout the field

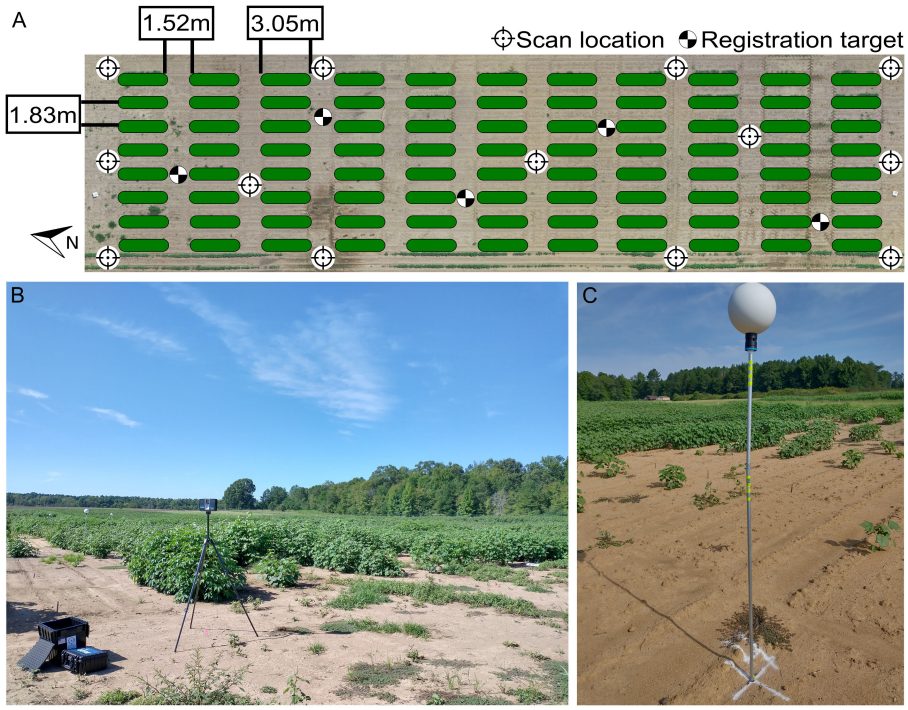


FIGURE 1
Experimental field layout. **(A)** Distribution of plots (green ovals), scan locations (circular target icons), and registration targets (checkered circles) for LiDAR-based phenotyping. **(B)** LiDAR scanner performing TLS scan at a scan location. **(C)** Registration target deployed in the field.

to facilitate the spatial co-registration of point clouds from different scan locations into a common coordinate system. These targets were mounted on aluminum rods with heights ranging from 1.4 m to 2 m (Figure 1C). The spherical targets were deployed before each TLS survey session. A mallet was used to drive the aluminum rods into the ground, ensuring each one of them went in straight and deep enough to provide stability during the survey. Since permanent structures in the field were not feasible due to ongoing experiments, the aluminum rods were taken down and stored between data collection sessions. Therefore, the position of the spheres was not necessarily consistent across different data collection sessions.

Manual measurements of key plant traits were conducted directly in the field at three different stages of crop development and served as ground truth data for validation purposes. The first session took place immediately after the onset of the blooming stage, approximately at 60 DAP. The second ground truth session occurred during the peak blooming period, at around 80 DAP. The final session was performed at the canopy closure stage, 90 just before the first open cotton bolls became visible. The field measurements included the height of the canopy and the quantity of light intercepted by the canopy, which is directly proportional to crop growth (Baker and Meyer, 1966) and plays a key role in modeling crop evolution and yield (Loomis and Williams, 1969). Canopy height (CH) was determined by measuring the distance between the ground and the top of the canopy (the plant terminal) using a measuring tape. Two to five different plants within each plot were measured following established methods commonly used in cotton breeding. The final CH value was calculated by averaging these measurements. Light interception was estimated by the fraction of photosynthetically active radiation (PAR) intercepted by the canopy ($IPAR_f$), using Equation 1. Two $IPAR_f$ readings were taken per plot and then averaged. Light interception measurements were taken using an AccuPAR LP-80 ceptometer (METER Environment, Pullman, WA) under cloudless conditions between 1100 and 1300 h. Both the below-canopy photosynthetically active radiation (PAR_{below}) and the above-canopy irradiance (PAR_{above}) were measured simultaneously.

$$IPAR_f = \frac{PAR_{above} - PAR_{below}}{PAR_{above}} \quad (1)$$

2.3 Data processing pipeline for spatiotemporal alignment

Our TLS-based 4D field phenotyping methodology involved a sequential series of data processing steps (Figure 2). Initially, intra-session data processing helped co-register multi-scan point clouds into a common coordinate system and prepared the data collected in each data collection for subsequent analysis. Inter-session processing aligned consecutive point clouds in both space and time, facilitating the extraction of crop traits to analyze growth trends over the season.

The data processing pipeline was designed to run on Windows systems, using only CPU resources without requiring specialized GPUs. All experiments were conducted on a desktop computer

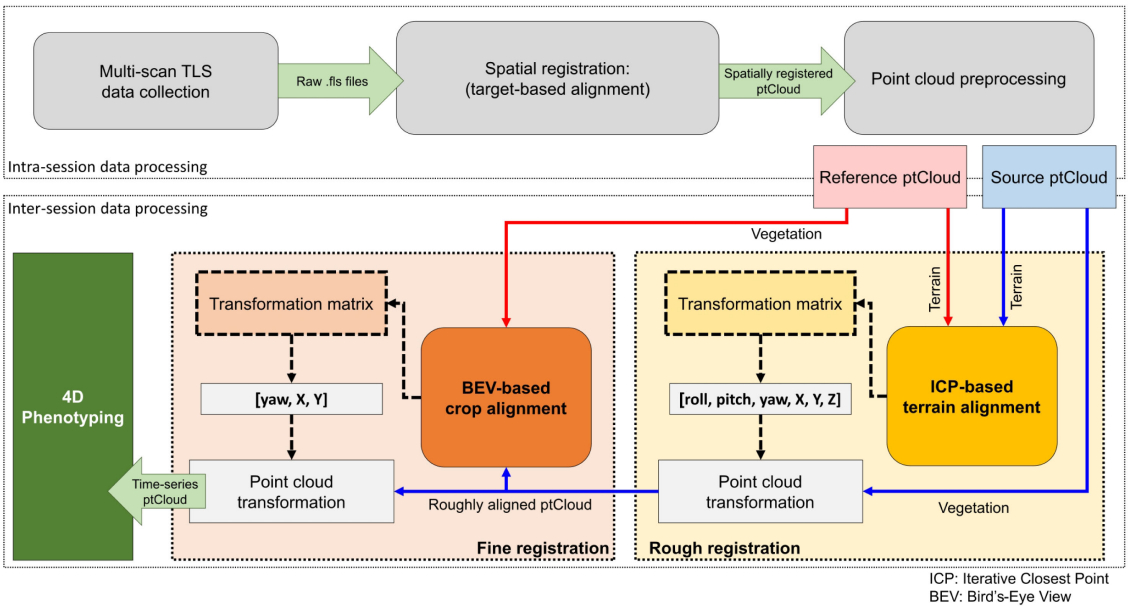


FIGURE 2
Overview of the proposed methodology for LiDAR-based 4D in-field phenotyping.

equipped with an 8-Core Intel(R) Core(TM) i7–9700K CPU running at 3.60 GHz and 64 GB of RAM.

2.3.1 Intra-session data processing

After each data collection session, point clouds from multiple scan locations were co-registered using SCENE software (FARO Technologies, Florida, US), version 2019.2. The raw. FLS files were imported into SCENE and preprocessed using the *'Edge artifact'* filter. Point clouds were automatically registered using the *'Target Based'* method. Registration results were validated through target-based and mean point error statistics. Then, SCENE's *'Clipping box'* tool was used to isolate and extract 3D points within designated field boundaries. Points within the clipping box were saved in the LASer (.LAS) file format.

Co-registered point clouds underwent denoising using a statistical outlier removal (SOR) filter with parameters $N = 20$ and ±2.5 standard deviations as outliers boundaries. Then, a subsampling step based on a 5 mm point-to-point distance threshold reduced the point clouds size while maintaining spatial information, reducing computational demands. After subsampling, point cloud height was normalized with respect to a local digital terrain model (DTM) ensured consistency in elevation data across the field. For a more detailed description, please refer to the Supplementary Materials.

2.3.2 Inter-session data processing

The terrain's elevation profiles and slopes are generally stable in the short term, and the arrangement of crop rows remains consistent despite canopy growth. Leveraging this stability, we introduced a two-phase alignment process for temporally registering successive point clouds: one as the fixed reference and the other as the source undergoing alignment.

In the initial phase, we conducted a preliminary alignment using the reconstructed terrain models to establish a rough correspondence between pairs of point clouds. We used the iterative closest point (ICP) pairwise matching algorithm (Besl and McKay, 1992) to align the DTM points. The ICP algorithm seeks to find the best-fitting transformation that minimizes the distance between points in two point clouds, in our case the terrain points. This iterative process continues until a satisfactory alignment with minimum root mean squared deviation (RMS) is achieved. The resulting transformation matrix, including full rotation and translation components (i.e., roll, pitch, yaw, X, Y, and Z), was applied to the source point cloud, providing an initial spatial alignment.

To further refine the alignment and minimize errors, we implemented a second phase based on bird's-eye view (BEV) alignment (Figure 3). This approach used the distribution of crop rows and plots, automatically identified from the point clouds, to match the orientation and position of both point clouds from a top-down perspective. This refinement step ensured greater accuracy in aligning the point clouds collected over time.

The BEV alignment process started with the identification of plots centroids in both point clouds. A height threshold, set at half the maximum height value within each plot, was applied to separate points belonging to the upper canopy section. After thresholding, the canopy points were projected onto the X-Y plane, resembling a 2D top-down orthographic view of the field. Using contour detection, points corresponding to each plot were clustered and filtered using Principal Components Analysis (PCA) to eliminate incomplete plots or noisy point clusters. We leveraged the typical growth pattern observed in field crops planted in plots, where the length along the row axis generally exceeds the width. Any clusters deviating from this expected growth pattern were identified as noise

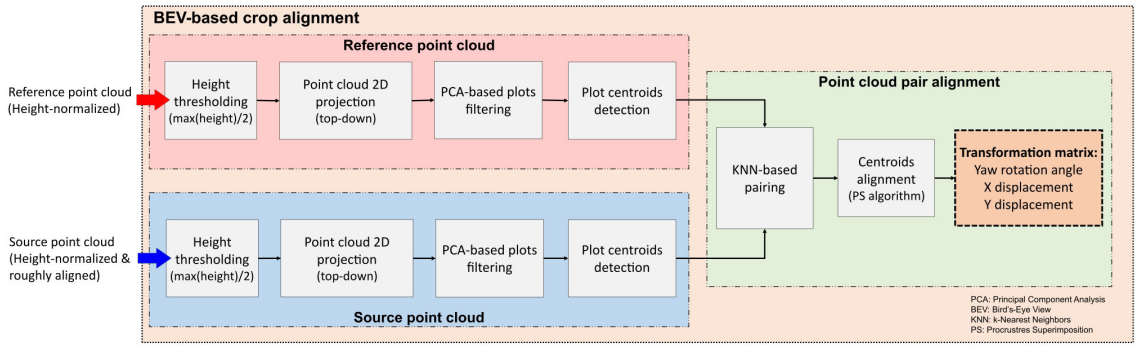


FIGURE 3
Bird's-eye view alignment steps for multi-temporal point cloud pairs registration.

and excluded from subsequent processing stages. The pixel coordinates corresponding to the centroid of each validated cluster were retained for further analysis.

The detected centroids served to improve the alignment between the point clouds. However, certain plots may not appear in the 2D projection in the XY plane during specific growth stages. To address this issue, a k-Nearest Neighbor (KNN) algorithm (Fix and Hodges, 1951) identified pairs of points common to both point clouds, excluding any missing centroids from the BEV alignment process. The rigid transformation between the source and reference point clouds was formulated as a Procrustes superimposition (PS) problem (Rohlf and Slice, 1990). This method involves determining the transformation needed to optimally align two sets of points, effectively overlaying one onto the other. Using a custom implementation of Sneath's method (Sneath, 1967) (Algorithm 1), we calculated the translation and rotation required to align the centroids from the source point cloud with those from the reference point cloud.

Given a reference set R of m pairs of coordinates (x, y) representing the plot centroids detected in the reference point cloud, and a source set S of n pairs of coordinates (x, y) representing the plot centroids detected in the source point cloud, the PS algorithm returns the optimal translation and rotation angle that minimize the sum of the squared distances between corresponding points. First, k pairs of points

Input:

$R = [(x_1, y_1), ..., (x_j, y_j)]$; $j = 1, 2, ..., m$

$S = [(x_1, y_1), ..., (x_l, y_l)]$; $l = 1, 2, ..., n$

Output:

Optimal translation T

Optimal rotation angle θ

Initialization:

1: $R' = [\,]$
2: $S' = [\,]$
3: $T = (0, 0)$
4: $\theta = 0$
5: $[R', S'] = KNN(R, S)$
6: $(\bar{x}_{R'}, \bar{y}_{R'}) = (\frac{\sum_{i=1}^{k} x_{R'i}}{k}, \frac{\sum_{i=1}^{k} y_{R'i}}{k})$
7: $(\bar{x}_{S'}, \bar{y}_{S'}) = (\frac{\sum_{i=1}^{k} x_{S'i}}{k}, \frac{\sum_{i=1}^{k} y_{S'i}}{k})$
8: $T \leftarrow (\bar{x}_{R'} - \bar{x}_{S'}, \bar{y}_{R'} - \bar{y}_{S'})(T_x, T_y)$
9: **for** $i = 1, 2, ..., k$ **do**
10: $R'(k) \leftarrow (x_{R'i} - \bar{x}_{R'}, y_{R'i} - \bar{y}_{R'})$
11: $S'(k) \leftarrow (x_{S'i} - \bar{x}_{S'}, y_{S'i} - \bar{y}_{S'})$
12: **end for**

13: $\theta = \arctan\left(-\frac{\sum_{i=1}^{k}(x_{S'i} \cdot y_{R'i} - y_{S'i} \cdot x_{R'i})}{\sum_{i=1}^{k}(x_{S'i} \cdot x_{R'i} - y_{S'i} \cdot y_{R'i})}\right)$
14: **return** T, θ

Algorithm 1. Procrustes superimposition algorithm for plot centroids alignment.

common to both point clouds were identified using the KNN algorithm (Algorithm 1, line 5). Then, the center of gravity (i.e., mean or average point) for these matched points was computed for each point cloud (Algorithm 1, lines 6 and 7). The distance between centers of gravity coordinates in the X and Y axes was used to compute the components of the optimal translation (T_x, T_y) (Algorithm 1, line 8). Then, both sets of points were translated to the origin (Algorithm 1, lines 9 to 12). Finally, the rotation angle was calculate using the sums of the cross products between x and y components for both sets to minimize the distances between corresponding points (Algorithm 1, line 13).

After fine registration, all point clouds were aligned under a common coordinate system. It is important to note that the initial normalization conducted in the previous section was performed locally for each data collection session. Discrepancies in DEM quality across different growth stages may introduce height errors

between sessions, potentially impacting crop growth analysis. To mitigate these errors and minimize bias in trait estimation, height values for every data collection were renormalized using a common reference plane across all collected point clouds. The terrain mesh derived from the reference point cloud served as the global reference ground level (Z = 0 meters) for all subsequent point clouds. This approach ensured that traits measured relative to each plot's unique characteristics had a consistent reference, enabling an unbiased assessment of crop growth over time.

2.3.3 Performance analysis for spatiotemporal registration

To evaluate our methodology's effectiveness in aligning point clouds over time, we needed accurately aligned reference models for each data collection session. We achieved this by manually aligning each point cloud dataset with the baseline reference, which was the initial point cloud collected at 35 DAP. Working in pairs and using the CloudCompare tool *'Align (point pairs picking)'*, we selected 12 key points common to both point clouds. We then refined the alignment using the *'Translate/Rotate'* tool until we achieved a root mean square (RMS) error of approximately 3 mm. The resulting transformation matrix served as the basis for comparing the temporal registration performance of our method.

To measure the performance of our registration process, we compared each point cloud spatiotemporally registered by our data processing pipeline with its manually aligned counterpart from each data collection session. The Hausdorff distance (Rote, 1991) was considered as performance metric to measure the dissimilarity between both point sets. Given two different point sets $PC_A = a_1, a_2, \ldots, a_n$ and $PC_B = b_1, b_2, \ldots, b_n$, the Hausdorff distance from PC_A to PC_B can be computed using Equation 2:

$$d_H(PC_A, PC_B) = \max_{a \in PC_A} \min_{b \in PC_B} (d(a, b)) \tag{2}$$

where a and b are points belonging to the point clouds PC_A and PC_B respectively, and $d(a, b)$ is the Euclidean distance between a and b.

2.4 Crop traits estimation

2.4.1 Individual-plot point clouds preparation

To isolate individual plots within the registered point clouds, we initially generated a polygon grid covering the entire field area and saved it using the ESRI shapefile (.shp) spatial data format. A region of interest (ROI) was manually defined for the first plot, then replicated and uniformly spaced to create a grid pattern aligning with crop rows and plots. The shapefile grid was used as the spatial guide to segment individual plots from each data collection session. As the point clouds were spatially and temporally co-registered with the reference point cloud from the first session, this grid generation was a one-time task applied throughout the growing season

2.4.2 Plot-level traits extraction

After segmenting individual plots, key morphological traits, including canopy height (CH) and canopy volume (CV), were extracted using the *'laspy'* and *OpenCV* libraries in Python. In addition, we evaluated the use of CV estimations as a proxy for estimating light interception under field conditions, which has been identified as a key input for process-based growth and yield models in cotton (Ermanis et al., 2020; Pokhrel et al., 2023). This indirect approach for estimating light interception has been explored previously in almond orchards (Zhang et al., 2021).

To estimate CH at the individual plot level, we analyzed normalized height values within each plot point cloud. We explored two percentile values, 95^{th} (CH95) and 99^{th} (CH99), as well as the maximum height (CHmax) derived from the histogram of Z-coordinates. We also compared CH estimations post-4D registration and those from individual point clouds before temporal registration.

For CV estimation, we initially calculated the per-plot projected canopy area (CA). Vegetation points were differentiated from terrain points based on a 5 cm threshold for Z-values. The identified vegetation points were then projected onto the XY plane, generating a 2D binary image mask for each plot. CA was determined by counting the number of pixels per unit area within the projected vegetation points. Subsequently, CV was estimated by multiplying the projected area by the corresponding CH values using Equation 3:

$$CV_{projArea} = CA \cdot CH = numPixels \cdot \frac{shapeSize_m}{shapeSize_{px}} \cdot CH \tag{3}$$

where $shapeSize_m$ represents the shapefile area for the plot in m^2, and $shapeSize_{px}$ represents the area in pixels.

2.4.3 Crop trait estimation performance metrics

The overall performance of our methodology at estimating crop traits for field phenotyping was assessed by comparing trait values estimated from TLS data with ground truth values. The selection of the optimal height percentile for CH estimations was based in three metrics: the coefficient of determination (R^2), the root mean squared error (RMSE), and the mean absolute percentage error (MAPE), as defined in Equations 4–6, respectively. Higher values of R^2 approaching 1 and lower RMSE and MAPE values indicated more accurate estimations.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

$$MAPE(\%) = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100 \tag{6}$$

where n is the total number of data points used for regression analysis, y_i is the actual value of CH for the i^{th} ground truth plot, $\hat{y}_i$

is the predicted CH value obtained from the LiDAR point cloud associated with the i^{th} ground truth plot, and $\bar{y}$ is the mean CH value calculated from the total of ground truth measurements.

To assess the use of CV estimations as a proxy for estimating canopy light interception, a nonlinear regression analysis was performed. This analysis compared the estimated CV values with field ground truth measurements of $IPAR_f$, using nonlinear least-squares fitting with the *'curve fit'* function from the *'scipy'* Python library. The standard error of regression (*SER*) was calculated using Equation 7 to evaluate the goodness of fit, accounting for the number of independent variables.

$$SER = \sqrt{\frac{1}{n-(p+1)} \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (7)$$

Here, y_i and $\hat{y}_i$ are the actual and predicted values of $IPAR_f$ for the i^{th} ground truth plot, respectively, n is the total number of data points used for nonlinear regression, and p is the number of coefficients in the model not counting the intercept.

2.4.4 Growth modeling and parameters estimation

The growth cycle of cotton plants follows an S-shaped curve, with lower and upper bounds indicating the period of growth equilibrium (Snider et al., 2021). Among the mathematical models for growth modeling, the three most commonly used for asymptotic growth are the logistic model, the Gompertz model, and the Richards growth model (Tjørve and Tjørve, 2010) (Supplementary Table S1). These models capture plant growth trajectories with different levels of asymmetry and flexibility. We defined W as the trait value at a given time t, A as the maximum value for the trait (upper asymptote), k as the growth rate, and T_i as the time at inflection when the maximum growth rate occurs.

The three sigmoid growth models were used to analyze the estimated trait data and reveal the underlying dynamics of CH and CV evolution over time. Nonlinear mixed-effect models (NLME) were employed to capture trait evolution, considering *time* and *genotype* as fixed effects, while *replicate* was treated as a random effect. For the model fitting process, we used the statistical computing and graphics software R (R Core Team, 2023), version 4.2.3, and the *'nlme'* package (Pinheiro and Bates, 2000). For the sake of simplicity and to ensure convergence, we assumed that the growth rate parameter k remained constant and independent of genotype, without showing any random variability across replicates.

The likelihood ratio test (*'anova.lme'*) from the *'nlme'* package was used to identify the most suitable model based on the Akaike information criterion (AIC) (Akaike, 1992) and the Bayesian information criterion (BIC) (Schwarz, 1978) metrics. The model with the highest AIC and BIC support was fitted to the trait data, estimating growth curve parameters. A *post-hoc* analysis was conducted to test for mean differences in growth parameters by genotype using the *'emmeans'* package (Lenth, 2023), with Bonferroni corrections for multiple comparisons. This comprehensive approach enabled us to discern significant variations in growth characteristics across different genotypes.

3 Results

3.1 Development of the 4D field model for validation

We successfully reconstructed detailed point clouds for each data collection session using multi-scan TLS approaches and automatic registration (Figure 4). The average mean point distance between matched target pairs was 1.6 mm, with a maximum error of 3.2 mm across all sessions. On the whole, the average mean scan point error remained consistently below 6 mm, peaking at 8.5 mm for data collected at 84 DAP. After point cloud preprocessing, the point cloud density was reduced by 8 to 18 times (Supplementary Table S2), significantly decreasing LAS file sizes to approximately 0.5 GB each and facilitating subsequent processing. This reduction also enabled simultaneous processing of all point clouds in a single CloudCompare session.

The presence of consistent and distinguishable 3D features across all collected point clouds allowed for the creation of an accurate 4D field model (Figure 5). This model integrated individually reconstructed 3D point clouds from each session, aligning them to the reference point cloud under a unified coordinate system. As a result, we obtained almost perfectly aligned point clouds for each session that served as a crucial reference for validating the accuracy of our registration process. By comparing the individual point clouds after alignment using our methodology to the reference model, we were able to assess the performance of our registration method.

3.2 Terrain-based registration of point clouds

The rasterization of terrain points and the posterior meshing process allowed us to obtain consistent DTM models for each data collection session during the growing season (Figure 6). However, the reconstruction process tended to overestimate terrain elevation in areas with excessive vegetative growth and denser canopies. As the growing season progressed and plant canopies began to overlap with neighboring plots, the number of terrain points collected by the LiDAR scanner notably decreased, leading to data gaps in the rasterized point clouds (Figures 6B–D). These gaps presented a challenge in accurately modeling the terrain. Notably, when comparing the initial terrain model from the first data collection session (Figure 6A) to the final session (Figure 6D), disparities in elevation reached nearly 37 cm in regions with dense vegetation. Despite these challenges, our method consistently produced accurate terrain morphology results up to 84 DAP when all plots in the field effectively reached the canopy closure stage.

Leveraging internal sensors data gathered during LiDAR data collection allowed us to position consecutive point clouds in close proximity to each other and facilitated convergence of the ICP algorithm, minimizing its risk of getting trapped in local minima (Supplementary Figure S2). The resulting RMS values for ICP

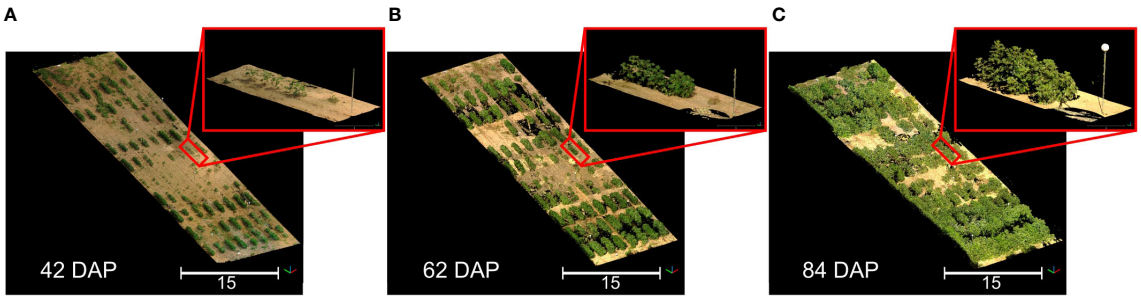


FIGURE 4
3D representation of the field at three different growth stages. **(A)** Data collected 42 days after planting (DAP); **(B)** Data collected 62 DAP; **(C)** Data collected 84 DAP. Point clouds were colorized using information from the sensor's camera. Black areas represent spaces without 3D point information. Insets show a close-up of the same plot in each point cloud.

registration between the DTM for each data collection session and the reference DTM were consistently less than 10 mm.

3.3 Bird's-eye view-based alignment for spatiotemporal registration refinement

The developed process enabled the isolation of prominent plots in the field, allowing for an accurate identification of their centroids (Figure 7). Through local height normalization and thresholding, only the upper section of the canopy was retained (Figure 7A), revealing the distinct pattern of crop rows by focusing solely on vegetation points (Figure 7B). Following plot clustering, PCA-based filtering reduced the potential impact of spurious points or poorly defined plots, such as noise, small plots, or tall weeds, thereby enhancing the consistency and robustness of plot centroid identification (Figures 7C, D). For a comprehensive visualization of the BEV alignment process for two point clouds, refer to Supplementary Figure S3.

The BEV alignment step consistently enhanced the accuracy of point cloud registration for data collected during the vegetative crop growth stage (Figure 8). Initially, errors ranging between 10 cm and 35 cm with respect to the reference point cloud were observed after the first alignment step (Figure 8A, inset 1). However, after the second alignment step, errors for points above the terrain plane notably decreased to values around 2 to 3 centimeters (Figure 8B, inset 1). Examination of alignment results at terrain level revealed that registration errors in the Z direction remained relatively constant, increasing slightly from 2.2 cm for the initial DTM-

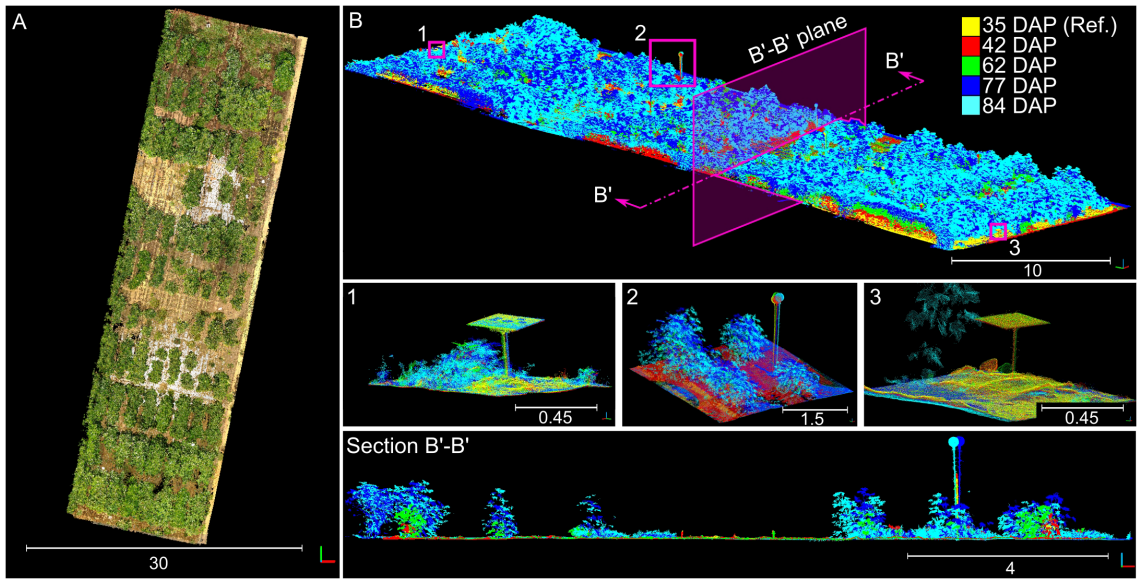


FIGURE 5
Manually aligned time-series point cloud used for benchmarking. **(A)** Overhead view of the reconstructed time-series point cloud; 3D points were colorized using information from the sensor's camera. **(B)** Perspective view of the point cloud colorized by data collection session; numbered insets show close-up views for the distinctive objects enclosed in pink boxes. Section B'-B' shows a slice of the point cloud taken from the direction indicated by the pink arrows. Different solid colors denote point clouds collected at different dates.

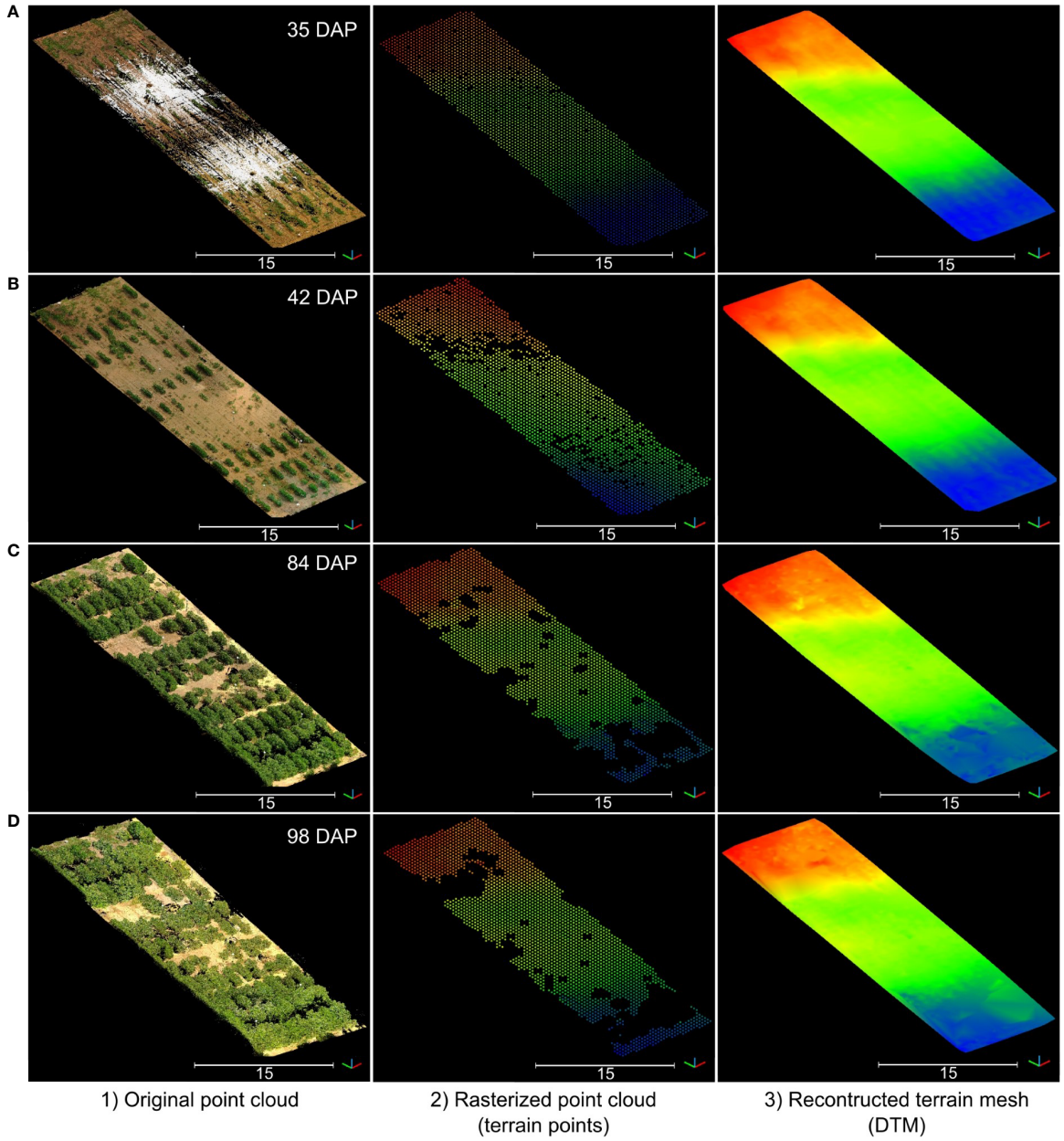


FIGURE 6
Reconstructed digital terrain model (DTM) for four data collection sessions during the vegetative growth stage. Rows **(A–D)** correspond to data collected at 35 days after planting (DAP), 42 DAP, 84 DAP, and 98 DAP, respectively. For each row, from left to right: (1) RGB point cloud; (2) Rasterized point cloud colorized by elevation; and (3) Reconstructed DTM colorized by elevation. Elevation is represented using a color map ranging from blue (the lowest point) to red (the highest point).

based alignment step (Figure 8A, inset 2) to 2.3 cm for the refinement step based on BEV (Figure 8B, inset 2).

In general, the alignment of point clouds remained consistent throughout the growing season, with final alignment errors comparable to those achieved by manual alignment of the point cloud pairs. In the manually aligned point clouds, the distance between point clouds at the overall level ranged between 0.5 and 1 cm, while at the ground level, the distance in the Z axis between point clouds reached approximately 6 mm. To see a detailed visual comparison between the manually aligned point clouds and those registered using our methodology, refer to Supplementary Figure S4.

3.4 Quantitative analysis of errors for spatiotemporal registration

Our quantitative analysis confirmed significant reductions in the main distances between point clouds registered using our methodology and their manually aligned counterparts (Table 1).

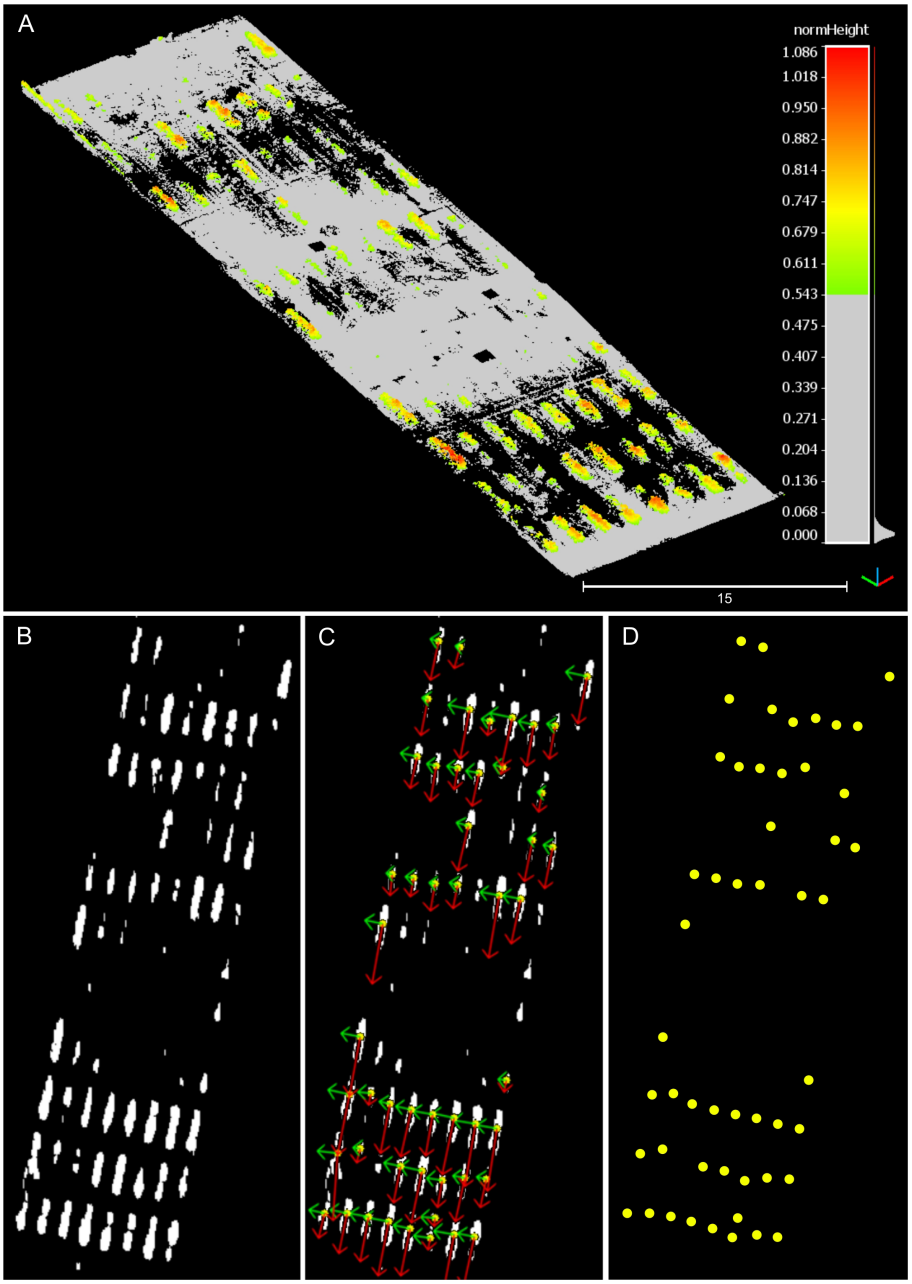


FIGURE 7
Partial results during the bird's-eye view alignment process. The depicted point cloud data pertains to the data collection session conducted at 42 DAP. **(A)** Canopy points isolation after height thresholding. **(B)** Projection of canopy points onto the X-Y plane in 2D. **(C)** Identified pixel clusters and principal component analysis results. Red arrows depict the first principal component's direction, while green arrows illustrate the second principal component's direction. **(D)** Visual depiction of identified plot centroids.

After the initial alignment based on DTM matching, the average Hausdorff distance between point clouds was approximately 31 cm, consistent with our previous qualitative analysis (Figure 8). With the second alignment step, the average Hausdorff distance was reduced to 5.5 cm, representing an 83% reduction from the initial alignment errors. However, despite this improvement, the alignment error for the final data collection session remained relatively high, reaching 13.5 cm even after the refinement step, in contrast to the approximately 3 cm observed for the other point clouds. This difference can be attributed to the limitations of the initial alignment based on the DTM in advanced growth stages. The denser and taller canopy obstructed the laser beam during the survey more frequently, resulting in limited terrain data for ICP registration. This contributed to larger initial errors that the refinement step, based on the BEV alignment approach, was not able to completely rectify.

A deeper analysis of registration errors revealed a clear trend: as the cotton canopy grew and became denser, the registration errors increased both in magnitude and variability (Figure 9). Early stages showed relatively low and uniform errors, while later stages

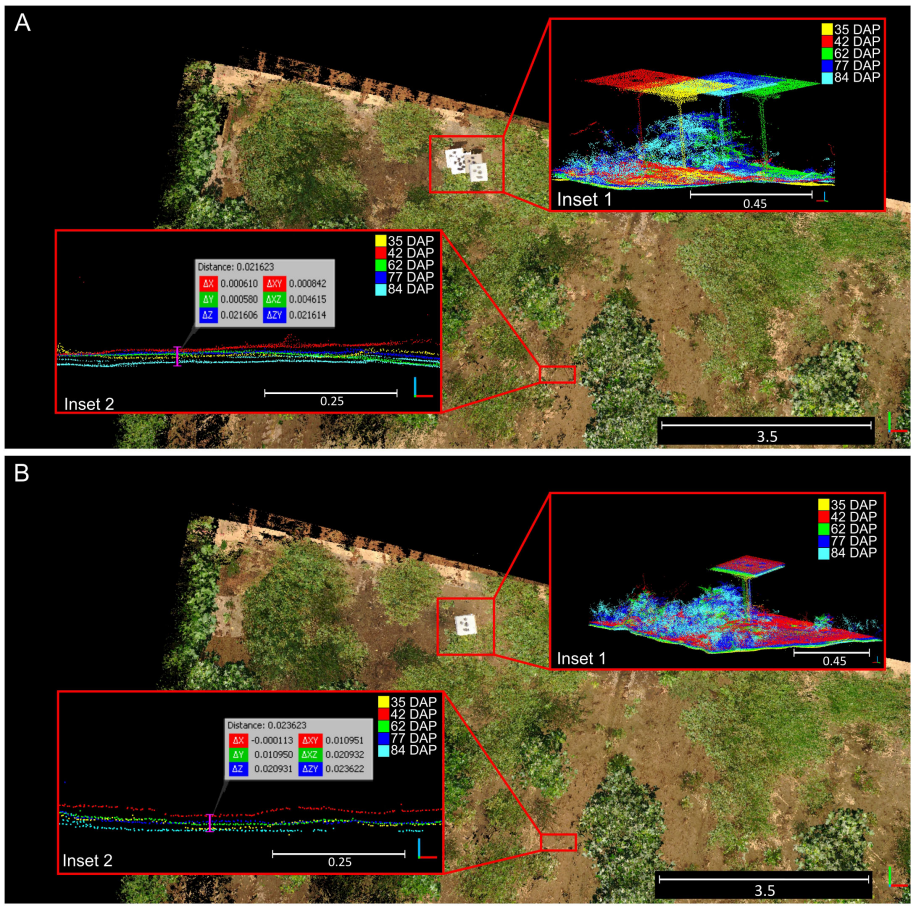


FIGURE 8
Qualitative results for the two-step registration process for point clouds collected over time. **(A)** Rough alignment results after the first alignment step based on digital terrain model matching. **(B)** Final alignment achieved after bird's-eye view refinement. Insets (1) demonstrate generalized alignment errors; Insets (2) highlight errors in the Z direction at the terrain level. Different solid colors denote point clouds collected at different dates.

exhibited higher errors due to the complexity of the mature canopy. This progression of errors from the early to maturity stages demonstrated the increasing difficulty in point cloud registration as the crop canopy developed, highlighting the importance of robust registration methodologies to handle increased canopy density and complexity.

At 42 DAP, during the early-stage canopy development with minimal plant overlap, distance errors peaked at 3.24 cm, with most errors being moderate. The majority of errors ranged between 0.38 cm and 1.13 cm, suggesting a low error spread. At 62 DAP, with increased canopy coverage and uniform growth, the maximum error reached 2.20 cm, with errors more uniformly distributed

TABLE 1 Spatiotemporal registration errors after each alignment step.

	Step 1 Terrain model-based alignment								Step 2 Bird's-eye view-based alignment							
			d_{H_X}		d_{H_Y}		d_{H_Z}				d_{H_X}		d_{H_Y}		d_{H_Z}	
DAP	RMS↓ (std)	d_H	max	min	max	min	max	min	RMS↓ (std)	d_H	max	min	max	min	max	min
42	2.1 (1.9)	19.3	18.3	-16.9	16.1	-18.3	17.6	-15.8	1.2 (0.6)	3.2	3.0	-3.1	3.0	-3.1	3.2	-3.1
62	4.9 (6.5)	32.5	32.0	-32.3	31.0	-30.0	31.8	-29.4	0.5 (0.3)	2.4	2.1	-2.1	2.1	-2.1	2.1	-2.1
84	3.6 (3.6)	24.2	22.3	-24.0	21.8	-22.6	23.0	-22.4	0.8 (0.5)	3.5	3.2	-3.0	3.0	-3.1	3.2	-3.0
98	5.7 (5.7)	46.8	36.3	-46.1	43.5	-36.9	39.7	-40.1	2.5 (1.8)	13.5	11.7	-12.5	12.2	-11.4	12.3	-11.3

All values are given in centimeters.

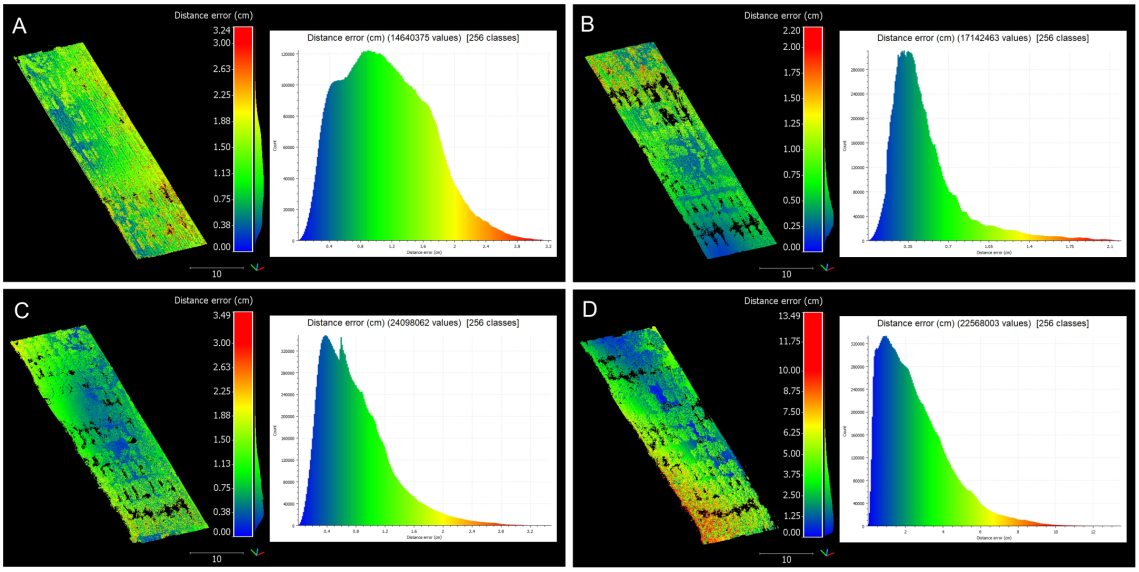


FIGURE 9
Distribution of registration errors across data collection sessions. Comprehensive analysis of registration errors showcasing both the spatial distribution of distance errors and their respective histograms. **(A)** 42 days after planting (DAP). **(B)** 62 DAP. **(C)** 84 DAP. **(D)** 98 DAP. The distribution of distance errors is visualized with a color gradient ranging from blue (low error) to red (high error).

across the field compared to the earlier stage. Errors were primarily concentrated between 0.25 cm and 0.75 cm, reflecting improved registration accuracy during this stage. This improvement could be attributed to better performance of the BEV refinement step, where increased canopy density provided more information to identify plot centroids, enhancing alignment accuracy. Additionally, as the crop grew taller, interference from weeds diminished, aiding in more accurate plot identification.

By 84 DAP, when the crop already reached canopy closure stage, the maximum error increased to 3.49 cm, with higher variability in areas of dense canopy growth. Errors ranged more broadly, with a significant portion between 0.75 cm and 2.63 cm, highlighting the challenges of maintaining accuracy as the canopy became denser. This stage reflected the increasing complexity of the registration process as the canopy structure became more intricate and plots overlapped, posing significant challenges to maintaining low error rates. At 98 DAP, errors reached a maximum of 13.49 cm, demonstrating significant challenges due to dense and overlapping canopies. A wide distribution of errors, clustering between 2 cm and 10 cm, indicated substantial variability and the presence of outliers caused by canopy occlusion and overlapping. At this point, the refinement step was no able to reduce alignment errors and optimize registration.

3.5 Phenotypic traits estimation

Our spatiotemporal registration methodology significantly enhanced CH estimations compared to conventional TLS data analysis without temporal registration (Supplementary Table S3). All analyzed percentiles for CH estimation—95^{th}, 99^{th}, and maximum height—exhibited a strong relationship with actual measurements for both methods. However, after the two-step spatiotemporal registration process, CH estimations consistently demonstrated stronger correlations with ground truth measurements and reduced errors compared to estimating CH from individual point clouds without temporal co-registration.

Estimations of CH using the maximum height value consistently exhibited the strongest correlation with actual CH values, explaining almost 95% of the total variance. While CH95 and CH99 also demonstrated strong correlations with ground truth measurements, CHmax consistently outperformed them across various evaluation metrics. Notably, CHmax showed reduced average deviation from the actual values and lower errors compared to CH95 and CH99, with an RMSE below 8 cm and average deviation of about 5%. These results are in line with findings from previous LiDAR-based studies in cotton (Sun et al., 2017, 2021), indicating the reliability of CHmax for estimating canopy height in cotton crops from multi-scan TLS, particularly in the context of time-series data analysis. CHmax captures the tallest point in the canopy, akin to manual field measurements, making it less susceptible to variability in canopy structure compared to percentile-based estimations, especially after denoising the point clouds.

3.5.1 Overall performance analysis

Our methodology for processing and analyzing in-field time-series TLS data demonstrated robust performance in estimating key traits versus ground truth measurements (Figure 10). The regression analysis showed a strong agreement between estimated values and ground truth measurements for both CH and $IPAR_f$. Despite the complex and evolving nature of the crop, our methodology allowed for the precise estimation of these traits, accurately capturing the dynamic changes in crop traits over time.

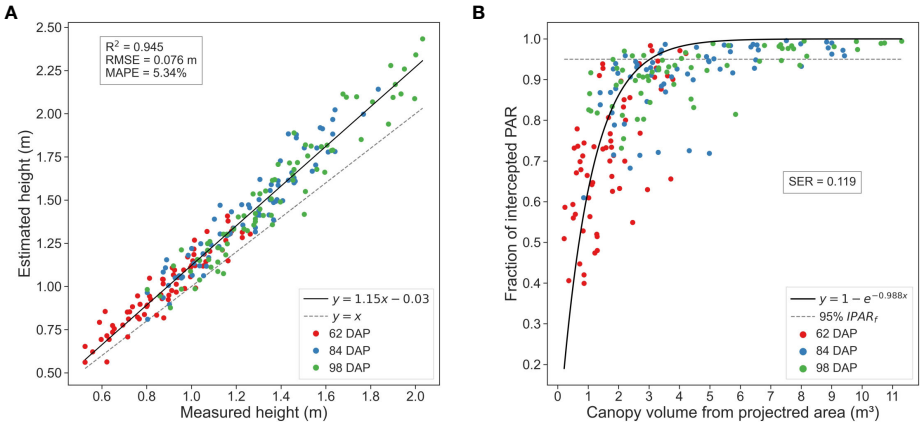


FIGURE 10
Phenotypic traits estimation results. **(A)** Canopy height estimations computed using the maximum height value versus ground truth measurements. **(B)** Canopy volume estimations versus ground truth measurements of fraction of intercepted photosynthetic active radiation (PAR).

Regarding CH estimations, we observed a general slight overestimation of canopy height compared to the actual values, which became more pronounced as the crop matured (Figure 10A). The linear regression model displayed a slight deviation from the 1:1 line, a trend consistent across other percentiles, albeit with slight variations. For instance, analysis of the 95^{th} percentile revealed a more pronounced tendency to overestimate canopy height during early growth stages, suggesting it may be less reliable for estimating in-field cotton plots height.

Analyzing the results for each data collection session individually, error values showed an increasing trend with crop development, leading to more dispersed CH estimations for larger canopies during later growth stages. RMSE values varied approximately 2 cm from the initial session to the last. At 62 DAP, the RMSE surpassed 6 cm, increasing to approximately 7 cm at 84 DAP and 8 cm at 98 DAP. This trend may be attributed to diverse canopy development of plants amongst plots or plant lodging during the season. Other potential causes of errors included underestimation in terrain elevation in regions with dense vegetative growth, as discussed previously (Section 3.2), or potential human error, given the challenges of measuring tall (reaching more than 2 m) and dense crops in the field. Nevertheless, examination of MAPE values revealed a consistent performance of our methodology, with only a 1.42% difference between the maximum and minimum MAPE values. Specifically, at 62 DAP, MAPE reached 6.17%, at 84 DAP it was 4.76%, and at 98 DAP, 5.11%. This suggests that, despite the observed increase in RMSE, our estimations remained relatively close to the actual values across different growth stages.

Upon examining the estimated CV in relation to field measurements of $IPAR_f$, we observed a pattern of exponential saturation with a distinct asymptote as the crop canopy approached its maximum volume (Figure 10B). This pattern aligns with the classical Beer-Lambert's law of attenuation of light through the canopy, offering a straightforward approach for predictive crop physiological traits estimation. Notably, the critical leaf area index at which 95% of radiation is intercepted was reached at canopies volumes between 3 and 4 m^3. The crop attained these volumes in a generalized manner between the first and second ground truth data collection dates, suggesting the likely canopy closure between 62 DAP and 84 DAP, a timeframe consistent with bibliography on cotton crop physiology (Snider et al., 2021). These findings provide valuable insights into the dynamic interplay between canopy structure and light interception efficiency, key for optimizing crop productivity.

3.5.2 Validation of CH estimates over time

Our spatiotemporal registration of TLS data provided reliable estimations of CH over time (Figure 11). The comparison between TLS-based estimates and ground truth measurements indicated that the temporal variation of predicted and observed data was generally within 7 cm for most genotypes and data collection sessions. This agreement was particularly strong for canopies under 1.5 meters, where CH estimations closely matched actual values. However, for canopies exceeding 1.5 meters, our estimations tended to slightly overestimate CH, which is consistent with our previous regression analysis.

The proximity of our estimated values to actual measurements indicated the accuracy of our methodology. CH estimations generally followed trends of steady growth, maintaining close alignment with ground truth values. In several cases, such as UA 48 or ST5020, the estimated CH closely matches the ground truth, demonstrating the reliability of our scanning technique. However, we found some discrepancies between estimated and ground truth values for certain genotypes, such as T0018MDN or MDN0101, highlighting areas for potential refinement in our scanning process. These discrepancies can be attributed to two main factors. Firstly, these two genotypes showed the highest growth rate, and as the canopies grew taller and denser, measuring their height in the field became more challenging and error-prone, introducing noisy points that could mask the performance of our methodology. Secondly, the increased complexity of the canopy structure at later growth stages might have affected the accuracy of the TLS data registration, leading to slight overestimations in CH.

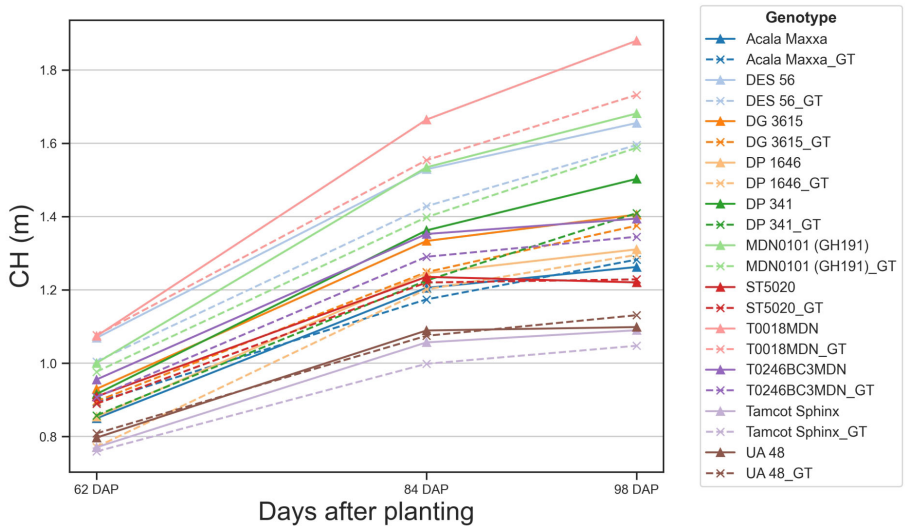


FIGURE 11
Temporal variation of predicted and observed canopy height. Solid lines represent canopy height (CH) values estimated using our methodology. Dashed lines represent ground truth (GT) CH values measured manually in the field. Each color and symbol combination represents a different genotype, with corresponding pairs (e.g., Acala Maxxa and Acala Maxxa_GT) for each genotype.

Our methodology effectively captured the variations in CH among different genotypes over time. Some genotypes exhibited considerable growth, and our methodology tracked this evolution relatively accurately. At 62 DAP, CH among genotypes showed minimal variation, mostly clustered between 0.8 and 1.0 meters, and our estimation errors were below 5 cm. However, as the crop matured, our CH estimates began to slightly diverge from the actual values. By 98 DAP, there was greater diversity in CH, ranging from nearly 1.8 meters for some genotypes to around 1.2 meters for others. At this stage, TLS estimates tended to be higher than the actual measurements, with estimation errors reaching approximately 20 cm for certain genotypes. These observations were consistent when analyzed at the individual plot level (Supplementary Figure S5). For genotypes such as DES 56 or DG 3615, average estimation errors remained under 7 cm. However, in plots of genotypes like T0018MDN, errors approached 25 cm at later DAP, indicating challenges in accurately measuring CH as canopies became taller and denser.

3.6 Analysis of phenotypic traits over time

Our methodology for collecting time-series TLS data provided highly detailed 3D information, enabling clear tracking of canopy evolution at the plot level across the distinct crop growth phases (Figure 12). The application of multi-scan TLS proved to be a robust technique for crop trait modeling over time, facilitating the reconstruction of detailed 4D models to successfully identify the primary growth phases leading up to crop maturity. The progression of canopy expansion is illustrated in Supplementary Figure S6, which shows a detailed profile of the distribution of canopy area per height layer over time for all genotypes. In general, during the initial growth stages, steady stem elongation and leaf area expansion were observed. With increasing resource availability, plant canopy expanded significantly both vertically and horizontally, resulting in exponential leaf area growth for almost all genotypes between 40 and 70 DAP. As the crop approached canopy closure, the canopy height tended to peak, yet vegetative branches continued to develop, significantly increasing lateral growth and the overall occupied volume until reaching canopy closure, where the canopy nears its upper size limit. This approach facilitated the extraction of morphological traits at the plot level and enabled comparative analysis across different genotypes throughout the growing season.

3.6.1 Crop growth model selection

The logistic model consistently emerged as the best-performing model across the analyzed traits, exhibiting the lowest AIC and BIC scores for modeling CH and CV evolution (Supplementary Table S4). Notably, the variations in performance among the models were relatively modest, indicating that all selected models provided a reasonably good fit to the data. The ΔAIC and ΔBIC values, which represent the differences in Akaike Information Criterion and Bayesian Information Criterion values, respectively, offer a comparative perspective on each model's performance relative to the best-fitting model.

In our study, the logistic and 3P-Richards models excelled in estimating CH growth parameters. For CH estimations, the logistic model was closely followed by the 3P-Richards model, with only a 4-point difference. The Gompertz model performed notably worse, with differences exceeding 50 points in both AIC and BIC scores. Conversely, the logistic model outperformed the others in CV growth modeling. Here, the 3P-Richards model had the worst AIC and BIC scores, while the Gompertz model's performance was intermediate, with AIC and BIC scores 10 points higher than those of the logistic model.

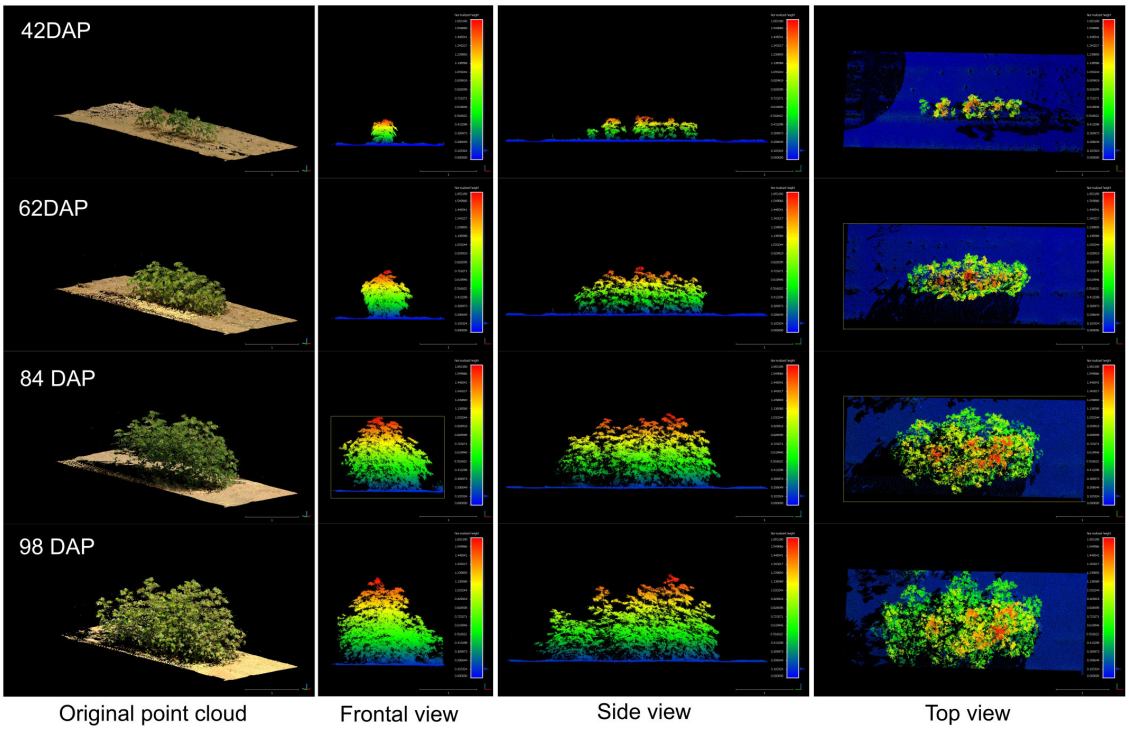


FIGURE 12
Growth comparison for a sample plot (Plot ID 906) during the vegetative crop growth. The left column shows the original RGB point cloud. The rest of columns show the frontal view, side view, and top view for each respective data collection session colorized by plant height. The figure is color-coded based on plant height, normalized with respect to the terrain (DTM), with blue tones indicating points closer to the ground.

3.6.2 Traits evolution modeling

Our findings revealed significant variability in CH growth parameters across different cotton genotypes, including both maximum CH value and inflection point (Table 2). In order to ensure the convergence of the mathematical model and facilitate the estimation of other growth parameters using the available data, we assumed a constant logistic growth rate (k) across all genotypes.

TABLE 2 Comparison of logistic growth parameters for canopy height per genotype.

Genotype	k_{CH}	A_{CH}†	SE	T_{iCH}†	SE
Tamcot Sphinx	0.06	1.24^{a}	0.112	53.3^{abcd}	1.153
UA 48	0.06	1.26^{a}	0.097	51.2^{ab}	0.875
ST5020	0.06	1.40^{ab}	0.097	49.1^{a}	0.792
Acala Maxxa	0.06	1.40^{ab}	0.104	53.2^{abc}	1.049
DP 1646	0.06	1.48^{ab}	0.104	55.7^{cde}	0.918
T0246BC3MDN	0.06	1.58^{abc}	0.136	53.3^{bcd}	1.028
DG 3615	0.06	1.59^{ab}	0.104	52.5^{abc}	0.941
DP 341	0.06	1.62^{abc}	0.105	56.1^{cde}	1.145
DES 56	0.06	1.81^{bc}	0.104	54.4^{bcd}	0.804
MDN0101 (GH191)	0.06	1.84^{bc}	0.099	57.7^{de}	0.909
T0018MDN	0.06	2.06^{c}	0.100	59.3^{e}	0.871

Means and standard error (SE) for maximum height (A_{CH}) in meters and inflection point (T_{iCH}) in days after planting. The logistic growth rate (k_{CH}) was assumed constant for all the genotypes. †Means not sharing any letter are significantly different by the Tukey-test at the 5% level of significance.

The maximum value in a sigmoidal growth curve represents the upper limit or saturation point of growth for the crop. In our analysis, genotype T0018MDN exhibited the highest maximum CH among all genotypes, reaching nearly 2 meters. Genotypes MDN0101 (GH191) and DES 56 closely followed, with a maximum CH of approximately 1.8 meters. In contrast, genotypes Tamcot Sphinx and UA48 were significantly smaller, with maximum CH values around 1.2 meters. The remaining genotypes fell in between, with CH values ranging from 1.4 to 1.6 meters. Notably, ST5020, Acala Maxxa, DP 1646, and DG 3615 had significantly shorter canopies compared to T0018MDN.

The inflection point is a critical feature that marks the change in crop growth dynamics. This point, where the curve's slope is at its maximum, represents the stage where the rate of growth transitions from being exponential to linear, signifying the phase of most rapid change in the growth rate. In our experiment, T0018MDN genotype was the last in reaching maximum growth rate, occurring after 59 DAP (T_i = 59.3 DAP). This was significantly later than the other genotypes, except for MDN0101 (GH191), DP 341, and DP 1646. In contrast, ST5020 and UA48 took around 50 days to reach their maximum growth rates. The rest of the genotypes exhibited inflection points of 53 DAP or more. Notably, MDN0101 (GH191) was significantly more slowly in reaching its peak than

TABLE 3 Comparison of logistic growth parameters for canopy volume (CV) per genotype.

Genotype	k_{CV}	A_{CV}†	SE	T_{iCV}†	SE
Tamcot Sphinx	0.106	2.76^{a}	0.92	64.4^{abc}	1.435
UA 48	0.106	3.68^{ab}	0.80	64.8^{ab}	1.058
Acala Maxxa	0.106	4.28^{ab}	0.86	64.0^{ab}	1.116
ST5020	0.106	4.28^{ab}	0.80	65.3^{abc}	1.139
DP 1646	0.106	4.77^{ab}	0.85	70.1^{c}	1.157
T0246BC3MDN	0.106	4.84^{ab}	1.12	66.1^{abc}	1.308
MDN0101 (GH191)	0.106	5.40^{ab}	0.80	62.6^{a}	0.846
DP 341	0.106	6.45^{ab}	0.85	68.3^{bc}	1.218
DES 56	0.106	6.49^{ab}	0.85	66.7^{abc}	0.962
DG 3615	0.106	6.67^{ab}	0.85	67.4^{bc}	1.087
T0018MDN	0.106	7.25^{b}	0.79	67.0^{bc}	0.849

Means and standard error (SE) for maximum canopy volume (A_{CV}) in cubic meters and inflection point (T_{iCV}) in days after planting. The logistic growth rate (k_{CV}) was assumed constant for all the genotypes. †Means not sharing any letter are significantly different by the Tukey-test at the 5% level of significance.

DG 3625, Acala Maxxa, UA48, and ST5020, taking approximately 54 DAP.

Our analysis of CV estimations over time revealed substantial genotype-dependent differences in CV growth (Table 3). Similar to the CH analysis, we assumed a constant logistic growth rate (k) for all genotypes to ensure model convergence. Among the genotypes assessed, T0018MDN exhibited the highest plant volume, surpassing 7 m^3. Following closely were genotypes DG 3615, DES 56, and DP 341, each with volumes exceeding 6 m^3. In contrast, Tamcot Sphinx attained a maximum volume of less than 3 m^3, significantly smaller than T0018MDN. The remaining genotypes displayed CV values ranging between 3 and 7 m^3. Regarding the inflection point for CV growth modeling, genotype DP 1646 demonstrated the highest value among all genotypes, occurring at approximately 70 DAP. Most other genotypes exhibited inflection points between 64 and 68 DAP. MDN0101 (GH191) emerged as the most precocious in reaching its maximum growth rate, at around 63 DAP, indicating a significant earlier maturation compared to genotypes T0018MDN, DG 3615, and DP 341.

The complex interplay between genotype, environment, and crop development is evident in the diverse morphological traits analyzed and the corresponding evolution curves of CH (Figure 13) and CV (Figure 14). The evolution of both CH and CV reflects the gradual pace of growth of the crop during the initial stages. As the season progressed and resources become more available, plant canopy expanded both in height and laterally, resulting in exponential CV growth across all genotypes between 40 and 70 DAP. This period of rapid development indicated the crop's increasing capacity to capture light energy. Around 75 DAP, a visible slowdown in stem elongation is observed across nearly all genotypes, signaling the approach to canopy closure. This phenomenon suggests that the crop is nearing its upper size limit, with further CH growth becoming increasingly constrained. Tracking growth traits over time provides valuable insights into the dynamic evolution of cotton crop development, including genotype-specific responses to environmental cues and management practices. These insights can guide breeders and researchers in selecting genotypes with desired traits to improve yield, enhance pest resistance, and ensure adaptability to diverse growing conditions.

4 Discussion

4.1 Time-series TLS data for field phenotyping

Using time-related data in plant phenotyping represents a valuable approach to gaining a deeper understanding of the temporal dynamics of plant growth. Plants respond to changing environmental conditions, and field data collected over time can capture these variations, providing continuous insights into trait evolution during the course of a growing season. This can allow for a better understanding of crop development (Pauli et al., 2016), holding promise for advancing our knowledge of plant biology and supporting the development of crops adapted to varying environmental conditions (Miao et al., 2020). In plant breeding, this information can be key to support genotype selection based on new traits that allow for a more efficient use of resources (Li et al., 2022).

Traditional plant phenotyping methods often limit analyses to a single time point, potentially overlooking vital changes and interactions between the crop and its environment during critical developmental stages (Tardieu et al., 2017). Unlike conventional static measurements, time-series data are composed of a sequence of data collected at different timestamps. This temporal context facilitates the identification of key stages of development and growth patterns, offering valuable information about the rate of growth, periods of stability, and potential stress responses (Chen et al., 2014).

Integrating time-series data with advanced 3D technologies like LiDAR enhances phenotyping for precision agriculture. LiDAR technology has contributed to the advancement of field phenotyping, offering direct access to complex 3D morphological trait information and allowing for detailed reconstructions of plant structures (Guo et al., 2018). Time-series TLS data can improve the accuracy of growth models by capturing intricate details of plant development during the whole growing season (Jin et al., 2021b). Our work demonstrated that TLS time-series data can provide consistent information on traits such as canopy height and volume, enabling precise crop traits tracking. This information can increase our understanding of critical growth parameters to optimize management strategies and make more informed decisions in crop breeding.

4.2 Importance of 4D registration of TLS data for crop growth tracking

Detailed 4D models, incorporating the temporal dimension into 3D spatial data from TLS scans, provide a powerful tool for

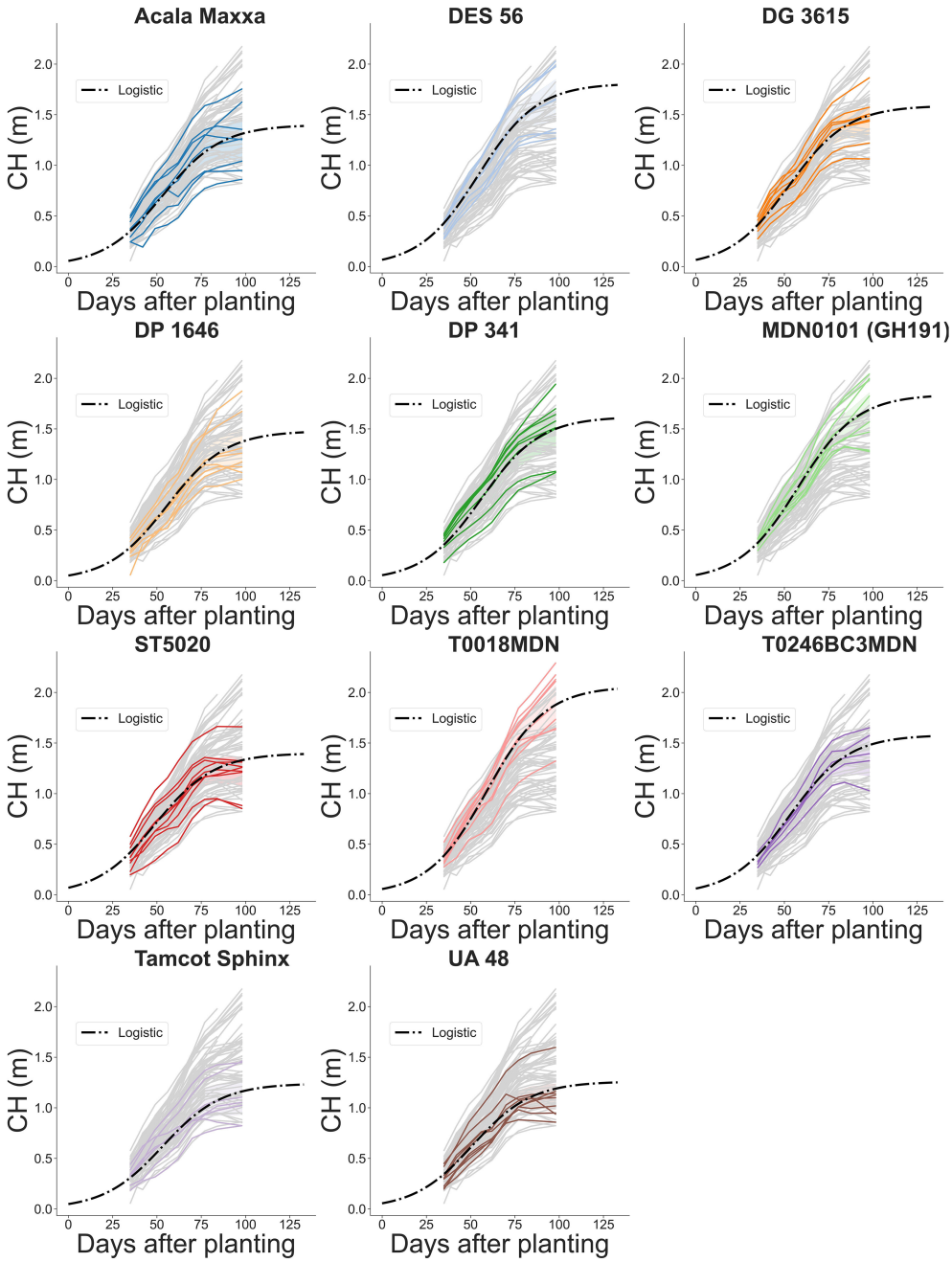


FIGURE 13
Temporal evolution of plot-level canopy height (CH) across the growing season for 11 genotypes. Grey lines indicate estimated CH values for all plots in the field. Colored solid lines highlight the estimated CH for a specific genotype. Dash-dotted lines depict fitted logistic growth curves.

analyzing and tracking morphological traits across the growing season. However, generating these models poses unique challenges, primarily from the need to align repeated measurements over time. The complex transformations that crops undergo during the growing season complicate the registration of point clouds collected at different time points. Ensuring the consistency and reliability of TLS data collection and processing becomes essential in field conditions (Pieruschka and Schurr, 2019).

Accurate data alignment is crucial for tracking changes in plant structure. While some methods use fixed targets to aid co-registration (Tilly et al., 2014; Friedli et al., 2016; Su et al., 2019), this approach can be labor-intensive and error-prone. The consistent positioning of targets in agricultural fields, where machinery needs to operate or other experiments need to be executed, can be problematic. Deviations in the target placement between surveys can impact the temporal alignment accuracy, and hence the estimation of traits over time.

Inspired by the positive results from a previous study in maize, soybean, and wheat using semi-permanent targets (Friedli et al., 2016), we initially adopted this concept to benchmark our

10.3389/fpls.2024.1436120

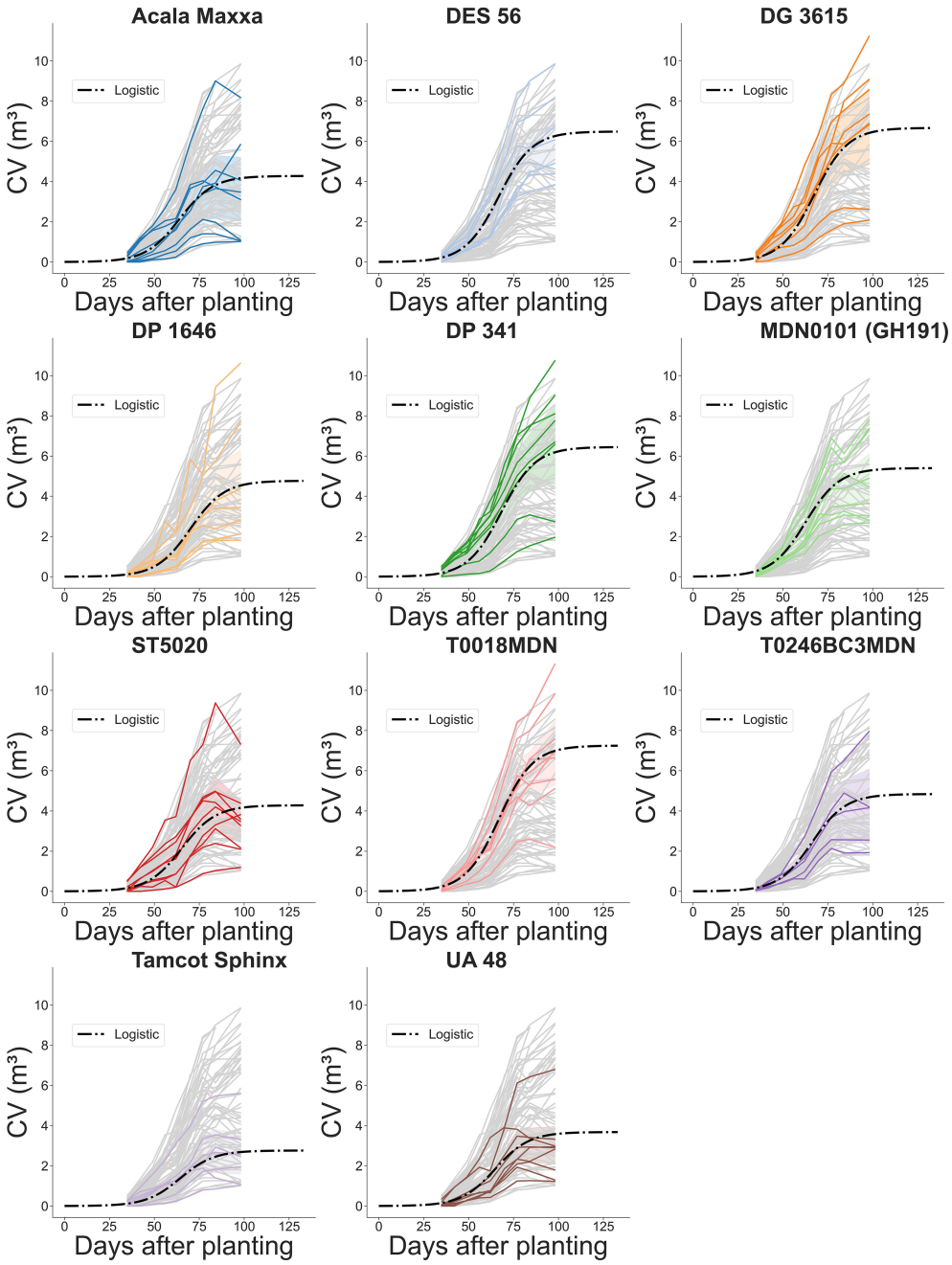


FIGURE 14
Temporal evolution of plot-level canopy volume (CV) during the growing season for 11 genotypes. Grey lines represent estimated CV values for every plot computed using the method based on projected area. Colored solid lines highlight the estimated CV of individual plots for a specific genotype. Dash-dotted lines depict fitted logistic growth curves.

methodology. This study reported deviations in the positions of spherical targets between 2.5 mm and 10 mm. However, after preliminary processing of data from our initial collection sessions, we found deviations in target locations exceeding 20 mm. These larger deviations were not due to the registration process itself but were likely caused by external factors beyond our control. Our experimental field was also used for other studies involving autonomous robot navigation, and areas with heavy foot and tractor traffic seemed to impact the soil around the spheres, causing the structures to shift.

To address this issue, we proposed an innovative approach that relies solely on the collected data for registration, reducing dependence on artificial targets and minimizing the impact of such external factors. By not relying on physical markers that can be displaced, our approach maintains accuracy over long periods and in environments subject to change. This flexibility makes it especially suitable for dynamic field conditions where traditional fixed targets might fail to provide consistent accuracy. This streamlines field setups and ensures consistent and reproducible results.

Our approach has demonstrated effectiveness in overcoming these challenges, offering an accessible and robust methodology for TLS-based phenotyping in dynamic field environments. Unlike approaches requiring continuous acquisition for ensuring common features for registration (Li et al., 2023), our multi-step TLS data alignment leveraged invariant elements naturally present in the collected point clouds to roughly align them. This allowed us to decouple the registration process from changing elements such as the crop, facilitating the registration of point clouds collected at distant time intervals. The observed registration errors, comparable to expected errors in rigid registration at the organ level (Chebrolu et al., 2021) and falling within the range of RMSE values for height estimation, indicate the accuracy of our approach in capturing time-series TLS data. This highlights the potential of our method for LiDAR-based crop phenotyping under dynamic field conditions.

4.3 Challenges in growth modeling for crop breeding

Understanding the growth patterns of cotton plants is essential for effective crop management and informed breeding strategies (Ritchie et al., 2007). Cotton plant development typically follows a sigmoid function, characterized by slow initial growth during establishment, followed by exponential vegetative growth that gradually slows as the crop approaches canopy closure (Snider et al., 2021). Growth modeling plays a key role in estimating key parameters defining these growth curves (Gregorczyk, 1998), providing a systematic approach to incorporate insights into crop phenotyping (Costa et al., 2019).

Growth modeling from field data can be challenging, and many of the growth parameters extracted with our methodology could not be obtained any other way on a large scale. Focused on the analysis of three key morphological traits, CH, CA, and CV, across diverse cotton genotypes, our study unraveled the dynamics of cotton growth and the relationships between growth parameters, providing valuable insights for genotype selection tailored to specific requirements. However, fitting maximal models with all random effects may fail to converge because the random effects structure has a complexity not supported by the underlying data (Barr et al., 2013). Simplification of the random effects structure can help the model to converge (Bates et al., 2015).

Trade-offs are often necessary due to computational limitations and the need for efficient model fitting (McCrea et al., 2023). Challenges may arise across the data analysis pipeline, including processing, model development, and information extraction. Fitting maximal models that include all random effects in growth parameters can face convergence challenges due to the complexity of the random effects structure (Barr et al., 2013). Our study revealed that growth rate could be one of the most complex parts of the random effects structure for crop growth modeling. Primarily, we dropped covariance terms for some of the random effects, as suggested in one paper (Seedorff et al., 2019) to try to achieve convergence. However, it was not sufficient for our model to converge, and we opted for a more drastic approach of fixing the growth rate for all the genotypes, removing the slope entirely, in order to reduce the complexity of our model. This led us to simplify the model by fixing the growth rate for all genotypes, emphasizing the importance of balancing model complexity and convergence (Bates et al., 2015).

4.4 Limitations and future work

The adaptability of TLS technology to capture 3D structural information makes it inherently suitable for studying a wide range of plant species. While our methodology has demonstrated effectiveness in a cotton breeding field, highlighting its potential application in similar agricultural contexts, its generalizability across diverse crops and environmental conditions remains to be fully explored.

Additionally, uncertainties in the use of estimated CV as a proxy for estimating canopy light interception should be acknowledged. Variability in environmental conditions can affect the accuracy of $IPAR_f$ measurements and validating its relationship with our CV estimation method warrants further research. Understanding these uncertainties is key for enhancing the reliability and applicability of our methodology across different agricultural settings.

Further experimentation across different field settings is imperative to comprehensively assess the adaptability and generalizability of our methodology. Expanding these efforts to accommodate various research environments will be key to establishing a scalable methodology for consistent field phenotyping over time. Understanding how factors such as canopy structure, plant density, and environmental influences interact with our methodology is crucial for evaluating its scalability and robustness across different crop types and field conditions. Our initial findings indicated that our methodology could create accurate 4D crop models under conditions of minimal plant overlap. However, maintaining model accuracy became challenging as canopy density increased. Different crop types may require tailored strategies for BEV-based alignment due to their unique growth characteristics. For example, maize plants, which are individually planted and in early stages may project only a small footprint, may require a more precise clustering process. Similarly, densely planted crops like wheat can introduce additional complexities in data interpretation and analysis.

Despite the effectiveness and accuracy achieved in our cotton field, challenges such as laser beam shadowing and occlusions were not entirely eliminated, particularly in advanced growth stages. These issues are common in many remote and proximal sensing studies conducted in the field, demanding further research for resolution. The complex structure of crops like cotton poses a challenge, where increasing the number of scan locations for multi-scan TLS-based field analysis may not completely eliminate occlusion effects. Striking a balance between the need for more scans and efficient resource utilization, including time and computing power, necessitates thorough site studies for TLS site planning to optimize data collection. We are actively exploring the potential of physics-based simulators for knowledge-guided TLS site planning. These simulations can provide valuable insights into

the optimal distribution of scan locations, especially in complex field environments, enhancing the efficiency of TLS-based field phenotyping.

Moreover, the increased number of scan locations introduces new challenges, particularly in terms of time consumption and logistics. Traditional practices for LiDAR-based scanning involve manually moving the scanner from one scan location to another, contributing to the time-intensive nature of data collection. In a prior study, we demonstrated the potential use of a ground robot to automate TLS data acquisition in a breeding field (Rodriguez-Sanchez and Li, 2022). We are working on improving this system to autonomously determine the number of scans and their distribution throughout the field, thereby fully automating the TLS-based phenotyping process. This work is currently under review in a reputed journal, and a preprint is available on arXiv (Rodriguez-Sanchez et al., 2024). This innovative approach aims to enhance the efficiency of TLS data collection for plant breeding, streamlining the phenotyping workflow and aligning with broader trends in automation and robotics within the field of agriculture.

5 Conclusions

In this study, we introduced an innovative methodology for precisely registering point clouds collected under field conditions, enabling LiDAR-based crop phenotyping over time. This work emphasizes the critical importance of precise point cloud data collection, accurate registration, and precise modeling for TLS-based field phenotyping. Our two-phase TLS data registration approach has demonstrated its effectiveness in aligning point clouds captured up to two months apart during the vegetative growth season, significantly reducing alignment errors. By leveraging terrain points and crop row distribution, our method provides a reliable and efficient solution for monitoring crop morphological growth, enabling breeders to consistently acquire accurate phenotypic traits from the same physical locations at different time points.

As digital technologies advance, the refinement of current procedures for in-field data collection and processing will strengthen our ability to enhance crop improvement in a more efficient manner. Despite challenges such as late-season occlusions, our approach presents a promising solution to enhance cotton breeding programs by offering a reliable digital approach to monitor traits over time. It provides a foundation for informed decision-making and genotype selection based on desirable growth characteristics. The findings of this study can contribute significantly to understanding cotton plant growth and genotype variations, offering valuable insights for optimizing cotton crop management and advancing plant phenomics.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

JR-S: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. JS: Methodology, Resources, Writing – review & editing. KJ: Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. CL: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by the National Science Foundation Growing Convergence Research (Award No. 1934481), and the Georgia Cotton Commission. The authors declare that this study received funding from Cotton Incorporated. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Acknowledgments

The authors would like to express their sincere gratitude to Dr. Andrew Paterson and his research group for providing the plant materials and their dedication in maintaining the experimental field. We also extend our thanks to Gary Pierce for his assistance during the field trials, and to Navneet Kaur for her efforts in collecting essential ground truth data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1436120/full#supplementary-material

References

Akaike, H. (1992). Information theory and an extension of the maximum likelihood principle. In *Springer series in statistics* (New York, NY: Springer New York), 610–624. doi: 10.1007/978-1-4612-0919-538

Baker, D. N., and Meyer, R. E. (1966). Influence of stand geometry on light interception and net photosynthesis in cotton. *Crop Sci.* 6, 15–19. doi: 10.2135/cropsci1966.0011183X000600010004x

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Memory Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D. M., Kliegl, R., Vasishth, S., and Baayen, H. (2015). Parsimonious mixed models. *ArXiv*. doi: 10.48550/arXiv.1506.04967

Besl, P., and McKay, N. D. (1992). A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* 14, 239–256. doi: 10.1109/34.121791

Carlone, L., Dong, J., Fenu, S., Rains, G., and Dellaert, F. (2015). Towards 4D crop analysis in precision agriculture: Estimating plant height and crown radius over time via expectation-maximization. In *ICRA workshop on robotics in agriculture*. (Seattle, WA, USA: International Conference on Robotics and Automation (ICRA)).

Chawade, A., van Ham, J., Blomquist, H., Bagge, O., Alexandersson, E., and Ortiz, R. (2019). Highthroughput field-phenotyping tools for plant breeding and precision agriculture. *Agronomy* 9, 258. doi: 10.3390/agronomy9050258

Chebrolu, N., Labe, T., and Stachniss, C. (2018). Robust long-term registration of UAV images of crop fields for precision agriculture. *IEEE Robotics Automation Lett.* 3, 3097–3104. doi: 10.1109/LRA.2018.2849603

Chebrolu, N., Labe, T., and Stachniss, C. (2020). Spatio-temporal non-rigid registration of 3D point clouds of plants. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. (Paris, France: IEEE), 3112–3118. doi: 10.1109/icra40945.2020.9197569

Chebrolu, N., Magistri, F., Läbe, T., and Stachniss, C. (2021). Registration of spatio-temporal point clouds of plants for phenotyping. *PloS One* 16, e0247243. doi: 10.1371/journal.pone.0247243

Chen, D., Neumann, K., Friedel, S., Kilian, B., Chen, M., Altmann, T., et al. (2014). Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *Plant Cell* 26, 4636–4655. doi: 10.1105/tpc.114.129601

Costa, J. M., Marques da Silva, J., Pinheiro, C., Barón, M., Mylona, P., Centritto, M., et al. (2019). Opportunities and limitations of crop phenotyping in Southern European countries. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01125

Crommelinck, S., and Höfle, B. (2016). Simulating an autonomously operating low-cost static terrestrial LiDAR for multitemporal maize crop height measurements. *Remote Sens.* 8, 205. doi: 10.3390/rs8030205

Dong, J., Burnham, J. G., Boots, B., Rains, G., and Dellaert, F. (2017). 4D crop monitoring: Spatio-temporal reconstruction for agriculture. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. (Singapore: IEEE), 3878–3885. doi: 10.1109/icra.2017.7989447

Eitel, J. U., Magney, T. S., Vierling, L. A., Greaves, H. E., and Zheng, G. (2016). An automated method to quantify crop height and calibrate satellite-derived biomass using hypertemporal LiDAR. *Remote Sens. Environ.* 187, 414–422. doi: 10.1016/j.rse.2016.10.044

Ermanis, A., Gobbo, S., Snider, J. L., Cohen, Y., Liakos, V., Lacerda, L., et al. (2020). Defining physiological contributions to yield loss in response to irrigation in cotton. *J. Agron. Crop Sci.* 207, 186–196. doi: 10.1111/jac.12453

Fix, E., and Hodges, J. L. (1951). *Discriminatory analysis: Nonparametric discrimination: Consistency properties*. (Randolph Field, TX: USAF School of Aviation Medicine). doi: 10.1037/e471672008-001

Friedli, M., Kirchgessner, N., Grieder, C., Liebisch, F., Mannale, M., and Walter, A. (2016). Terrestrial 3D laser scanning to track the increase in canopy height of both monocot and dicot crop species under field conditions. *Plant Methods* 12, 1–15. doi: 10.1186/s13007-016-0109-7

Furbank, R. T., and Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644. doi: 10.1016/j.tplants.2011.09.005

Gelard, W., Herbulot, A., Devy, M., and Casadebaig, P. (2018). 3D leaf tracking for plant growth monitoring. In *2018 25th IEEE International Conference on Image Processing (ICIP)*. (Athens, Greece: IEEE), 3663–3667. doi: 10.1109/icip.2018.8451553

Gregorczyk, A. (1998). Richards plant growth model. *J. Agron. Crop Sci.* 181, 243–247. doi: 10.1111/j.1439-037X.1998.tb00424.x

Großkinsky, D. K., Svensgaard, J., Christensen, S., and Roitsch, T. (2015). Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. *J. Exp. Bot.* 66, 5429–5440. doi: 10.1093/jxb/erv345

Günder, M., Ispizua Yamati, F. R., Kierdorf, J., Roscher, R., Mahlein, A.-K., and Bauckhage, C. (2022). Agricultural plant cataloging and establishment of a data framework from UAV-based crop images by computer vision. *GigaScience* 11. doi: 10.1093/gigascience/giac054

Guo, T., Fang, Y., Cheng, T., Tian, Y., Zhu, Y., Chen, Q., et al. (2019). Detection of wheat height using optimized multi-scan mode of LiDAR during the entire growth stages. *Comput. Electron. Agric.* 165, 104959. doi: 10.1016/j.compag.2019.104959

Guo, Q., Wu, F., Pang, S., Zhao, X., Chen, L., Liu, J., et al. (2018). Crop 3D-a LiDAR based platform for 3D high-throughput crop phenotyping. *Sci. China. Life Sci.* 61, 328—339. doi: 10.1007/s11427-017-9056-0

Herrero-Huerta, M., Lindenbergh, R., and Gard, W. (2018). Leaf movements of indoor plants monitored by terrestrial LiDAR. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00189

Hosoi, F., and Omasa, K. (2009). Estimating vertical plant area density profile and growth parameters of a wheat canopy at different growth stages using three-dimensional portable LiDAR imaging. *ISPRS J. Photogrammetry Remote Sens.* 64, 151–158. doi: 10.1016/j.isprsjprs.2008.09.003

Hosoi, F., and Omasa, K. (2012). Estimation of vertical plant area density profiles in a rice canopy at different growth stages by high-resolution portable scanning LiDAR with a lightweight mirror. *ISPRS J. Photogrammetry Remote Sens.* 74, 11–19. doi: 10.1016/j.isprsjprs.2012.08.001

Hui, F., Zhu, J., Hu, P., Meng, L., Zhu, B., Guo, Y., et al. (2018). Image-based dynamic quantification and high-accuracy 3D evaluation of canopy structure of plant populations. *Ann. Bot.* 121, 1079–1088. doi: 10.1093/aob/mcy016

Jiménez-Berni, J. A., Deery, D. M., Rozas-Larraondo, P., Condon, A. G., Rebetzke, G. J., James, R., et al. (2018). High throughput determination of plant height, ground cover, and above-ground biomass in wheat with LiDAR. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00237

Jin, S., Su, Y., Zhang, Y., Song, S., Li, Q., Liu, Z., et al. (2021a). Exploring seasonal and circadian rhythms in structural traits of field maize from LiDAR time series. *Plant Phenomics* 2021. doi: 10.34133/2021/9895241

Jin, S., Sun, X., Wu, F., Su, Y., Li, Y., Song, S., et al. (2021b). LiDAR sheds new light on plant phenomics for plant breeding and management: Recent advances and future prospects. *ISPRS J. Photogrammetry Remote Sens.* 171, 202–223. doi: 10.1016/j.isprsjprs.2020.11.006

Kaur, N., Snider, J. L., Paterson, A. H., Grey, T. L., Li, C., Virk, G., et al. (2023). Variation in thermotolerance of photosystem II energy trapping, intersystem electron transport, and photosystem i electron acceptor reduction for diverse cotton genotypes. *Plant Physiol. Biochem.* 201, 107868. doi: 10.1016/j.plaphy.2023.107868

Kim, J. Y. (2020). Roadmap to high throughput phenotyping for plant breeding. *J. Biosyst. Eng.* 45, 43–55. doi: 10.1007/s42853-020-00043-0

Lenth, R. V. (2023). *Emmeans: Estimated marginal means, aka least-squares means*, R package version 1.8.8. Retrieved from https://CRAN.R-project.org/package=emmeans

Li, D., Bai, D., Tian, Y., Li, Y.-H., Zhao, C., Wang, Q., et al. (2022). Time series canopy phenotyping enables the identification of genetic variants controlling dynamic phenotypes in soybean. *J. Integr. Plant Biol.* 65, 117–132. doi: 10.1111/jipb.13380

Li, Y., Fan, X., Mitra, N. J., Chamovitz, D., Cohen-Or, D., and Chen, B. (2013). Analyzing growing plants from 4D point cloud data. *ACM Trans. Graphics* 32, 1–10. doi: 10.1145/2508363.2508368

Li, Y., Wen, W., Fan, J., Gou, W., Gu, S., Lu, X., et al. (2023). Multi-source data fusion improves time-series phenotype accuracy in maize under a field high-throughput phenotyping platform. *Plant Phenomics* 5, 0043. doi: 10.34133/plantphenomics.0043

Lin, Y. (2015). LiDAR: An important tool for next-generation phenotyping technology of high potential for plant phenomics? *Comput. Electron. Agric.* 119, 61–73. doi: 10.1016/j.compag.2015.10.011

Loomis, R., and Williams, W. (1969). Productivity and the morphology of crop stands: Patterns with leaves. In *Physiological aspects of crop yield*. Eds. J. D. Eastin, F. A. Haskins, C. Y. Sullivan and C. H. M. van Bavel (Madison, WI: American Society of Agronomy, Crop Science Society of America), 27–47. doi: 10.2135/1969.physiologicalaspects.c3

Madec, S., Baret, F., de Solan, B., Thomas, S., Dutartre, D., Jézéquel, S., et al. (2017). High-throughput phenotyping of plant height: Comparing unmanned aerial vehicles and ground LiDAR estimates. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.02002

Magistri, F., Chebrolu, N., and Stachniss, C. (2020). Segmentation-based 4D registration of plants point clouds for phenotyping. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. (Las Vegas, NV, USA: IEEE), 2433–2439. doi: 10.1109/iros45743.2020.9340918

Malambo, L., Popescu, S., Horne, D., Pugh, N., and Rooney, W. (2019). Automated detection and measurement of individual sorghum panicles using density-based clustering of terrestrial LiDAR data. *ISPRS J. Photogrammetry Remote Sens.* 149, 1–13. doi: 10.1016/j.isprsjprs.2018.12.015

Malambo, L., Popescu, S., Murray, S., Putman, E., Pugh, N., Horne, D., et al. (2018). Multitemporal field-based plant height estimation using 3D point clouds generated from small unmanned aerial systems high-resolution imagery. *Int. J. Appl. Earth Observation Geoinformation* 64, 31–42. doi: 10.1016/j.jag.2017.08.014

McCrea, R., King, R., Graham, L., and Börger, L. (2023). Realising the promise of large data and complex models. *Methods Ecol. Evol.* 14, 4–11. doi: 10.1111/2041-210X.14050

Medic, T., Bömer, J., and Paulus, S. (2023). Challenges and recommendations for 3D plant phenotyping in agriculture using terrestrial lasers scanners. *ISPRS Ann. Photogrammetry Remote Sens. Spatial Inf. Sci.* X-1/W1-2023, 1007–1014. doi: 10.5194/isprs-annals-X-1-W1-2023-1007-2023. X-1/W1-2023.

Miao, C., Xu, Y., Liu, S., Schnable, P. S., and Schnable, J. C. (2020). Increased power and accuracy of causal locus identification in time series genome-wide association in sorghum. *Plant Physiol.* 183, 1898–1909. doi: 10.1104/pp.20.00277

Paproki, A., Sirault, X., Berry, S., Furbank, R., and Fripp, J. (2012). A novel mesh processing based technique for 3D plant analysis. *BMC Plant Biol.* 12, 63. doi: 10.1186/1471-2229-12-63

Pauli, D., Andrade-Sanchez, P., Carmo-Silva, A. E., Gazave, E., French, A. N., Heun, J., et al. (2016). Field-based high-throughput plant phenotyping reveals the temporal patterns of quantitative trait loci associated with stress-responsive traits in cotton. *G3 Genes—Genomes—Genetics* 6, 865–879. doi: 10.1534/g3.115.023515

Paulus, S., Schumann, H., Kuhlmann, H., and Léon, J. (2014). High-precision laser scanning system for capturing 3D plant architecture and analysing growth of cereal plants. *Biosyst. Eng.* 121, 1–11. doi: 10.1016/j.biosystemseng.2014.01.010

Pieruschka, R., and Schurr, U. (2019). Plant phenotyping: Past, present, and future. *Plant Phenomics* 2019. doi: 10.34133/2019/7507131

Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS* (New York: Springer). doi: 10.1007/b98882

Pokhrel, A., Snider, J. L., Virk, S., Sintim, H. Y., Hand, L. C., Vellidis, G., et al. (2023). Quantifying physiological contributions to nitrogen-induced yield variation in field-grown cotton. *Field Crops Res.* 299, 108976. doi: 10.1016/j.fcr.2023.108976

R Core Team (2023). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing).

Ritchie, G. L., Bednarz, C. W., Jost, P. H., and Brown, S. M. (2007). Cotton growth and development. In *Bulletin 1952* (Athens, GA: University of Georgia).

Rodriguez-Sanchez, J., Johnsen, K., and Li, C. (2024). A ground mobile robot for autonomous terrestrial laser scanning-based field phenotyping. *ArXiv*. doi: 10.48550/arXiv.2404.04404

Rodriguez-Sanchez, J., and Li, C. (2022). An autonomous ground system for 3D LiDAR-based crop scouting. In *2022 Houston, TX July 17-20, 2022* (St. Joseph, MI: American Society of Agricultural and Biological Engineers), 1–10. doi: 10.13031/aim.202200142

Rohlf, F. J., and Slice, D. (1990). Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Zoology* 39, 40. doi: 10.2307/2992207

Rote, G. (1991). Computing the minimum Hausdorff distance between two point sets on a line under translation. *Inf. Process. Lett.* 38, 123–127. doi: 10.1016/0020-0190(91)90233-8

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat* 6, 461–464. doi: 10.1214/aos/1176344136

Seedorff, M., Oleson, J., and McMurray, B. (2019). Maybe maximal: Good enough mixed models optimize power while controlling type I error. *PsyArXiv*. doi: 10.31234/osf.io/xmhfr

Singh, A., Ganapathysubramanian, B., Singh, A. K., and Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* 21, 110–124. doi: 10.1016/j.tplants.2015.10.015

Sneath, P. H. A. (1967). Trend-surface analysis of transformation grids. *J. Zoology* 151, 65–122. doi: 10.1111/j.1469-7998.1967.tb02866.x

Snider, J., Bange, M., and Heitholt, J. (2021). Cotton. In *Crop physiology case histories for major crops* (Cambridge, MA: Elsevier), 714–746. doi: 10.1016/b978-0-12-819194-1.00022-0

Su, Y., Wu, F., Ao, Z., Jin, S., Qin, F., Liu, B., et al. (2019). Evaluating maize phenotype dynamics under drought stress using terrestrial LiDAR. *Plant Methods* 15 (11). doi: 10.1186/s13007-019-0396-x

Sun, S., Li, C., Chee, P. W., Paterson, A. H., Meng, C., Zhang, J., et al. (2021). High resolution 3D terrestrial LiDAR for cotton plant main stalk and node detection. *Comput. Electron. Agric.* 187, 106276. doi: 10.1016/j.compag.2021.106276

Sun, S., Li, C., and Paterson, A. (2017). In-field high-throughput phenotyping of cotton plant height using LiDAR. *Remote Sens.* 9, 377. doi: 10.3390/rs9040377

Sun, S., Li, C., Paterson, A. H., Jiang, Y., Xu, R., Robertson, J. S., et al. (2018). In-field high throughput phenotyping and cotton plant growth analysis using LiDAR. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00016

Tao, H., Xu, S., Tian, Y., Li, Z., Ge, Y., Zhang, J., et al. (2022). Proximal and remote sensing in plant phenomics: 20 years of progress, challenges, and perspectives. *Plant Commun.* 3, 100344. doi: 10.1016/j.xplc.2022.100344

Tardieu, F., Cabrera-Bosquet, L., Pridmore, T., and Bennett, M. (2017). Plant phenomics, from sensors to knowledge. *Curr. Biol.* 27, R770–R783. doi: 10.1016/j.cub.2017.05.055

Tilly, N., Hoffmeister, D., Cao, Q., Huang, S., Lenz-Wiedemann, V., Miao, Y., et al. (2014). Multitemporal crop surface models: accurate plant height measurement and biomass estimation with terrestrial laser scanning in paddy rice. *J. Appl. Remote Sens.* 8, 83671. doi: 10.1117/1.JRS.8.083671

Tjørve, E., and Tjørve, K. M. (2010). A unified approach to the Richards-model family for use in growth analyses: Why we need only two model forms. *J. Theor. Biol.* 267, 417–425. doi: 10.1016/j.jtbi.2010.09.008

Voss-Fels, K. P., Stahl, A., and Hickey, L. T. (2019). Q&A: Modern crop breeding for future food security. *BMC Biol.* 17, 1–7. doi: 10.1186/s12915-019-0638-4

Wang, X., Singh, D. S. K., Marla, S. R., Morris, G. P., and Poland, J. A. (2018). Field-based high-throughput phenotyping of plant height in sorghum using different sensing technologies. *Plant Methods* 14 (53). doi: 10.1186/s13007-018-0324-5

Xavier, A., Hall, B., Hearst, A. A., Cherkauer, K. A., and Rainey, K. M. (2017). Genetic architecture of phenomic-enabled canopy coverage in glycine max. *Genetics* 206, 1081–1089. doi: 10.1534/genetics.116.198713

Zhang, X., Pourreza, A., Cheung, K. H., Zuniga-Ramirez, G., Lampinen, B. D., and Shackel, K. A. (2021). Estimation of fractional photosynthetically active radiation from a canopy 3D model; case study: Almond yield prediction. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.715361

Zhou, X., Wang, P., Tansey, K., Zhang, S., Li, H., and Tian, H. (2020). Reconstruction of time series leaf area index for improving wheat yield estimates at field scales by fusion of sentinel-2, -3 and MODIS imagery. *Comput. Electron. Agric.* 177, 105692. doi: 10.1016/j.compag.2020.105692

frontiers | Frontiers in Plant Science

TYPE Systematic Review
PUBLISHED 09 August 2024
DOI 10.3389/fpls.2024.1401246

Check for updates

OPEN ACCESS

EDITED BY
Parvathaneni Naga Srinivasu,
Amrita Vishwa Vidyapeetham University, India

REVIEWED BY
Yang Liu,
Shihezi University, China
Tariq Hussain,
Zhejiang Gongshang University, China

*CORRESPONDENCE
Yuhong Liu
yhliu@hhu.edu.cn

RECEIVED 15 March 2024
ACCEPTED 15 July 2024
PUBLISHED 09 August 2024

CITATION
Mustafa G, Liu Y, Khan IH, Hussain S, Jiang Y, Liu J, Arshad S and Osman R (2024) Establishing a knowledge structure for yield prediction in cereal crops using unmanned aerial vehicles. *Front. Plant Sci.* 15:1401246. doi: 10.3389/fpls.2024.1401246

COPYRIGHT
© 2024 Mustafa, Liu, Khan, Hussain, Jiang, Liu, Arshad and Osman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Establishing a knowledge structure for yield prediction in cereal crops using unmanned aerial vehicles

Ghulam Mustafa[1,2], Yuhong Liu[1*], Imran Haider Khan[2], Sarfraz Hussain[3], Yuhan Jiang[1], Jiayuan Liu[1], Saeed Arshad[2] and Raheel Osman[4]

[1]Key Laboratory of Integrated Regulation and Resource Development on Shallow Lakes, Ministry of Education, College of Environment, Hohai University, Nanjing, China, [2]College of Agriculture, Nanjing Agricultural University, Nanjing, China, [3]College of Physics and Optoelectronic Engineering, Shenzhen University, Shenzhen, China, [4]Department of Agronomy, Iowa State University, Ames, IA, United States

Recently, a rapid advancement in using unmanned aerial vehicles (UAVs) for yield prediction (YP) has led to many YP research findings. This study aims to visualize the intellectual background, research progress, knowledge structure, and main research frontiers of the entire YP domain for main cereal crops using VOSviewer and a comprehensive literature review. To develop visualization networks of UAVs related knowledge for YP of wheat, maize, rice, and soybean (WMRS) crops, the original research articles published between January 2001 and August 2023 were retrieved from the web of science core collection (WOSCC) database. Significant contributors have been observed to the growth of YP-related research, including the most active countries, prolific publications, productive writers and authors, the top contributing institutions, influential journals, papers, and keywords. Furthermore, the study observed the primary contributions of YP for WMRS crops using UAVs at the micro, meso, and macro levels and the degree of collaboration and information sources for YP. Moreover, the policy assistance from the People's Republic of China, the United States of America, Germany, and Australia considerably advances the knowledge of UAVs connected to YP of WMRS crops, revealed under investigation of grants and collaborating nations. Lastly, the findings of WMRS crops for YP are presented regarding the data type, algorithms, results, and study location. The remote sensing community can significantly benefit from this study by being able to discriminate between the most critical sub-domains of the YP literature for WMRS crops utilizing UAVs and to recommend new research frontiers for concentrating on the essential directions for subsequent studies.

KEYWORDS

yield prediction, unmanned aerial vehicles, VOSviewer, wheat, maize, rice, soybean heading 3, left

1 Introduction

Wheat, maize, and rice were responsible for 30% of the world's crop production by the year 2020. Furthermore, when it came to the exports of the commodities that were produced from cereal crops, wheat, maize, and rice contributed 45%, 32%, and 9% respectively. Simultaneously, soybean contributed 28% of oil production (FAO, 2022). These crops are of keen importance to address the food security issue. Hence, the global research developments can be examined to update and for self-evaluation in fulfilling the current food demands. By delving deeply into a particular study's academic history and knowledge structure, it is possible to identify the research themes, knowledge base, research frontiers, and hotspots worldwide. The extensive background of scientific investigations can be categorized using a range of distinct specialties, including institutions, collaborating authors, countries, co-occurring keywords, publications, hot research topics, cited references, and knowledge clusters. Leveraging the co-occurring keywords in articles from databases to conduct a literature clustering helps to reveal the knowledge structure and domains (Qian et al., 2019; Chen and Liu, 2020). These maps can be represented as networks to create a structure of these analyses known as bibliometric (or scientormetric) analysis (Ahmed et al., 2021; Azam et al., 2021; Hussain et al., 2023). Consequently, this review systematically interprets the intellectual history of a research topic or scientific literature through this novel approach. Moreover, it provides insight into new breakthroughs relevant for researchers, business investors, engineers, and the author or institutional collaboration knowledge structures and emerging research topics. This review article focusses on the use of UAVs in yield prediction (YP), and particularly with applications to wheat, maize, rice and soybean (WMRS) cropping systems: given its importance for food security at global scale.

Numerous approaches are being made for WMRS's sustainable production, protection, monitoring, and estimation, and one use is UAVs. YP analysis has been increasingly popular as big data technologies have advanced (Barnetson et al., 2020). Consulting firms and policymakers frequently employ such models when developing effective strategies (Astor et al., 2020). One of the most critical challenges in agriculture is grain yield regarding personal living standards and food security. The quantity of crop produced in a certain year is referred as the yield (Fischer et al., 2014). Several factors affect here, like genetic features, soil, weather, cultivation, and a wide range of varietals crop yields. Remote sensing (RS), a relatively new technique is expected to help calculate rice yield, especially on a regional scale. With the emergence of UAVs, a unique strategy for RS has been provided, and high spatiotemporal resolution imaging on a regional scale is now possible (Zheng et al., 2019). However, it wasn't till 2000s that the UAV technology took off in the agricultural sector, thanks to ground-breaking innovations that made the technology more accessible, user-friendly, and affordable. Examples of agricultural-grade UAVs with sensors available in 2019 include the Matrice 100 (Dà-Jiang Innovations, Shenzhen, China) and MicaSense Red Edge MX (Seattle, Washington) as well as the Pix4D Fields (Lausanne, Switzerland) and hardware from Pix4D (Puget Systems, 2021) could be purchased for less than $30,000. In addition, using UAVs' sensors has facilitated the collection of visual data, which can now be analyzed through cloud-based computing services. This advancement has decreased the requirement for costly image processing software and hardware that formerly had to be present on-site. Notably, Pix4D Fields, a software developed by Pix4D in Lausanne, Switzerland, is an example of such cloud-based computing services. Typically, the acquisition of UAVs-based data involves several stages: the formulation of a mission plan, the collection of imagery during the operation, the subsequent stitching, processing, and extraction of the data, and ultimately, the uploading of the resulting data output to precision agriculture machines or for statistical analysis in the context of research. Figure 1 demonstrates a conventional method for acquiring quantifiable and actionable data from multispectral and hyperspectral UAVs.

Additionally, satellites, manned airplanes, and handheld sensors can all be used for RS (Melesse et al., 2007). However, satellites' drawbacks include lesser resolutions, cloud blockage, scheduling, and location issues. Manned aircraft can cover large areas, but the expense is overpriced. While handheld sensors are highly accurate, they have far smaller coverage areas than aerial RS. UAVs are appealing for agricultural applications due to their ability to efficiently cover large areas, unaffected by cloud interference, unrestricted by location or timing constraints, and reasonably cost-effective (Muchiri and Kimathi, 2022). However, some disadvantages of UAVs-based RS are data degradation due to lighting conditions, data collection's tendency to occur close to solar noon, airspace restrictions, and poor weather grounding. Regulations regarding the usage of UAVs vary greatly throughout countries and often encompass limitations on velocity and elevation, nocturnal operations, visual range, proximity to airports, and densely populated regions. Additional issues to take into account during secure flight operations include the presence of other manned aircraft, prey birds, and disruptions in controller communication that could result in out of control.

Crop improvement through genetics and plant breeding and the use of precision technologies such as UAVs and remote sensors are potential solutions to meet this demand (Yang et al., 2017). Such technologies are vital for strategic management and can lead to specific breeding decisions and ensure the maximum agricultural outputs (Tilman et al., 2011). The technologies like UAVs in RS, enable rapid collection of phenotypic data with an efficient and non-destructive way for agronomists and plant breeders (Yang et al., 2017). Examples include estimating leaf color (Graham et al., 2009), lodging (Zhao et al., 2019), plant height (Han et al., 2018), stand count (Zhao et al., 2018), canopy cover (Lee and Lee, 2011), fruit count (Dorj et al., 2017) and flower count (Adamsen et al., 2000). The spectral sensors can be used to calculate yield (Shanahan et al., 2001), leaf area index (Boegh et al., 2013), leaf chlorophyll content, indirect leaf nitrogen content, and plant biomass. Finally, thermal sensors gather information to calculate canopy temperature, stomatal conductance, plant water potential and water use efficiency (Santesteban et al., 2017).

Phenotyping estimations rely significantly on the duration of sensing, and as crops reach maturity, phenotyping assessments

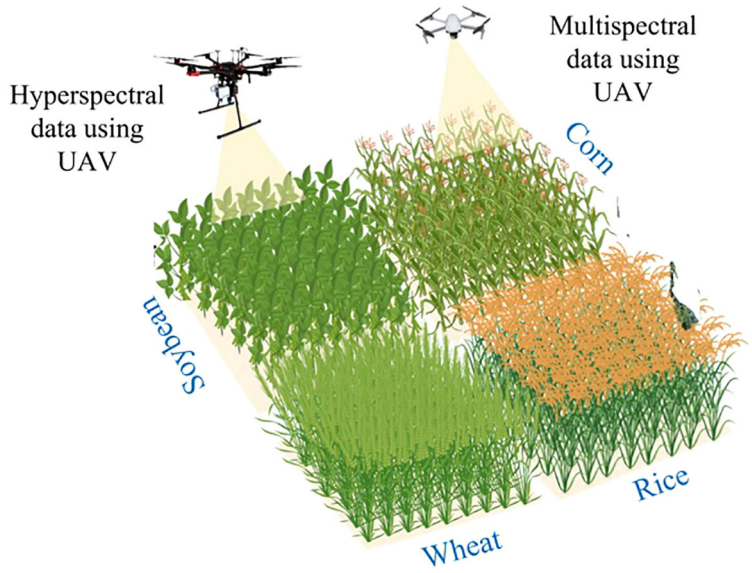


FIGURE 1
Pictorial view of the application of the UAV for crop yield prediction.

improve in accuracy. In most cases, the difference between the terrain model and the surface model must be calculated to predict crop height; this yields the so-called digital crop model (Han et al., 2018). Low-resolution cameras (3 megapixels) cannot catch the fine details of complex crop surfaces, hence only cameras with superior resolution (>10 megapixels) should be utilized. Additionally, it has been demonstrated that the "Scale Constraint" function of some imaging processing software and ground control points improves crop height estimates. Refraining from conducting sensing activities on days with wind speeds over 1-10 kilometers per hour is advisable (Sziroczak et al., 2022). This precaution is recommended due to the propensity of wind to induce movement in plants and result in image blur, hence diminishing the precision of crop height estimation.

1.1 Background for yield assessment

Numerous parties rely on crop output and quality estimates, including consultants, producers, academics, insurance agents, commodities merchants, governments, and non-governmental organizations (Rembold et al., 2013). This data helps to determine how much crop insurance to purchase, how much to deliver, when to harvest to maximize quality, how much space is needed for storage, and how much money will be needed. Traditionally, managers have relied on historical yield data and seasonal variables to inform their yield and quality assessments for the remaining season (Raun et al., 2001). In contrast, the ultimate outcomes of crop production and the resulting quality are characterized by a lack of predictability and often subject to factors such as crop genetics, weather conditions, soil composition, proficiency, and choices made in crop management. The assessment of crop production and quality through the utilization of UAVs technologies commonly depends on data acquired from color and spectral sensors. The utilization of UAVs imagery, particularly in conjunction with machine learning (ML) techniques, currently holds the capacity to enhance assessment precision and potentially diminish or eradicate the need for terrestrial surveys. The accurate estimation of crop yield and quality is contingent upon the ability to effectively sense time, as the accuracy of estimation tends to improve as the crop progresses through its life cycle (Ballester et al., 2017; Zhou et al., 2017).

The green normalized difference vegetation index (GNDVI) yielded more precise biomass estimations at the stages of anthesis and full crop development than at earlier stages of development (Ostos-Garrido et al., 2019). According to other studies, it has been observed that the GNDVI yielded more accurate estimations of crop production at the early stage (5 weeks) of a crop's growth cycle (Wahab et al., 2018). Consequently, great care must be taken in the selection of factors (color component and saturation indices), as this choice greatly influences the efficacy and precision of crop yield and quality estimates using spectral reflectance obtained at peak sites.

However, environmental factors and crop water stress are likely to hinder the performance reliability of spectral indices where NDVI (Normalized Difference Vegetation Index) and RENDVI (Red Edge Normalized Vegetation Difference Index), based on physiological behavior measured in lab conditions rather than actual meteorological variations under field environments for predicting wheat production accurately (Hassan et al., 2019). Consequently, forthcoming investigations will examine these elements within stress-controlled experiments to cultivate further understanding. Incorporating correction factors for quality and

yield assessments may be seen as a potential approach to enhance the accuracy of spectrum consumption measurements in indices. The application of ML, namely artificial neural networks, to spectral bands has demonstrated potential (Sarkar et al., 2018). For instance, these protocols have successfully predicted the grain protein content of rice (*Oryza sativa*) and the total soluble solids content of grapes (*Vitis*). Relatively inexpensive color cameras have been utilized to capture color characteristics, proving to be a dependable method for generating accurate yield estimations (Kefauver et al., 2017). This approach offers a feasible alternative to utilizing more expensive sensors. However, previous investigations that have employed color cameras have yielded unsatisfactory outcomes (Bura et al., 2018). This could be attributed to the inadequate development of yield estimation protocols, or the studies conducted under unfavorable environmental conditions (Bura et al., 2018). In conclusion, numerous studies have been undertaken to estimate crop production (Zhou et al., 2017; Bura et al., 2018; Hassan et al., 2019; Zheng et al., 2019). However, numerous studies have shown the concerns regarding accuracy, reliability, and scalability for the analysis of the UAVs data for different crop phonemics based studies.

1.1.1 Accuracy

1.1.1.1 Sensor restrictions

The UAVs usually have thermal, multispectral, or hyperspectral sensors (Fei et al., 2023). Even though these sensors can record much information about crops, they might not be precise enough for analyzing individual cereal crops, especially in environments with varying lighting and weather conditions (Lee et al., 2010).

1.1.1.2 Weather and flight stability

Cloud cover, wind, and fluctuating sunlight are external elements that might add inaccuracies into sensor data (Aasen et al., 2018), impacting YP's accuracy.

1.1.1.3 Data quality

UAVs gather information at different speeds and altitudes, which could cause inconsistencies in the resolution and quality of the data collected from the sensors (Pádua et al., 2017). YP and analyses that rely on this ambiguity may be imperfect.

1.1.2 Reliability

UAVs express difficulty flying in unstable air and are vulnerable to bad weather like wind, rain, and fog, which can delay data collection and even equipment damage (Ubina and Cheng, 2022). This unpredictability may make YP models less reliable (Müller et al., 2016).

1.1.2.1 Data processing and integration

UAVs produce massive amounts of data that must be processed correctly to conduct real-time analyses (Guimarães et al., 2020). Because of inherent variances in data structure and format, integrating UAV data with other data sources, including weather, soil, and ground truth measures is not easy (Samaras et al., 2019).

1.1.3 Scalability

1.1.3.1 Field coverage

UAVs have trouble collecting data at scale due to their short flight times and short battery lives, which limits their capacity to survey vast cereal crop fields in a single flight (Mohsan et al., 2023).

1.1.3.2 Data storage, integration and management

For analyses that span several growing seasons or include long-term monitoring, it is necessary to have a robust infrastructure to store and manage massive datasets acquired by UAVs (Nabwire et al., 2021). When applied to new locations with varied soil types, temperatures, and crop types, ML models built on data from one field or area might not perform as well as when trained on data from another (Benos et al., 2021). Hence, it limits the models' ability to be applied at various agricultural scales.

1.1.4 Addressing challenges

Diverse methods can be considered to overcome these previously mentioned difficulties.

1.1.4.1 Sensor calibration and data fusion

The application of advanced filtering techniques and regularly calibrating sensors can help to increase the quality of data (Concas et al., 2021). A more complete picture of crop conditions and improved yield projections can be achieved through data fusion and integration (Ahmad et al., 2022), which involves combining data from UAV with data from other sources (such as satellite imaging and ground-based sensors).

1.1.4.2 Hybrid machine learning models

By integrating ML with more conventional statistical approaches, scientists can create more robust models and generalize to new datasets (Elavarasan et al., 2018).

1.1.4.3 Improved technology for UAV

Longer flight times and more automated flight controls are just two examples of how UAV technology is constantly evolving to better cover more ground and collect more data (Chaurasia and Mohindru, 2021).

1.1.4.4 Data management solutions with cloud computing

Using the cloud to store and analyses data can solve scalability problems and make data analysis more effective (Delgado et al., 2019). The background reveals proof of adaptation of the multiple approaches for YP in different cereal crops.

This study employs UAVs for YP and provides a bibliometric analysis of existing research conducted at WMRS crops. The WOSCC extracted pertinent scientific information published between 2005 and 2023 in high caliber publications. VOSviewer was used for co-occurrence, co-citation, and co-authorship analyses. The following are the goals of the current study: (a) To establish a knowledge structure for WMRS crops using UAVs for the YP (b) To evaluate the use of ML, different sensors and limitations for WMRS crops. (c) To identify the knowledge gaps

and research frontiers for YP in WMRS crops employing UAVs. The article's structure includes an introduction to WMRS crops and their importance, publication collection, a description of the results of the bibliometric analysis, review on WMRS crops, constraints and prospects for future study, and contribution of the current study for the scientific community.

2 Retrieval of publications' information and outline for bibliometric methods

2.1 Publications' information retrieval from web of science

The most reliable literature-indexing site is the Web of Science (WOS), which includes scientific, social, health, and economic information. Therefore, it is acknowledged that the global WOS is the ideal source for gathering data for bibliometric analysis (Chen and Liu, 2020). The pertinent information was gathered from the WOS core collection (WOSCC) databases. Many iterations were employed to find the best keyword search code to download the most pertinent RS-related publications for YP in WMRS crops utilizing UAVs. Table 1A shows a list of more specific keyword codes that can be used to probe the WOS database systematically. The most useful search keywords were as follows: ("UAV" OR "multispectral" OR "hyperspectral" OR "RGB") (Topic) and ("yield prediction" OR "yield estimation") (Topic) and ("maize" OR "rice" OR "wheat" OR "soybean"). Be noted that the published articles were examined for the presence of words within their titles, abstracts, or keywords. All searched documents were validated for relevant material, and manually irrelevant articles were removed. The extraction process involved selecting solely peer-reviewed, original research publications that were published in the English language. The data acquiring period spanned from January 1, 2000, to August 3, 2023 (Table 1B), and the relevant scholarly material and the scope of inquiry were restricted to science and technology.

2.2 Work flow for the study

An overall number of 226 original research publications were retrieved using the methods described above. The entire document and all referenced materials were saved as "Tab-delimited" as the preferred file type. The schematic shows the actions performed to carry out the investigation in Figure 2.

2.3 VOSviewer based bibliometric analysis

A map or network analysis and the visualization of scientific literature are examples of more advanced bibliometric analysis skills. Nees Jan van Eck and Ludo Waltman developed a robust program called VOSviewer (Version 1.6.18) for investigating and visualizing bibliometric networks (Van Eck and Waltman, 2010). It makes it possible to do analyses such as co-citation analysis, cluster analysis, and bibliometric mapping, all of which highlight the current research collaborations and patterns. The scientific community relies heavily on VOSviewer to visualize and project their data. The scientific community can learn more and get the program at http://www.vosviewer.com.

TABLE 1 Selection of optimized keywords for WOS publications' information, and VOSviewer values and parameters for bibliometric examination for yield prediction in WMRS crops utilizing UAVs.

(A) Selection of optimized keywords for WOS publications' information			
No.	Searching code	Results	Quality
1	("hyperspectral" OR "reflectance" AND "wheat" AND "yield prediction")	50,159	very generic, very rough, highly irrelevant
2	("UAV" AND "wheat" AND "yield prediction" OR "yield estimation")	3,478	Very rough, highly irrelevant
3	("UAV" AND "wheat" AND "yield prediction")	75	yet irrelevant, Improved,
4	("UAV" AND "wheat" AND "yield prediction" OR "yield estimation") (Topic) and ("UAV" AND "wheat" AND "yield prediction" OR "yield estimation")	448	Very generic and moderately irrelevant
5	("UAV" AND "wheat" AND "yield prediction" OR "yield estimation") (Topic)	1907	Improved, yet irrelevant
6	("UAV" OR "multispectral" OR "hyperspectral" OR "RGB") (Topic) and ("yield prediction" OR "yield estimation") (Topic) and ("maize" OR "rice" OR "wheat" OR "soybean")	226	Highly improved, fully relevant
(B) VOSviewer values and parameters for bibliometric examination of advanced research			
No.	Parameters	Definition	
1	Time slicing	2001-01-01 to 2023-08-03	
2	Term source	Abstract, keywords, title, author, and keywords plus	
3	Node type	Country, institution, cited author, Author, cited journal & keywords cited reference.	
4	Selection criteria	Top 15%	
5	Links	Default	
6	Visualization	Cluster static and combined network views are displayed	

As per the specified parameters, utilizing VOSviewer facilitated the execution of co-citation analysis and keywords co-occurrence analysis. These analyses generated networks visually representing the co-citations seen among authors, documents, journals, and keywords. Furthermore, these networks provided insights into prominent study themes and emerging areas of investigation. In conclusion, the relevant data and mapping networks were thoroughly analyzed, and the results of the visualization inquiry in the present research study were presented and discussed.

3 Interpretation and discussion of the bibliometric results

3.1 Publications' information analyses retrieved from web of science

The distribution of citations and matching publications in the domain of YP using RS in WMRS crops utilizing UAVs throughout the years 2005-01-01 to 2023-08-03 is shown in Figure 3A. It can be seen that over the first ten years (from 2005 to 2011), there was a prominently gradual increase in the number of notable publications. In 2012-2014 there was a decrease in publications. However, from 2014 to onward, there is a continuous and gradual increase in publications. In total, 44 research were published in 2022, which is the highest. Of the whole, 60.25% of the total publications are published from 2019-2022, 19.89% only in 2022, while 2023 has contributed 10.52% so far. The concerns about the research dramatically increased starting from 2019 to onward. This can be related to shifting funding preferences or research priorities during that time. The number of citations for publications remains gradually growing, while it has shown a fall in 2023, which might be due to the incompletion of this year. The highest citations (2291) have been reported in 2022, reflecting the impact of the research in this field. This upsurge could be attributed to developments in ML and UAV technology, which facilitate the conduct of creative and significant research.

The WOS based subject category area (Figure 3B) shows the publication distribution in the top 10 scientific study areas. Where "Remote Sensing" is leading than all others, having 39.04% (89 Articles) of total articles. While Agriculture, Imaging Science, Photographic Technology, Geosciences Multidisciplinary, Environmental Sciences, Agronomy and Plant Sciences share 89, 66, 60, 55, 43 and 38 publications, respectively. This diversity draws attention to the research's interdisciplinary nature. The more influenced 15 authors for the YP in WMRS crops utilizing UAVs are ranked in Figure 3C where Yang GJ ranked at top with the maximum number of records in the acquired publications' dataset. Given their prominence, it appears that they have given this field of study a lot of attention in their work.

Table 2A lists the top 15 journals in which the most articles relevant to the research area of YP in WMRS crops utilizing

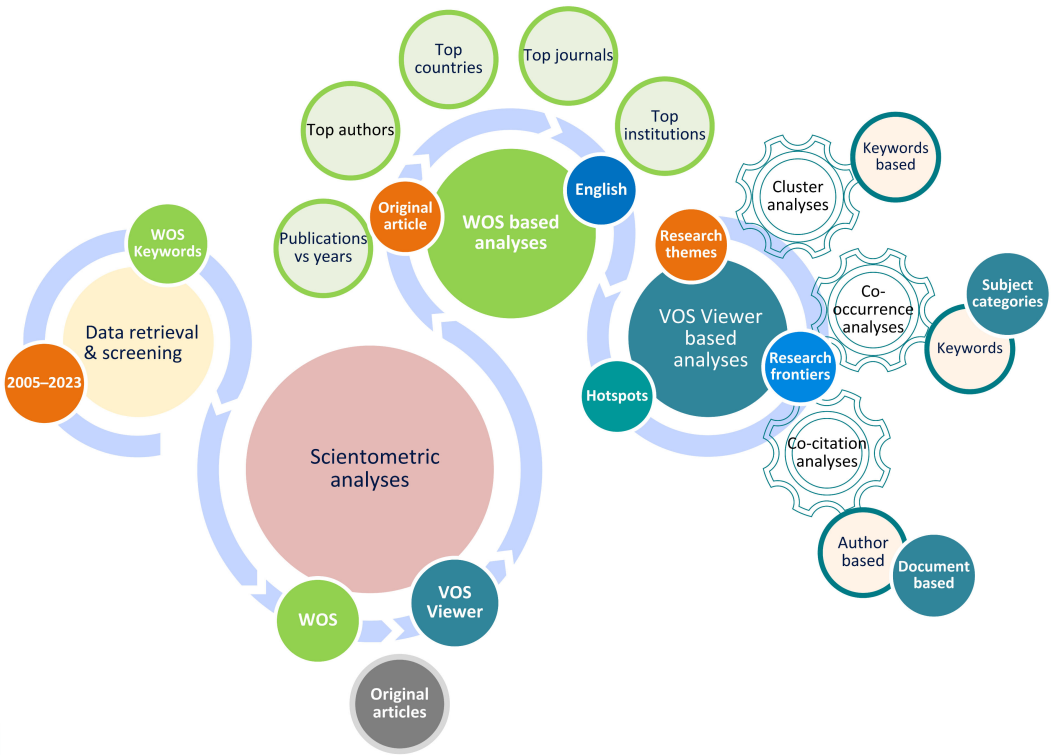


FIGURE 2
Outline for the data retrieval and VOS-viewer based bibliometric analysis.

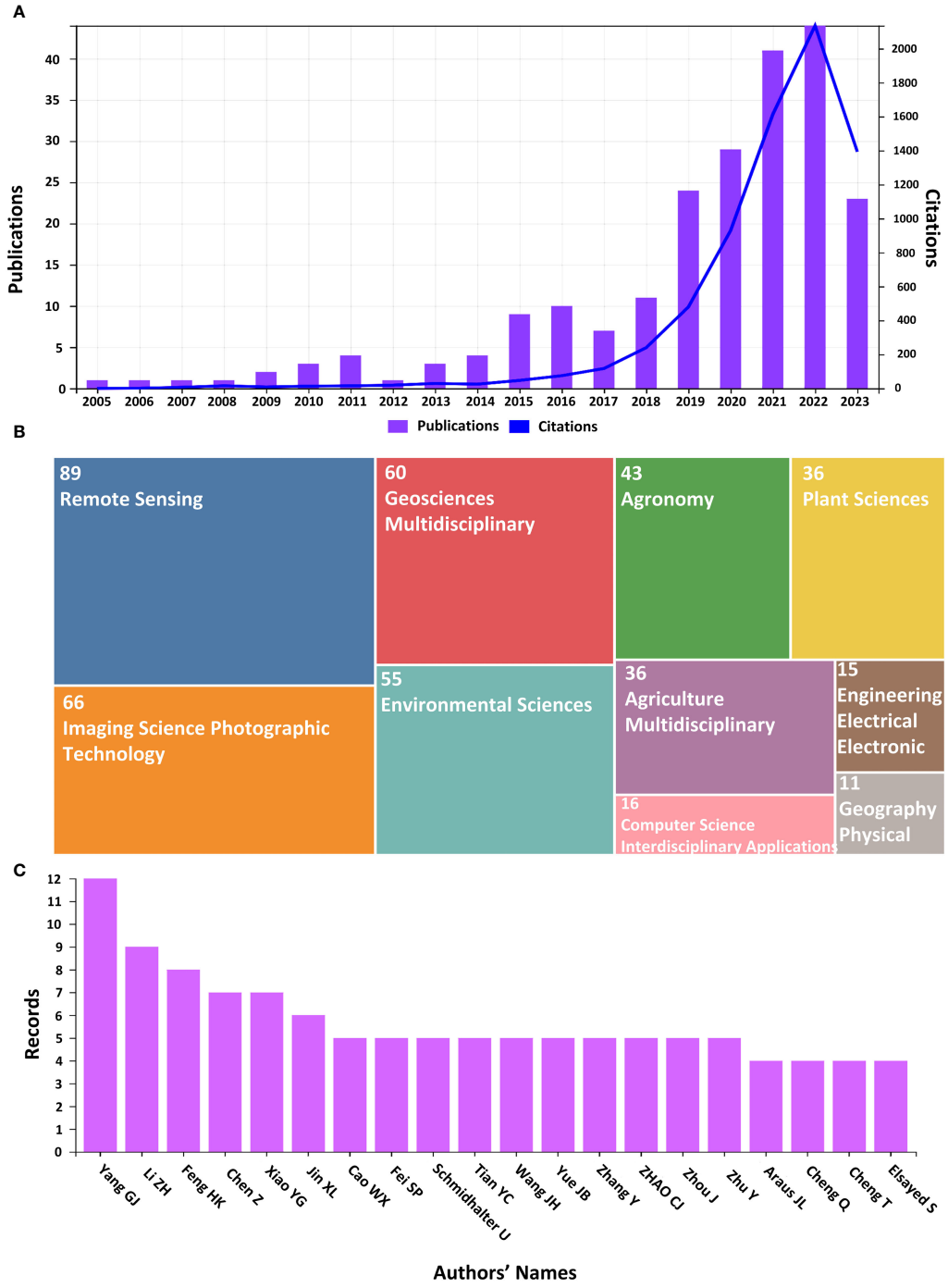


FIGURE 3
(A) Illustrates the number of publications and citations, **(B)** Number of publications according to subject categories, **(C)** Number of records for different authors.

UAVs are published, along with the number of publications and their percentage contribution. Table 2B lists the top 15 research institutions for YP in WMRS crops utilizing UAVs. Supplementary Tables S1, S2 indicate the 15 most important countries and prolific writers with the most studies in the specified research subject, respectively. Supplementary Table S3 shows the distribution of the top fifteen funding agencies involved in relevant publications collected from the WOS in YP research utilizing UAVs. This systematic methodology provides insightful information and enables researchers to narrow down areas of potential studies for future.

3.2 Citations' analyses

When a third author or document cites two or more authors or documents simultaneously, this is called a co-citation link

10.3389/fpls.2024.1401246

TABLE 2A The 15 best journals for publishing research on yield prediction in WMRS crops utilizing UAVs.

No.	Journals	Records	% of total
1	REMOTE SENSING	46	20.354
2	FRONTIERS IN PLANT SCIENCE	15	6.637
3	COMPUTERS AND ELECTRONICS IN AGRICULTURE	13	5.752
4	PRECISION AGRICULTURE	11	4.867
5	IEEE INTERNATIONAL SYMPOSIUM ON GEOSCIENCE AND REMOTE SENSING IGRASS	8	3.540
6	AGRICULTURE BASEL	7	3.097
7	AGRONOMY BASEL	7	3.097
8	PLANT METHODS	7	3.097
9	SENSORS	6	2.655
10	AGRICULTURAL AND FOREST METEOROLOGY	5	2.212
11	BIOSYSTEMS ENGINEERING	5	2.212
12	DRONES	5	2.212
13	PROCEEDINGS OF SPIE	5	2.212
14	FIELD CROPS RESEARCH	4	1.770
15	ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING	4	1.770

TABLE 2B The 15 best institutions for publishing research on yield prediction in WMRS crops utilizing UAVs

No.	Affiliations	Records	% of total
1	BEIJING ACADEMY OF AGRICULTURE FORESTRY SCIENCES BAAFS	19	8.407
2	MINISTRY OF AGRICULTURE RURAL AFFAIRS	19	8.407
3	CHINESE ACADEMY OF AGRICULTURAL SCIENCES	17	7.522
4	CHINA AGRICULTURE UNIVERSITY	12	5.310
5	CHINESE ACADEMY OF SCIENCES	12	5.310
6	INSTITUTE OF CROP SCIENCES CAAS	10	4.425
7	NANJING AGRICULTURAL UNIVERSITY	10	4.425
8	ZHEJIANG UNIVERSITY	9	3.982
9	UNITED STATES DEPARTMENT OF AGRICULTURE USDA	8	3.540
10	WUHAN UNIVERSITY	8	3.540
11	UNIVERSITY OF MISSOURI COLUMBIA	7	3.097
12	UNIVERSITY OF MISSOURI SYSTEM	7	2.097
13	UNIVERSITY OF NEBRASKA LINCOLN	6	2.655
14	UNIVERSITY OF NEBRASKA SYSTEM	6	2.655
15	EGYPTIAN KNOWLEDGE BANK EKB	5	2.212

(Chang et al., 2015). VOSviewer uses three main types of co-citation analyses to show how documents, writers who cite each other, and journals are connected and how they map to each other. Co-citation analysis is a good way to determine how many connections between journals, authors, and papers. It builds a mapping framework and tracks how scientific research areas change over time (Behrend and Eulerich, 2019).

3.2.1 Co-citation analysis for documents

The documents or articles are the most important parts of the knowledge repository or database in the area of predicting crop yields using UAVs for WMRS crops. Reference co-citation analysis, also called document co-citation analysis, is a good way to look at how a study area has grown and changed over time (Liao et al., 2018). After the scientometric study in VOS viewer was done, a network for displaying cited documents was made (Figure 4). Every node represents a reference or article cited, and the links between the nodes demonstrate how the mentioned references and articles relate to one another. Larger nodes represent more essential documents, and documents often referenced by other documents are close to each other. Figure 4 shows the most important works completed in this field are (Zhou et al., 2017) (Haboudane et al., 2004), and (Maimaitijiang et al., 2020).

3.2.2 Co-citation analysis for authors

The distribution of authors with more citations in a particular field of study is also examined by the co-citation analysis for the author, which is used to identify the most productive writers in that subject. The co-citation analysis also makes the visualization of associated writers' subject areas and research interests feasible. The visualization network that was produced after the author's co-citation analysis of the study into YP in WMRS crops using UAVs is shown in Figure 5. The nodes represent authors, the connecting lines between two nodes indicate their relationship regarding co-citations. An author's importance is increased when a node gets larger because that particular author makes more citations.

Similarly, the distance between two successive nodes or writers is inversely associated with the amount that one author is cited by the other. The authors' areas of interest will be more strongly correlated with the size of the distance between the nodes. Author co-authorship analysis confirms the extensive analysis of the visualization network, which shows that most authors collaborate to a very high degree. It offers vital insights into the organization

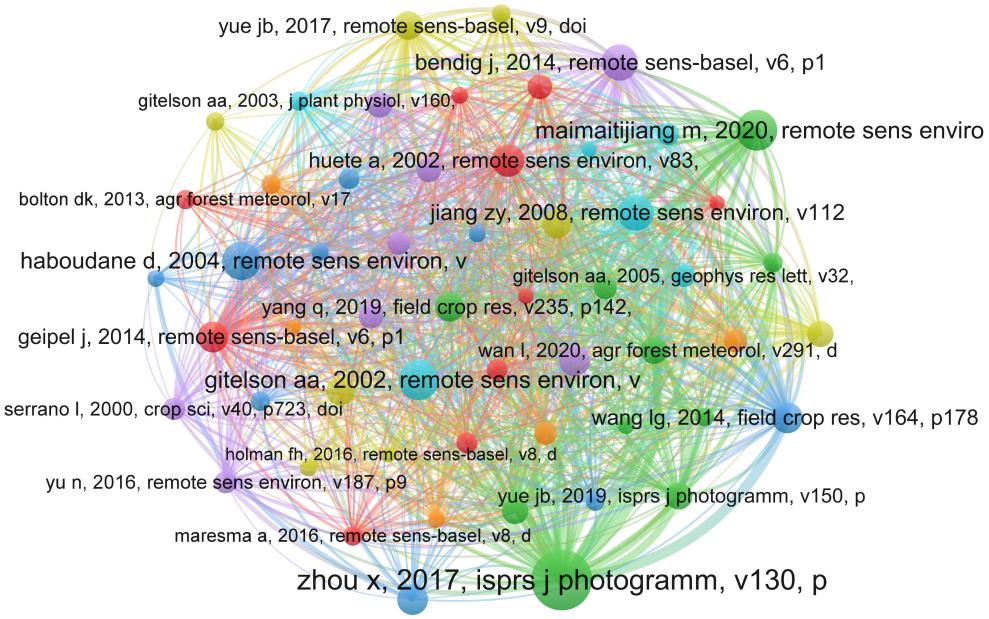


FIGURE 4
The document co-citation network visualization map for research on yield prediction in WMRS crops utilizing UAVs.

and movement of a research field, aiding researchers in comprehending significant contributors, patterns, and possibilities for collaboration.

The 15 best most co-cited authors are listed in Table 3A along with their respective authors, counts of citations, years of citation counts, and rankings based on the times their scientific literature have been cited. According to the data, the aforementioned authors' work made a significant contribution to the field of YP in WMRS crops utilizing UAVs, making them significant participators to the future growth of YP research. According to the findings, the authors (Zhou et al., 2017), (Maimaitijiang et al., 2020), and (Yue et al., 2019) were the most prolific in the research field,

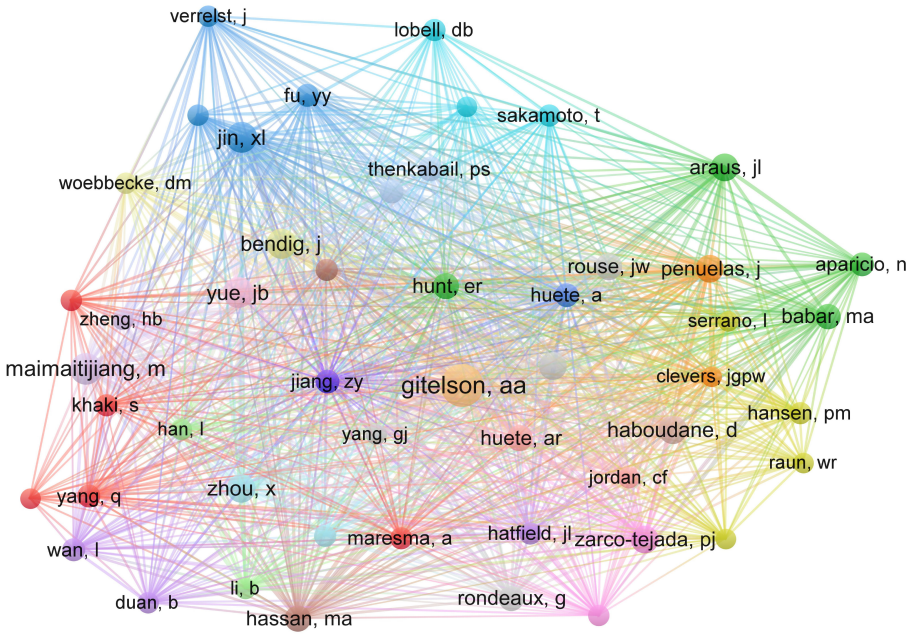


FIGURE 5
The authors' co-citation network visualization map for research on yield prediction in WMRS crops utilizing UAVs.

TABLE 3 The 15 best co-cited authors and top fifteen keywords for research on yield prediction in WMRS crops utilizing UAVs.

Table 3A The 15 best co-cited authors for publishing research			Table 3B Top fifteen keywords in the domain of yield prediction		
Sr. No.	Count	Cited Authors	Ranking	Counts	Keywords
1	68	ZHOU, X	1	110	Vegetation indices
2	63	MAIMAITIJIANG, M	2	84	UAV
3	63	YUE, JB	3	60	Wheat yield
4	59	JIN, XL	4	59	Yield estimation
5	50	BENDIG, J	5	46	Biomass estimation
6	49	HABOUDANE, D	6	44	Yield prediction
7	47	ARAUS, JL	7	43	Winter-wheat
8	46	TUCKER, CJ	8	43	Grain-yield
9	45	ZARCO-TEJADA, PJ	9	40	Remote sensing
10	45	PENUELAS, J	10	38	Leaf-area index
11	43	ROUSE, JW	11	27	Corn yield
12	41	HASSAN, MA	12	26	Machine learning
13	40	HUETE, AR	13	26	Chlorophyll content
14	39	RONDEAUX, G	14	25	Precision agriculture
15	39	BABAR, MA	15	25	Maize yield

3.2.3 Citations analysis for documents

Citation analysis for authors represents a bibliometric approach employed to assess the impact and influence of an individual author's scholarly output within the academic and relevant research field. This method entails comprehensive scrutiny of the frequency with which an author's publications have garnered citations in the works of other researchers. Through the examination of citation patterns, this analytical method provides valuable insights into an author's contributions to their field of study, the level of recognition their research receives, and the extent of their influence within the scholarly and academic community. Figure 6 visualizes the articles with more than 20 citations from different years. The size of nodes represents the highly cited documents, like (Chlingaryan et al., 2018), (Zhou et al., 2017), (Maes and Steppe, 2019), and (Shanahan et al., 2001) are most prominent with superior size of nodes.

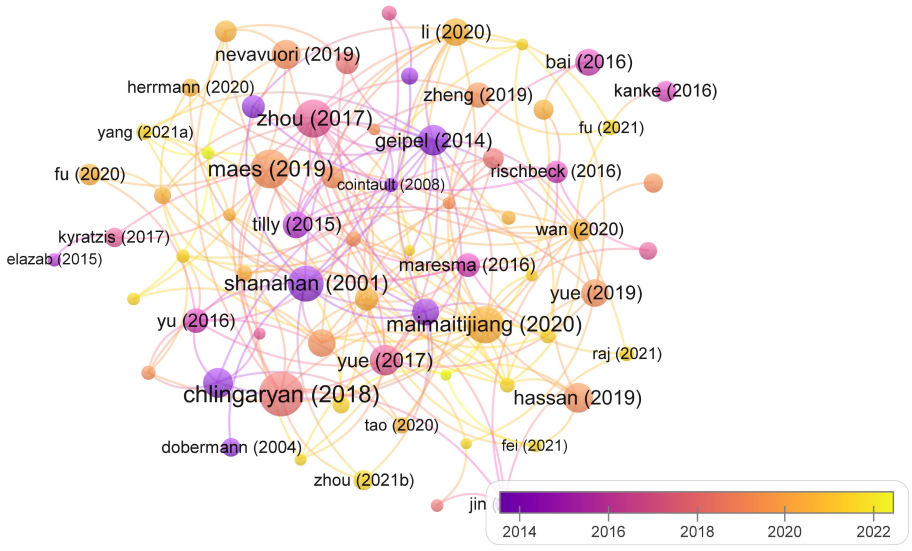


FIGURE 6
Citations' analysis of documents for yield prediction in WMRS crops utilizing UAVs.

3.3 Keywords' co-occurrence analysis

In a scholarly context, keywords serve as descriptors that elucidate the precise subject matter or overarching category to which an article pertains. Additionally, they encapsulate the primary information summarized within research papers. In essence, keyword co-occurrence analysis can be employed to discern prevailing focal points and emerging research frontiers. The most rapidly increasing citation rates for individual keywords indicate perennially popular issues or promising avenues for future study. Figure 7 is a visual representation of the VOSviewer keyword co-occurrence analysis results. Nodes represent the keywords, and the size of the nodes represents the frequency with which they occur together.

Table 3B presents a ranking of the top fifteen keywords in the domain of YP in WMRS crops utilizing UAVs. These keywords have been ordered based on their occurrence frequency. The keywords demonstrating the most pronounced co-occurrence frequencies, along with their respective counts, include Vegetation indices (110 occurrences), UAV (84 occurrences), Wheat yield (60 occurrences), and Yield estimation (59 occurrences). Citation frequency analysis offers a concise insight into the prevalence of frequently employed keywords within a specified timeframe. This analysis allows for the temporal depiction of these terms, drawing attention to the time frame in which they were employed and referenced most frequently.

3.4 Co-authorship analysis

Cooperative analysis represents a multifaceted approach that delves into various tiers of investigation, spanning from the broad macro scale to the intricate micro level. Its objective is to unveil the intricate distribution patterns characterizing the collaborative dynamics within scientific research. In this specific context, the formidable tool of choice is VOSviewer, a sophisticated visualization software adept at illuminating the intricate web of collaborative relationships among institutions, countries/regions, and authors operating in the specialized domain of YP in WMRS crops utilizing UAVs. The calculation of the depth of collaborative involvement utilizes both fractional and full counting approaches, where full counting assigns equal weight to each participant.

Within the analytical framework of literature studies, two distinctive counting methodologies come into play: fractional and full counting. In the full counting mode, each contributor is endowed with an equal weight of 1, which is symmetrically distributed in quantifying the depth of their collaborative involvement. VOSviewer, as the computational engine, undertakes the intricate task of computing scores that gauge the interrelationships among knowledge units. Employing the precision of the association strength algorithm, it harmonizes and standardizes the raw data. The result concludes in creating an enlightening visualization map of the literature, encompassing distance-based and graph-based

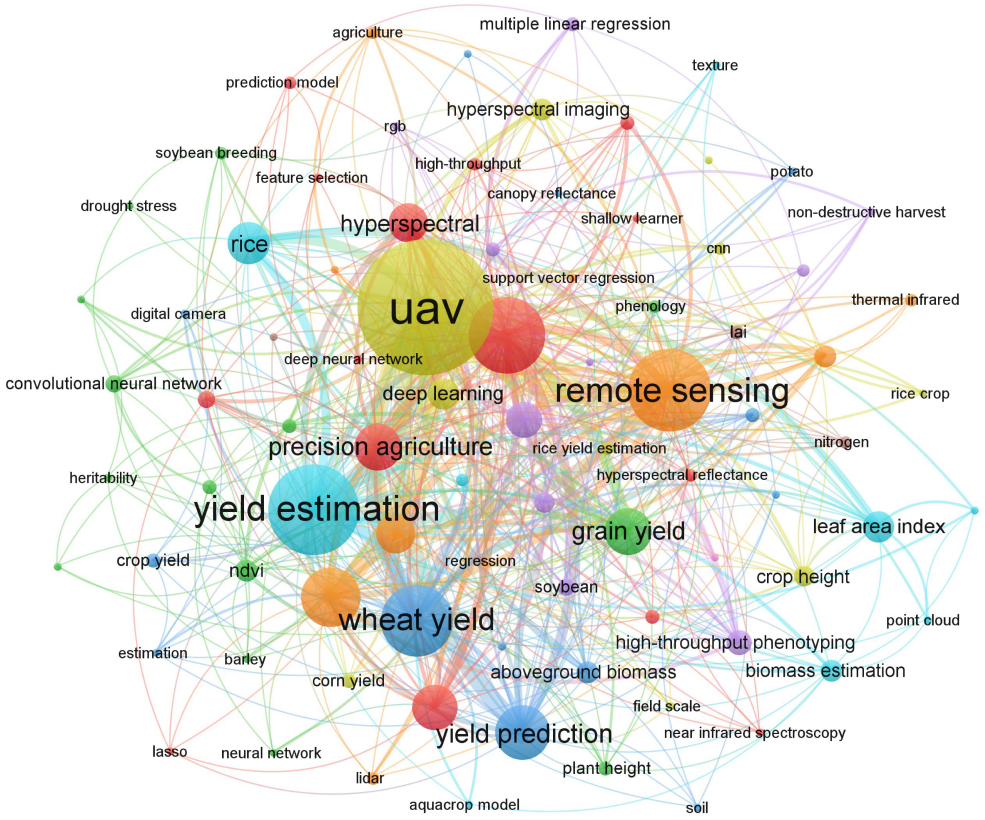


FIGURE 7
Keywords' co-occurrence analysis for yield prediction in WMRS crops utilizing UAVs.

representations, thereby offering a comprehensive insight into the complex web of scholarly interactions.

3.4.1 Co-authorship analysis for authors

The quantification of an author's productivity within a specific domain is a pivotal metric in assessing their impact within that particular field. The analysis of author collaboration serves as a valuable tool for investigating YP in WMRS crops utilizing UAVs and the associated social networks of cooperation. In Figure 8A, the study presents the author collaboration network within the domain of yield predictionYP in WMRS crops utilizing UAVs, focusing on authors who have contributed for at least two articles.

As depicted in Figure 8A, the collaboration landscape in this field exhibits a persistent spirit of cooperation and manifests a small-world effect. It is evident that each academic group maintains the capacity to engage in direct or indirect collaboration with other scientific research teams, illustrating the continuous transmission of information within the network. Notably, within specific clusters, the trajectory of YP in WMRS crops utilizing UAVs is influenced by high-impact authors, such as Yang Guijun, Li Zhenhai, Feng Haikuan, Chen Zhen, Xiao Yonggui and several others (Figure 8A).

3.4.2 Co- authorship analysis for regions/countries

Figure 8B visually represents international cooperation among countries and regions over time, incorporating dynamic elements. In this visualization, the size and color of nodes correspond to the volume of research documents produced and the average publication year within each respective country or region. Furthermore, the thickness of the connecting lines between nodes indicates the level of collaboration between these entities, with thicker links representing more significant cooperation. By analyzing Figure 8B, it is evident that China and the United States are the leading countries in deploying UAVs for YP in WMRS crops. Countries such as Germany, Australia, Spain, Japan, Saudi Arabia and Canada have also actively engaged in this field of research.

It is worth highlighting that China, owing to its remarkable research productivity, has forged robust collaborative ties with Unites States for YP in WMRS crops utilizing UAVs. Looking ahead, both the China and United States are poised to emerge as principal players in the field of RS. Progress in advancing RS for YP in these two nations is of global significance, given the far-reaching benefits it brings. The advancements and collaboration between

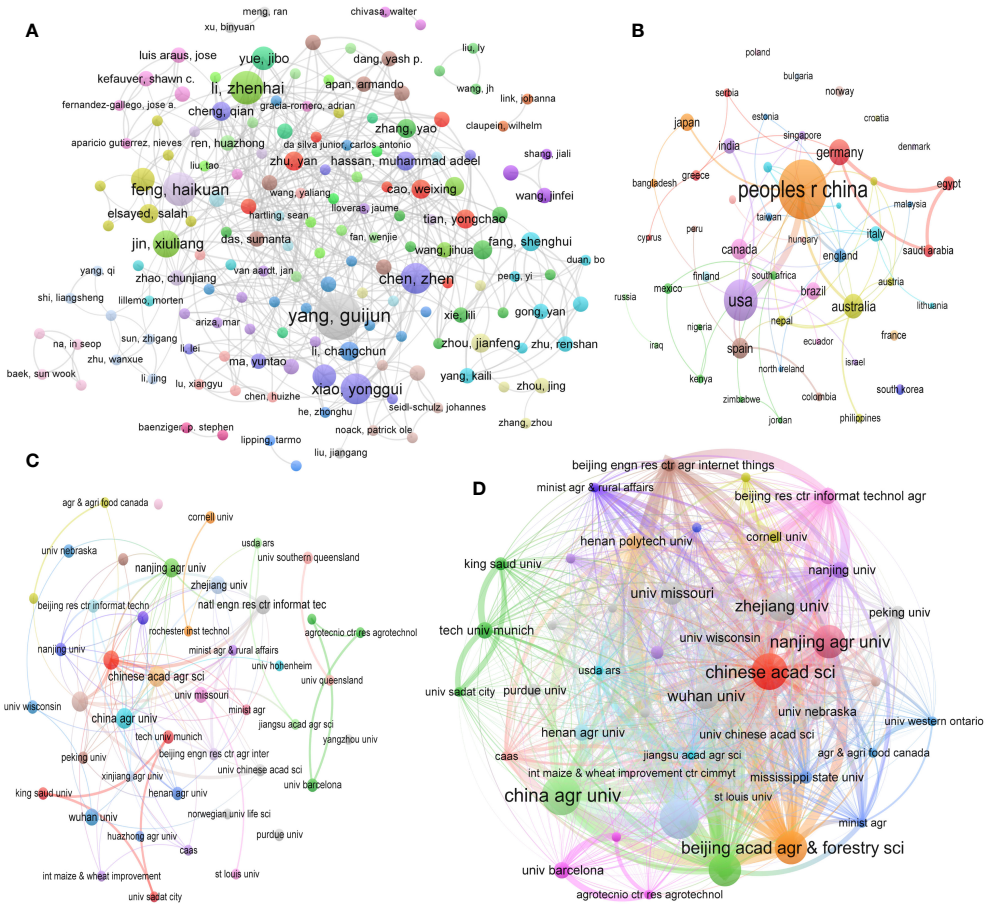


FIGURE 8
(A) Co-authorship analysis of authors, **(B)** Co-authorship analysis of regions/countries, **(C)** Co-authorship analysis of institutions, and **(D)** Coupling analysis of institutions for yield prediction in WMRS crops utilizing UAVs.

China and the United States, the two greatest developed and emerging nations in the world, significantly impact the Asia-Pacific region and the global scene. Motivated by same goals including guaranteeing food security and reducing the impact of sustainable crop production, both countries have agreed to make use of crop production innovations.

3.4.3 Co-authorship analysis for institutions

The examination of institutional cooperation yields valuable insights into organizations and groups that make significant contributions within a particular field. This analysis serves as a foundational step towards fostering enhanced future collaboration for organizations with 3 documents as co-authorship. To effectively depict the distribution of organizations and their collaborative relationships, this study employs VOSviewer, a visualization tool, to represent the network of institutional collaboration within the domain of RS for YP in WMRS crops utilizing UAVs, as presented in Figure 8C.

Each node's size in the below illustration represents the total number of scholarly articles published by that institution. The thickness of the connecting lines represents the degree of collaboration between universities. Nodes sharing the same color signify a higher level of cooperation than nodes with distinct colors. This analysis reveals notable collaborative connections among universities and colleges. For instance, the Chinese Academy of Sciences with China Agriculture University, and Nanjing Agricultural University with Zhejiang University exhibit a pronounced cooperative association, as evidenced by their placement within the red and green clusters. Additionally, in this study, we observe a strong collaborative relationship between King Saud University and the University of Sadat within the red cluster. Several other clusters have also merged around productive institutions, thereby contributing to forming a diversified and expansive cooperation network within the landscape of RS for YP in WMRS crops utilizing UAVs.

3.4.4 Coupling analysis for organizations

Organizations' coupling analysis is a sophisticated bibliometric method used to examine the interconnections and relationships between different organizations or institutions within the context of scientific research and publication activities. It primarily focuses on quantifying and understanding the collaborative patterns and knowledge exchange dynamics among these entities, often in the context of specific research fields or disciplines. The insights gained from organizational coupling analysis can inform policymakers, funding agencies, and researchers about the structure and dynamics of collaborative networks in specific research domains. This information is valuable for fostering interdisciplinary collaboration, optimizing research investments, and advancing scientific progress by promoting effective knowledge exchange among organizations. The organizations with a minimum 3 documents together are illustrated in Figure 8D, Chinese Academy, Nanjing Agricultural University, China Agricultural University, and Beijing Academy of Agriculture and Forestry Sciences are more prominent in respective clusters and have shown a prominent connection in connecting lines and node size. By highlighting collaborative networks and influencing choices on research investments and interdisciplinary collaboration, the results of this analysis can direct researchers, politicians, and funding agencies.

4 Recent studies on WMRS crops for UAV-based yield prediction

In agricultural research, precise early YP is paramount for individual farmers and the broader agricultural sector. UAVs have demonstrated commendable efficacy in enhancing YP accuracy through the utilization of various data sources (Hassan et al., 2019; Zheng et al., 2019; Kumar et al., 2023). For instance, such accuracy has been achieved by harnessing metrics such as RGB-derived plant height and canopy cover (Chu et al., 2016), VIs (Gracia-Romero et al., 2017), and multispectral imagery (Kyratzis et al., 2017; Zheng et al., 2019; Su et al., 2023). It is worth noting that the temporal dimension plays a pivotal role in optimizing YPs, with multitemporal VIs, including those accumulated throughout the crop's growing season, exhibiting superior performance compared to single measurements (Zhou et al., 2017).The investigations conducted using UAVs for YP have predominantly adjusted on experimental fields characterized by substantial variations in final yield due to factors such as nitrogen levels (Zhou et al., 2017), phosphorous concentrations (Gracia-Romero et al., 2017), or irrigation practices (Vega et al., 2015). However, it remains imperative to scrutinize the efficacy of these methods within the context of precision agriculture conditions, where variations are predominantly driven by edaphic and microclimatic factors and are comparatively less extreme. Furthermore, it is crucial to acknowledge that prevailing UAV-based YP studies have primarily revolved around developing empirical regression models. While these models serve the purpose of extrapolating yield estimates to encompass entire fields, it is imperative to recognize their inherent limitation – regression coefficients derived from one year's data may not be transferrable to subsequent years at the same location, nor to different locations within the same year (Rembold et al., 2013).An alternative approach in this domain involves the estimation of yield based on crop growth models. For instance, the GRAMI growth model was successfully applied to rice YP using UAVs GRAMI growth model utilizing UAV-derived data (Kim et al., 2017). Nevertheless, it remains evident that substantial research endeavors are indispensable to delineate the optimal sensor configurations, flight timing, and refinement of crop growth models to harness UAVs information most effectively and reliably for YP purposes. Numerous studies conducted since 2020 are shown in Table 4 as a summary for YP in WMRS crops utilizing UAVs.

4.1 Utilization of machine learning techniques and factors influencing yield prediction

Various methodologies and techniques have been employed in previous findings for YP. Table 4 provides details of the ML

10.3389/fpls.2024.1401246

TABLE 4 Brief summary of the studies conducted for yield prediction in WMRS crops utilizing UAVs.

Crop	Data type	Algorithmic/ Mathematical expressions	Methodology	Accuracy	Country	References
Rice	MS	VIs, A	MSMA	R^2 = 0.75, RRMSE= 0.15	China	(Su et al., 2023)
Maize	MS, R	VIs	OM, DSF	95.75%	Nepal	(Sapkota and Paudyal, 2023)
Rice	MS, WD	LD	MMDL	RMSE= 0.86, RMSPE=14, R^2 = 0.65	Japan	(Mia et al., 2023)
Wheat	MS	VIs, RE, SA	LASSO-R	R^2 = 0.73	Norway	(Shafiee et al., 2023)
Wheat	3-D PC	LAI	GF	R^2 = 0.63	China	(Yang et al., 2023)
Maize	MS	VIs	GL, SVM, RF	88-89%	Lithuania	(Kavaliauskas et al., 2023)
Maize	MS	VIs	LR, KNN, RF, SVR, DNN	R^2 = 0.71, RMSE = 1.08 Mg/ha	USA	(Kumar et al., 2023)
Wheat	RGB, HS-NIR	SR, Th, Tx	ELM	R^2 = 0.74	China	(Ma et al., 2023)
Wheat	MS	VIs	RR	R^2 = 0.651	Spain	(Gracia-Romero et al., 2023)
Rice	HSI	ID	LR	R^2 = 0.858 and RMSPE = 7.52%	Japan	(Kurihara et al., 2023)
Wheat	SR	ND-RE	PLSR	R^2 = 0.81	Germany	(Prey et al., 2023)
Rice	MS	VIs, TIs	RF	R^2 = 0.795, RMSE = 0.298, RRMSE = 0.072	China	(Longfei et al., 2023)
Maize	RGB, MS	ID	RF	R^2 = 0.859, RMSE = 1086.412 kg/ha, RMSE = 13.1%	China	(Liu et al., 2023)
Maize	MS, SR	VIs	ANN, DT, REPT, RF, SVM	R = 0.58	Brazil	(Baio et al., 2022)
Wheat	MS	CIs	PLSR	R^2 = 0.81, RMSE = 1248.48, NRMSE = 21.77%,	China	(Wang et al., 2022a)
Wheat	RGB, MS	VIs	SVM, RF, PLSR, RR, MLR	R^2 = 0.85	Germany	(Prey et al., 2022)
Rice	MS	VIs, TIs	RF	RMSE = 0.94 t/ha, RRMSE = 9.37%	China	(Zheng et al., 2022)
Maize	MS	VIs	PCA, LR, R	R^2 = 0.61	Peru	(Saravia et al., 2022)
Wheat	MS	VIs	MLR, SMLR, PLSR	R^2 = 0.61, RMSE = 7.48 kg yield/kg N, MAE = 6.05 kg yield/kg N	China	(Liu et al., 2022)
Rice	SAR images	Ku band	WCM	92.7%	China	(Wang et al., 2022b)
Wheat	MS	VIs	RF	R^2 = 0.8516, RMSE = 0.0744 kg/m2	China	(Tian et al., 2022)
Wheat	SR, HSI	VIs	PLSR, ANN	RMSE = 599.63 kg/ha, NRMSE = 9.82%	China	(Feng et al., 2022)
Rice	MS	VIs	TCT	83%	China	(Luo et al., 2022)
Wheat	RGB, TIR, MS	VIs	SVM, DNN, RR, RF, EL	R^2 = 0.692	China	(Fei et al., 2023)
Rice	MS	VIs	XGB	R^2 = 0.83	Japan	(Bascon et al., 2022)
Maize	RGB, MS, SR	DF	RF, DCN, RF, DCN, SVM	RRMSE = 17.22%	China	(Yu et al., 2023)
Wheat	TIR, MS	VIs, WIs	NN, RF	R^2 = 0.78, RRMSE = 684.1 kg/ha	China	(Shen et al., 2022)
Wheat	OP	SEL	YOLOX-m	87.93%	China	(Zhaosheng et al., 2022)
Wheat	HSI	FFLR	PLSR	86.58%	USA	(Costa et al., 2022)

(Continued)

TABLE 4 Continued

Crop	Data type	Algorithmic/ Mathematical expressions	Methodology	Accuracy	Country	References
Maize	MS	SIs, TIs	RF	R^2 = 0.93	China	(Yang et al., 2022)
Maize	MS	3D-CNN, 2D-CNN	XGB	RMSE: 8.7–9.3%	USA	(Bellis et al., 2022)
Maize	MS, DI	SIs, TIs	TCM	R^2 = 0.82, RMSE = 38.53 g/m², RRMSE = 29.19%	China	(Meiyan et al., 2022)
Wheat	MS	VIs	GPR, SVR, RFR	R^2 = 0.88, RMSE = 49.18 g/m^2	China	(Bian et al., 2022)
Soybean	MS	VIs	gSW, BMT	R^2 = 0.98	Brazil	(Tavares et al., 2022)
Soybean	MS	OMI	LASSO, PCA,	76%	USA	(Zhou et al., 2022)
Wheat	HSI	SIs, FS	SVM, GP, LRR, RF	R^2 = 0.78	China	(Li et al., 2022)
Wheat	T, MS	VIs, DF, NRCT	ENR, EWF	R^2 = 0.729, RMSE = 0.831 t/ha	China	(Fei et al., 2021b)
Rice	RGB	VIs	RF	R^2 = 0.80	China	(Ge et al., 2021)
Soybean	MS	SBs, VIs	RF, SVM, LR	RMSE = 8.23, MAE = 6.65	Brazil	(Teodoro et al., 2021)
Maize	MS	VIs	DNN	RMSE = 1.07 t/ha, R^2 = 0.73, RRMSE = 7.60% t/ha	Australia	(Danilevicz et al., 2021)
Maize	MS	VIs	EM	R^2 = 0.97	USA	(Sunoj et al., 2021)
Rice	HSI	SIs, TIs	MLR	R^2 = 0.80, RMSE = 0.421 Mg/ha	China	(Wang et al., 2021b)
Wheat	MS, 3D point cloud, SR,	VIs	MTLR, SVM, GPR, ANN	R^2 = 0.88; RMSE = 11.8 g/m^2	Australia	(Roy Choudhury et al., 2021)
Rice	MS	VIs	MLR	R^2 = 0.95	Greece	(Perros et al., 2021)
Wheat	MS	WIs, VIs	LRM	R^2 = 0.87	Canada	(Song et al., 2021)
Rice	MS	VIs, FIs	MLR	R^2 = 0.869, RMSE = 396.02 kg/ha, MAPE= 3.98%	China	(Wang et al., 2021a)
Wheat	RGB	CIs	SVM	RMSE = 32.18 g/m^2, R^2 = 0.93	China	(Zeng et al., 2021)
Wheat	MS	VIs	RF, SVM, GP, RR	R^2 = 0.628	China	(Fei et al., 2021a)
Rice	MS	VIs	SMA, BMM	91.9%	China	(Yuan et al., 2021)
Wheat	TI	TeIs, WIs	CRT	RMSE = 16.7 g/m^2, R^2 = 0.78,	Australia	(Das et al., 2021)
Wheat	MS	VIs	MLR	R^2 = 0.807, RMSE = 781.59 kg/ha	China	(Han et al., 2021)
Rice	MS	VIs	NN	92.9%	China	(Duan et al., 2021)
Wheat	MS	GAI	RA	R^2 = 0.82	Germany	(Bukowiecki et al., 2021)
Maize	MS	VIs	MLR, DT	R = 0.86, RMSE = 0.32	South Africa	(Chivasa et al., 2021)
Maize	HSI	TI, SI	CNN	75.50%	China	(Yang et al., 2021b)
Wheat	MSI	VIs	SFS, LASSO-R, SVR	90%	Norway	(Shafiee et al., 2021)
Wheat	RGB, MS	VIs	DSF, OMI	R^2 = 0.70	USA	(Bhandari et al., 2021)
Maize	MS	VIs	EKF	R^2 = 0.855, RMSE = 692.8kg/ha	China	(Peng et al., 2021)
Soybean	MS	VIs	CNN, DNN, RNN	RMSE = 391 kg/ha	USA	(Zhou et al., 2021a)
Wheat	DI, HSI	VIs, TIs	PLSR, SVM	R^2 = 0.87, RMSE = 119.76 g/m^2	China	(Fu et al., 2021)
Wheat	HSI	VIs	RF	RMSE = 985.83 (kg/ha)	China	(Yang et al., 2021a)

(Continued)

TABLE 4 Continued

Crop	Data type	Algorithmic/ Mathematical expressions	Methodology	Accuracy	Country	References
Maize	MS	VIs	LMvR	R^2 = 0.62	Nigeria	(Adewopo et al., 2020)
Soybean	HSI	VIs, DN	PLSR	R^2 = 0.79,	China	(Li et al., 2021)
Maize	MS	VIs	RF	R = 0.78, MAE = 853.11kg/ha	Brazil	(Ramos et al., 2020)
Wheat	RGB	VIs	PSO	R^2 = 0.63, RMSE = 1.16 t/ha, MAE = 0.96 t/ha, NRMSE = 21.9%	China	(Yue et al., 2021)
Maize	RGB	3D-PC, OMI	LR	R^2 = 0.94	France	(Gilliot et al., 2021)
Rice	RGB, MS	VIs, SI	RF	R^2 = 0.85, RRMSE = 3.56%	China	(Wan et al., 2020)
Wheat	MS, SR	VIs	LR, RF, ANN	RMSE = 1.07%	Japan	(Zhou et al., 2021b)
Maize	RGB	VIs	BP, ExLM, SVM, RF	MAEs = 0.925 g/hundred grain weight	China	(Guo et al., 2020)
Wheat	RGB	DSF	LEER	R^2 = 0.73	Nepal	(Panday et al., 2020)
Wheat	RGB	LAI	SCE	RRMSE = 15.2%	Canada	(Song et al., 2020)
Maize	RGB, MS	VIs	LM, GA	R = 0.97, RMSE = 0.425 t/ha, MAE = 0.249 t/ha	Mexico	(García-Martínez et al., 2020)
Soybean	RGB, MS	PC	RF, XGB	91.36%	USA	(Herrero-Huerta et al., 2020)
Wheat	MS	VIs	LR, MLR, SMLR, PLSR, ANN, RF	R^2 = 0.78, RRMSE = 0.1030.	China	(Fu et al., 2020)
Maize	HSI	VIs	PLSR	RMSE = 2.07 ton/ha, R^2 = 0.73,	Israel	(Herrmann et al., 2020)
Soybean	HSI, Th, Tx	VIs	PLSR, RFR, SVR, DNN	R^2 = 0.720, RMSE= 15.9%	USA	(Maimaitijiang et al., 2020)
Wheat	HSI	SIs	PLSR, ANN, RF	R^2 = 0.77, NRMSE = 10.63%, RMSE = 648.90 kg/ha	China	(Tao et al., 2020)
Wheat	RGB	VIs	PCA,	R^2 = 0.67	Italy	(Marino and Alvino, 2020)
Maize	RGB, MS	VIs	NRM	85-94%	China	(Zhang et al., 2020)

approaches applied in recent studies. The RF technique is widely employed in the domain of YP in WMRS crops, as it is a commonly utilized method in technical approaches. The studies have used numerous techniques, i.e. SVM, KNN, PCA, LR, MLR, MMDL, RR, ANN, PLSR DCN, etc. SVM can be used with statistical techniques like ANOVA to estimate agricultural yield (Li et al., 2018). In wheat crop SVM manifested R^2 = 0.93 using RGB images acquired by UAVs (Wang et al., 2021a). Another study established the R^2 = 0.87 for wheat crops through MS images by UAVs calculating WIs and VIs using the LRM technique (Song et al., 2021). The MLR showed R^2 = 0.81 for wheat crop (Han et al., 2021) and R^2 = 0.95 for rice crop through MS images (Perros et al., 2021). Accordingly, different ML techniques have shown interesting results using raw image data acquired using UAVs (Table 4). The details of the abbreviations is given in Table 5.

The authors give similar importance to each factor for inclusion into the ML model (Zhang et al., 2019; Mustafa et al., 2022b). Osco et al. (2020) consider the significance of several elements like meteorological conditions, geographic location, and radiometric calibration. Each location's meteorological variables (daily daylight hours, daily solar radiation, daily temperature sum, and daily wind speed) were employed, along with field-specific rainfall data. They found that the number of neural layers and the learning rate of an ANN affect the prediction. As a result, the prediction model performs worse when too many or too few parameters are added (Adisa et al., 2019). The conclusion highlights the significance of neural networks in creating prediction models. On the other hand, by including irrelevant features to the model, its complexity may increase, leading to a drop in prediction accuracy concerning time complexity and irrelevant results (Kanning et al., 2018). The study Eugenio et al. (2020) indicates that the UAV is the most effective method for gathering image data from the location. They performed their experiment accounting for distinct weather scenarios. Overall performance shows the accuracy of the model,

10.3389/fpls.2024.1401246

TABLE 5 List of abbreviations for acronyms used in the manuscript.

Acronyms	Abbreviations	Acronyms	Abbreviations
3D-PD	3D-point cloud	MSMA	Member spectral mixture analysis
A	Abundance	MTLR	Multitarget linear regression
ANN	Artificial Neural Network	ND-RE	Normalized difference red edge
BMM	Bilinear mixing model	NN	Neural network
BMT	Box-M test	NRCT	Normalized relative canopy temperature
BP	Backpropagation neural network model	NRM	Nonlinear regression models
CIs	Color indices	NRMSE	Normalized root-mean-square error
CNN	Convolutional neural network	OMI	Orthomosaic images
CRT	Classification and regression tree	OP	Orthophotos
DCN	Deep convolutional network	PC	Point clouds
DF	Data fusion	PCA	Principal component analysis
DI	Digital images	PLS	Partial least squares
DN	Digital numbers	PLSR	Partial least squares regression
DNN	Deep Neural Network	PSO	Particle swarm optimization,
DSF	Digital surface model	R	Pearson correlation coefficient
DT	Decision Tree	R^2	Co-efficient of determination
EKF	Ensemble Kalman Filter	RA	Regression analysis
ELM	Ensemble learning model	RE	Red-edge
EM	Exponential models	REPT	REPTree Decision Tree
ENR	Elastic net regression	RF	Random forest
EWF	Entropy weight fusion	RFR	Random forest regression
ExLM	Extreme learning machine	RMSE	Root mean square error
FFLR	Function on function linear regression	RMSPE	Root mean square percentage error
FIs	Fluorescence Indices	RNN	Recurrent neural network
FS	Feature selection	RR	Ridge regression
GA	Garson's Algorithm	RRMSE	Relative root mean square error
GAI	Green area index	SA	Sun angles
GB	Gradient boost	SBs	Spectral bands
GF	Gap fraction	SCE	Shuffled Complex Evolution
GL	Generalized linear	SI	Spectral information
GP	Gaussian process	SEL	Squeeze-and-Excitation Layer
GPR	Gaussian process regression	SFS	Sequential forward selection
gSW	Generalized Shapiro–Wilk	SIs	Spectral Indices
HSI	Hyperspectral imaging	SMLR	Stepwise MLR
HS-NIR	Hyperspectral Near infrared	SR	Spectral reflectance
ID	Index development	SVM	Support vector machine
KNN	k-Nearest Neighbor	SVR	Support vector regression
LAI	Leaf area Index	TCM	Tridimensional concept mode
LASSO-R	Least absolute shrinkage and selection operator regression	TCT	Tasseled cap transformation
LD	Layer depths	TeIs	Temperature Indices
LEER	Linear, exponential and empirical regression	Th	Thermal
LM	Levenberg–Marquart	TI	Thermal imaging
LMvR	Linear multivariate regression	TxI	Texture Information
LR	Linear Regression	TIR	Thermal infrared
LRM	Linear regression model	TIs	Texture Indices
LRR	Linear ridge regression	Tx	Texture
MAE	Mean absolute Error	VIs	Vegetation Indices
MAPE	Mean absolute percentage error	WCM	water-cloud model

TABLE 5 Continued

Acronyms	Abbreviations	Acronyms	Abbreviations
MLR	Multiple linear regression	WD	Weather data
MMDL	Multimodal deep learning	WIs	Water indices
MS	Multispectral	XGBoost	Extreme gradient boost

although this study has indicated that the inclusion of either an excessive number or a small number of factors causes data sparsity. The main problem this study highlighted is that they employ constant weather conditions for every season. Because of this, a realistic strategy is required, considering the various weather-related influencing factors throughout the year. Peng et al. (2019) has mentioned that ML technology is typically used in precision agriculture to leverage the vast amounts of collected data. ML can estimate certain crop growth rate-related metrics, identify and differentiate objects in photos, and even detect diseases. To take advantage of the data, ML techniques such as CNN, ANN, regression modeling, RF, and deep learning have been employed (Islam et al., 2018). The use of ML has grown dramatically in recent years, partly because deep learning is developing at a rapid pace (Guo et al., 2019; Mustafa et al., 2022a, 2022; Hussain et al., 2023; Mustafa et al., 2024).

4.2 Emerging trends and the current research landscape in the domain of yield prediction in WMRS crops utilizing UAVs

Some of this study's most substantial implications are as follows:

- Regarding YP in WMRS crops utilizing UAVs research, Zhou X, Maimaitijiang M, Yue JB, Jin XL, Bendig J, Haboudane D, and other writers are the most productive at the micro level. Researchers like Shanahan et al. (2001); Chlingaryan et al. (2018); Maes and Steppe (2019); Zheng et al. (2019), and others have been referenced extensively in YP for WMRS crops utilizing UAVs.
- Beijing Academy of Agriculture Forestry Science, Ministry of Agriculture Rural Affairs, Chinese Academy of Agricultural Sciences and China Agriculture University are the most dynamic and productive research for YP in WMRS crops utilizing UAVs at the meso scale.
- China, the United States, Germany, Australia, Canada, Spain, and Brazil are the most active and productive contributors to YP in WMRS crops using UAVs at the macro level. Compared to the other countries on the list, China and the United States presumably have more publications because their governments provide more extensive financial support policies to the scientific community.
- The National Natural Science Fund of China, the National Key Research and Development Program of China, the National Science Foundation, the Fundamental Research Funds for The Central Universities, and others have provided the bulk of the funding for YP in WMRS crops using UAVs.
- Precision Agriculture, Computers and Electronics in Agriculture, Remote Sensing, and Frontiers in Plant Science supported the main journals most.
- Although promising progress has been made in predicting crop yields with WMRS employing UAVs, more definitive results are urgently needed.

These findings provide important insight into the state of research and development in YP in WMRS crops using UAVs, and they highlight the most recent advancements in the field.

5 Constraints and prospects for the future

Despite the extensive research performed on YP of WMRS crops using RS, a definitive technique or technological configuration can still be universally applied. Due to variations in ML approaches, the selected or retrieved specific features such as vegetation indices, color, and spectral properties exhibit fluctuations (Zheng et al., 2019; Baio et al., 2022). Table 4 displays the suitability and chaos of several algorithms. As a result, the primary future vision is to concentrate on certain methodology that can be generalized for accurate quantification and YP.

Examining complex interactions among many environmental conditions will enable more precise and real-time crop YPs (Jung et al., 2021; Mia et al., 2023). A more thorough dataset for enhancing prediction models can be produced by establishing data standards protocols and encouraging data exchange between farmers and researchers; collaboration can result in more accurate insights and forecasts (Wolfert et al., 2017). Additionally, efforts to make UAVs technology less expensive and sophisticated will enhance its use, which will be advantageous to small-scale farmers. While incentives and subsidies provided by the government may be helpful. Crop YP and agriculture management benefit from deep and ML. These new algorithms can more effectively analyze complicated environmental interactions and massive datasets than older methods, resulting in more accurate and real-time yield estimates. Deep learning algorithms can identify complex data patterns for better predictions and insights (Maimaitijiang et al., 2020; Bellis et al., 2022; Mia et al., 2023; Yu et al., 2023).

Additionally, improvements in sensor technology (hyperspectral and LiDAR sensors) can offer even more precise and detailed information, enabling crop health and yield potential (Maes and Steppe, 2019). Also, it will be necessary to push for simplified laws that balance safety issues and the potential advantages of UAVs technology in agriculture. This may encourage a broader uptake of UAVs (McCarthy et al., 2023). Most significantly, maximizing the

effectiveness of UAVs solutions for certain crops and areas, customized sensors and data analysis methods can give farmers and researchers more pertinent information.

6 Contribution of the study for the scientific community

An in-depth scientometric analysis of academic literature from 2001 to 2023 highlights research trends, boundaries and institutional relationships. The study identifies critical gaps, such as reliance on English-language articles and the need for transferable culture-specific features between sensors and algorithms. It proposes standardized data protocols for consistency and comparability, advanced hybrid models combining multiple machine learning (ML) algorithms and remote sensing (RS) techniques, and real-time monitoring systems using machine learning and deep learning to gain immediate insights. In addition, it recommends integrating advanced sensors such as hyperspectral and LiDAR with UAVs, promoting collaboration and data sharing between stakeholders, and advocating balanced government incentives and regulations for UAVs. These contributions aim to mitigate current limitations, ensure more accurate, reliable and widely applicable yield forecasts, and ultimately improve global agricultural productivity.

7 Conclusion

This study conducted a scientometric analysis of the academic literature pertaining to the use of unmanned aerial vehicles (UAVs) for yield prediction (YP) in Wheat, Maize, Rice, and Soybean (WMRS) crops. The retrieval of research published between 2001 and 2023 was conducted using co-citation, co-authorship, and co-occurrence analysis of phrases, utilizing the Web of Science (WOS) database. Research frontiers, trending issues, authors, innovative knowledge systems, and institutional relationships were all taken into account. Although the results of the present study's graphical analysis of related papers are outstanding, the study has many shortcomings. Since the databases that make up the WOS core collection only include English-language articles, the resulting reference footprint is limited. Although remote sensing can predict yields with high accuracy, more research is needed, focusing on identifying transferable crop-specific traits across sensors and algorithms. This would make it possible to estimate yields more precisely and consistently, enhancing crop management techniques, reducing financial losses, and guarantee food security. Thus, research must concentrate on extracting quantitative knowledge on various crop stages utilizing various data types to construct comprehensive decision support systems. These advancements in remote sensing and YP have the potential to increase agricultural productivity across the globe significantly. Moreover, application of the advanced hybrid models, data protocols and feature selection are appealing indicators in conjunction with ML and RS for WMRS crops' yield estimation. Government incentives for farmers, real time monitoring using ML and deep learning, hyperspectral and Lidar sensors, and UAV regulations can enhance the YP investigations.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

GM: Writing – original draft, Writing – review & editing. IK: Formal analysis, Methodology, Writing – original draft. SH: Project administration, Validation, Writing – review & editing. YJ: Methodology, Software, Supervision, Writing – original draft. JL: Data curation, Formal analysis, Investigation, Validation, Writing – review & editing. SA: Resources, Visualization, Writing – review & editing. RO: Software, Validation, Supervision, Writing – original draft. YL: Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. National Key Research and Development Program (2024YFE0103100) and National Natural Science Foundation of China: 3170474.

Acknowledgments

Thanks to all co-authors for best participation in literature review.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer HZ declared a past co-authorship with the author GM to the handling editor.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1401246/full#supplementary-material

References

Aasen, H., Honkavaara, E., Lucieer, A., and Zarco-Tejada, P. J. (2018). Quantitative remote sensing at ultra-high resolution with UAV spectroscopy: a review of sensor technology, measurement procedures, and data correction workflows. *Remote Sens.* 10, 1091. doi: 10.3390/rs10071091

Adamsen, F., Coffelt, T., Nelson, J. M., Barnes, E. M., and Rice, R. C. (2000). Method for using images from a color digital camera to estimate flower number. *Crop Sci.* 40, 704–709. doi: 10.2135/cropsci2000.403704x

Adewopo, J., Peter, H., Mohammed, I., Kamara, A., Craufurd, P., and Vanlauwe, B. (2020). Can a combination of UAV-derived vegetation indices with biophysical variables improve yield variability assessment in smallholder farms? *Agronomy* 10, 1934. doi: 10.3390/agronomy10121934

Adisa, O. M., Botai, J. O., Adeola, A. M., Hassen, A., Botai, C. M., Darkey, D., et al. (2019). Application of artificial neural network for predicting maize production in South Africa. *Sustainability* 11, 1145. doi: 10.3390/su11041145

Ahmad, U., Nasirahmadi, A., Hensel, O., and Marino, S. (2022). Technology and data fusion methods to enhance site-specific crop monitoring. *Agronomy* 12, 555. doi: 10.3390/agronomy12030555

Ahmed, A., Azam, A., Bhutta, M. M. A., Khan, F. A., Aslam, R., and Tahir, Z. (2021). Discovering the technology evolution pathways for 3D printing (3DP) using bibliometric investigation and emerging applications of 3DP during COVID-19. *Clean. Environ. Syst.* 3, 100042. doi: 10.1016/j.cesys.2021.100042

Astor, T., Dayananda, S., Nautiyal, S., and Wachendorf, M. (2020). Vegetable crop biomass estimation using hyperspectral and RGB 3D UAV data. *Agronomy* 10, 1600. doi: 10.3390/agronomy10101600

Azam, A., Ahmed, A., Wang, H., Wang, Y., and Zhang, Z. (2021). Knowledge structure and research progress in wind power generation (WPG) from 2005 to 2020 using CiteSpace based scientometric analysis. *J. Clean. Product.* 295, 126496. doi: 10.1016/j.jclepro.2021.126496

Baio, F. H. R., Santana, D. C., Teodoro, L. P. R., de Oliveira, I. C., Gava, R., de Oliveira, J. L. G., et al. (2022). Maize yield prediction with machine learning, spectral variables and irrigation management. *Remote Sens.* 15, 79. doi: 10.3390/rs15010079

Ballester, C., Hornbuckle, J., Brinkhoff, J., Smith, J., and Quayle, W. (2017). Assessment of in-season cotton nitrogen status and lint yield prediction from unmanned aerial system imagery. *Remote Sens.* 9, 1149. doi: 10.3390/rs9111149

Barnetson, J., Phinn, S., and Scarth, P. (2020). Estimating plant pasture biomass and quality from UAV imaging across Queensland's Rangelands. *AgriEngineering* 2, 523–543. doi: 10.3390/agriengineering2040035

Bascon, M. V., Nakata, T., Shibata, S., Takata, I., Kobayashi, N., Kato, Y., et al. (2022). Estimating yield-related traits using uav-derived multispectral images to improve rice grain yield prediction. *Agriculture* 12, 1141. doi: 10.3390/agriculture12081141

Behrend, J., and Eulerich, M. (2019). The evolution of internal audit research: a bibliometric analysis of published documents, (1926–2016). *Account. History Rev.* 29, 103–139. doi: 10.1080/21552851.2019.1606721

Bellis, E. S., Hashem, A. A., Causey, J. L., Runkle, B. R., Moreno-García, B., Burns, B. W., et al. (2022). Detecting intra-field variation in rice yield with unmanned aerial vehicle imagery and deep learning. *Front. Plant Sci.* 13, 716506. doi: 10.3389/fpls.2022.716506

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., and Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors* 21, 3758. doi: 10.3390/s21113758

Bhandari, M., Baker, S., Rudd, J. C., Ibrahim, A. M., Chang, A., Xue, Q., et al. (2021). Assessing the effect of drought on winter wheat growth using unmanned aerial system (UAS)-based phenotyping. *Remote Sens.* 13, 1144. doi: 10.3390/rs13061144

Bian, C., Shi, H., Wu, S., Zhang, K., Wei, M., Zhao, Y., et al. (2022). Prediction of field-scale wheat yield using machine learning method and multi-spectral UAV data. *Remote Sens.* 14, 1474. doi: 10.3390/rs14061474

Boegh, E., Houborg, R., Bienkowski, J., Braban, C. F., Dalgaard, T., Van Dijk, N., et al. (2013). Remote sensing of LAI, chlorophyll and leaf nitrogen pools of crop-and grasslands in five European landscapes. *Biogeosciences* 10, 6279–6307. doi: 10.5194/bg-10-6279-2013

Bukowiecki, J., Rose, T., and Kage, H. (2021). Sentinel-2 data for precision agriculture?—a uav-based assessment. *Sensors* 21, 2861. doi: 10.3390/s21082861

Bura, R. O., Apriyani, S. W., Ariwibawa, K., and Adharian, E. (2018). "UAV application for oil palm harvest prediction." in *Journal of Physics: Conference Series, Volume 1130, 6th International Seminar of Aerospace Science and Technology* 25–26 September 2018 Jakarta, Indonesia, 012001. doi: 10.1088/1742-6596/1130/1/012001

Chang, Y.-W., Huang, M.-H., and Lin, C.-W. (2015). Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. *Scientometrics* 105, 2071–2087. doi: 10.1007/s11192-015-1762-8

Chaurasia, R., and Mohindru, V. (2021). Unmanned aerial vehicle (UAV): A comprehensive survey. *Unmanned Aerial Vehicles Internet Things (IoT) Concepts Techn. Appl.* 1–27. doi: 10.1002/9781119769170.ch1

Chen, X., and Liu, Y. (2020). Visualization analysis of high-speed railway research based on CiteSpace. *Trans. Policy* 85, 1–17. doi: 10.1016/j.tranpol.2019.10.004

Chivasa, W., Mutanga, O., and Burgueno, J. (2021). UAV-based high-throughput phenotyping to increase prediction and selection accuracy in maize varieties under artificial MSV inoculation. *Comput. Electron. Agric.* 184, 106128. doi: 10.1016/j.compag.2021.106128

Chlingaryan, A., Sukkarieh, S., and Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput. Electron. Agric.* 151, 61–69. doi: 10.1016/j.compag.2018.05.012

Chu, T., Chen, R., Landivar, J. A., Maeda, M. M., Yang, C., and Starek, M. J. (2016). Cotton growth modeling and assessment using unmanned aircraft system visual-band imagery. *J. Appl. Remote Sens.* 10, 036018–036018. doi: 10.1117/1.JRS.10.036018

Concas, F., Mineraud, J., Lagerspetz, E., Varjonen, S., Liu, X., Puolamäki, K., et al. (2021). Low-cost outdoor air quality monitoring and sensor calibration: A survey and critical analysis. *ACM Trans. Sensor Networks (TOSN)* 17, 1–44. doi: 10.1145/3446005

Costa, L., McBreen, J., Ampatzidis, Y., Guo, J., Gahrooei, M. R., and Babar, M. A. (2022). Using UAV-based hyperspectral imaging and functional regression to assist in predicting grain yield and related traits in wheat under heat-related stress environments for the purpose of stable yielding genotypes. *Precis. Agric.* 23, 622–642. doi: 10.1007/s11119-021-09852-5

Danilevicz, M. F., Bayer, P. E., Boussaid, F., Bennamoun, M., and Edwards, D. (2021). Maize yield prediction at an early developmental stage using multispectral images and genotype data for preliminary hybrid selection. *Remote Sens.* 13, 3976. doi: 10.3390/rs13193976

Das, S., Christopher, J., Apan, A., Choudhury, M. R., Chapman, S., Menzies, N. W., et al. (2021). Evaluation of water status of wheat genotypes to aid prediction of yield on sodic soils using UAV-thermal imaging and machine learning. *Agric. For. Meteorol.* 307, 108477. doi: 10.1016/j.agrformet.2021.108477

Delgado, J. A., Short, N. M. Jr., Roberts, D. P., and Vandenberg, B. (2019). Big data analysis for sustainable agriculture on a geospatial cloud framework. *Front. Sustain. Food Syst.* 3, 54. doi: 10.3389/fsufs.2019.00054

Dorj, U.-O., Lee, M., and Yun, S. (2017). An yield estimation in citrus orchards via fruit detection and counting using image processing. *Comput. Electron. Agric.* 140, 103–112. doi: 10.1016/j.compag.2017.05.019

Duan, B., Fang, S., Gong, Y., Peng, Y., Wu, X., and Zhu, R. (2021). Remote estimation of grain yield based on UAV data in different rice cultivars under contrasting climatic zone. *Field Crops Res.* 267, 108148. doi: 10.1016/j.fcr.2021.108148

Elavarasan, D., Vincent, D. R., Sharma, V., Zomaya, A. Y., and Srinivasan, K. (2018). Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Comput. Electron. Agric.* 155, 257–282. doi: 10.1016/j.compag.2018.10.024

Eugenio, F. C., Grohs, M., Venancio, L. P., Schuh, M., Bottega, E. L., Ruoso, R., et al. (2020). Estimation of soybean yield from machine learning techniques and multispectral RPAS imagery. *Remote Sens. Applications: Soc. Environ.* 20, 100397. doi: 10.1016/j.rsase.2020.100397

FAO. (2022). *WORLD FOOD AND AGRICULTURE STATISTICAL YEARBOOK 2022* (Italy: FAO).

Fei, S., Hassan, M. A., He, Z., Chen, Z., Shu, M., Wang, J., et al. (2021a). Assessment of ensemble learning to predict wheat grain yield based on UAV-multispectral reflectance. *Remote Sens.* 13, 2338. doi: 10.3390/rs13122338

Fei, S., Hassan, M. A., Ma, Y., Shu, M., Cheng, Q., Li, Z., et al. (2021b). Entropy weight ensemble framework for yield prediction of winter wheat under different water stress treatments using unmanned aerial vehicle-based multispectral and thermal data. *Front. Plant Sci.* 12,730181. doi: 10.3389/fpls.2021.730181

Fei, S., Hassan, M. A., Xiao, Y., Su, X., Chen, Z., Cheng, Q., et al. (2023). UAV-based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat. *Precis. Agric.* 24, 187–212. doi: 10.1007/s11119-022-09938-8

Feng, H., Tao, H., Fan, Y., Liu, Y., Li, Z., Yang, G., et al. (2022). Comparison of winter wheat yield estimation based on near-surface hyperspectral and UAV hyperspectral remote sensing data. *Remote Sens.* 14, 4158. doi: 10.3390/rs14174158

Fischer, R., Byerlee, D., and Edmeades, G. (2014). Crop yields and global food security. *ACIAR: Canberra ACT,* 8–11. doi: 10.1093/erae/jbv034

Fu, Z., Jiang, J., Gao, Y., Krienke, B., Wang, M., Zhong, K., et al. (2020). Wheat growth monitoring and yield estimation based on multi-rotor unmanned aerial vehicle. *Remote Sens.* 12, 508. doi: 10.3390/rs12030508

Fu, Y., Yang, G., Song, X., Li, Z., Xu, X., Feng, H., et al. (2021). Improved estimation of winter wheat aboveground biomass using multiscale textures extracted from UAV-based digital images and hyperspectral feature analysis. *Remote Sens.* 13, 581. doi: 10.3390/rs13040581

García-Martínez, H., Flores-Magdaleno, H., Ascencio-Hernández, R., Khalil-Gardezi, A., Tijerina-Chávez, L., Mancilla-Villa, O. R., et al. (2020). Corn grain yield estimation from vegetation indices, canopy cover, plant density, and a neural network using multispectral and RGB images acquired with unmanned aerial vehicles. *Agriculture* 10, 277. doi: 10.3390/agriculture10070277

Ge, H., Ma, F., Li, Z., and Du, C. (2021). Grain yield estimation in rice breeding using phenological data and vegetation indices derived from UAV images. *Agronomy* 11, 2439. doi: 10.3390/agronomy11122439

Gilliot, J.-M., Michelin, J., Hadjard, D., and Houot, S. (2021). An accurate method for predicting spatial variability of maize yield from UAV-based plant height estimation: A tool for monitoring agronomic field experiments. *Precis. Agric.* 22, 897–921. doi: 10.1007/s11119-020-09764-w

Gracia-Romero, A., Kefauver, S. C., Vergara-Díaz, O., Zaman-Allah, M. A., Prasanna, B. M., Cairns, J. E., et al. (2017). Comparative performance of ground vs. aerially assessed RGB and multispectral indices for early-growth evaluation of maize performance under phosphorus fertilization. *Front. Plant Sci.* 8, 2004. doi: 10.3389/fpls.2017.02004

Gracia-Romero, A., Rufo, R., Gómez-Candón, D., Soriano, J. M., Bellvert, J., Yannam, V. R. R., et al. (2023). Improving in-season wheat yield prediction using remote sensing and additional agronomic traits as predictors. *Front. Plant Sci.* 14, 1063983. doi: 10.3389/fpls.2023.1063983

Graham, E. A., Yuen, E. M., Robertson, G. F., Kaiser, W. J., Hamilton, M. P., and Rundel, P. (2009). Budburst and leaf area expansion measured with a novel mobile camera system and simple color thresholding. *Environ. Exp. Bot.* 65, 238–244. doi: 10.1016/j.envexpbot.2008.09.013

Guimarães, N., Pádua, L., Marques, P., Silva, N., Peres, E., and Sousa, J. J. (2020). Forestry remote sensing from unmanned aerial vehicles: A review focusing on the data, processing and potentialities. *Remote Sens.* 12, 1046. doi: 10.3390/rs12061046

Guo, C., Tang, Y., Lu, J., Zhu, Y., Cao, W., Cheng, T., et al. (2019). Predicting wheat productivity: Integrating time series of vegetation indices into crop modeling via sequential assimilation. *Agric. For. Meteorol.* 272, 69–80. doi: 10.1016/j.agrformet.2019.01.023

Guo, Y., Wang, H., Wu, Z., Wang, S., Sun, H., Senthilnath, J., et al. (2020). Modified red blue vegetation index for chlorophyll estimation and yield prediction of maize from visible images captured by UAV. *Sensors* 20, 5055. doi: 10.3390/s20185055

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., and Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* 90, 337–352. doi: 10.1016/j.rse.2003.12.013

Han, X., Thomasson, J. A., Bagnall, G. C., Pugh, N. A., Horne, D. W., Rooney, W. L., et al. (2018). Measurement and calibration of plant-height from fixed-wing UAV images. *Sensors* 18, 4092. doi: 10.3390/s18124092

Han, X., Wei, Z., Chen, H., Zhang, B., Li, Y., and Du, T. (2021). Inversion of winter wheat growth parameters and yield under different water treatments based on UAV multispectral remote sensing. *Front. Plant Sci.* 12, 609876. doi: 10.3389/fpls.2021.609876

Hassan, M. A., Yang, M., Rasheed, A., Yang, G., Reynolds, M., Xia, X., et al. (2019). A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. *Plant Sci.* 282, 95–103. doi: 10.1016/j.plantsci.2018.10.022

Herrero-Huerta, M., Rodriguez-Gonzalvez, P., and Rainey, K. M. (2020). Yield prediction by machine learning from UAS-based multi-sensor data fusion in soybean. *Plant Methods* 16, 1–16. doi: 10.1186/s13007-020-00620-6

Herrmann, I., Bdolach, E., Montekyo, Y., Rachmilevitch, S., Townsend, P. A., and Karnieli, A. (2020). Assessment of maize yield and phenology by drone-mounted superspectral camera. *Precis. Agric.* 21, 51–76. doi: 10.1007/s11119-019-09659-5

Hussain, S., Mustafa, G., Haider Khan, I., Liu, J., Chen, C., Hu, B., et al. (2023). Global trends and future directions in agricultural remote sensing for wheat scab detection: insights from a bibliometric analysis. *Remote Sens.* 15, 3431. doi: 10.3390/rs15133431

Islam, T., Chisty, T. A., and Chakrabarty, A. (2018). "A deep neural network approach for crop selection and yield prediction in Bangladesh," in *2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Malambe, Sri Lanka, 1–6. doi: 10.1109/R10-HTC.2018.8629828

Jung, J., Maeda, M., Chang, A., Bhandari, M., Ashapure, A., and Landivar-Bowles, J. (2021). The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems. *Curr. Opin. Biotechnol.* 70, 15–22. doi: 10.1016/j.copbio.2020.09.003

Kanning, M., Kühling, I., Trautz, D., and Jarmer, T. (2018). High-resolution UAV-based hyperspectral imagery for LAI and chlorophyll estimations from wheat for yield prediction. *Remote Sens.* 10, 2000. doi: 10.3390/rs10122000

Kavaliauskas, A., Žydelis, R., Castaldi, F., Auškalnienė, O., and Povilaitis, V. (2023). Predicting maize theoretical methane yield in combination with ground and UAV remote data using machine learning. *Plants* 12, 1823. doi: 10.3390/plants12091823

Kefauver, S. C., Vicente, R., Vergara-Díaz, O., Fernandez-Gallego, J. A., Kerfal, S., Lopez, A., et al. (2017). Comparative UAV and field phenotyping to assess yield and nitrogen use efficiency in hybrid and conventional barley. *Front. Plant Sci.* 8, 1733. doi: 10.3389/fpls.2017.01733

Kim, M., Ko, J., Jeong, S., Yeom, J., and Kim, H. (2017). Monitoring canopy growth and grain yield of paddy rice in South Korea by using the GRAMI model and high spatial resolution imagery. *GIScience Remote Sens.* 54, 534–551. doi: 10.1080/15481603.2017.1291783

Kumar, C., Mubvumba, P., Huang, Y., Dhillon, J., and Reddy, K. (2023). Multi-stage corn yield prediction using high-resolution UAV multispectral data and machine learning models. *Agronomy* 13, 1277. doi: 10.3390/agronomy13051277

Kurihara, J., Nagata, T., and Tomiyama, H. (2023). Rice yield prediction in different growth environments using unmanned aerial vehicle-based hyperspectral imaging. *Remote Sens.* 15, 2004. doi: 10.3390/rs15082004

Kyratzis, A. C., Skarlatos, D. P., Menexes, G. C., Vamvakousis, V. F., and Katsiotis, A. (2017). Assessment of vegetation indices derived by UAV imagery for durum wheat phenotyping under a water limited and heat stressed Mediterranean environment. *Front. Plant Sci.* 8, 1114. doi: 10.3389/fpls.2017.01114

Lee, W.-S., Alchanatis, V., Yang, C., Hirafuji, M., Moshou, D., and Li, C. (2010). Sensing technologies for precision specialty crop production. *Comput. Electron. Agric.* 74, 2–33. doi: 10.1016/j.compag.2010.08.005

Lee, K.-J., and Lee, B.-W. (2011). Estimating canopy cover from color digital camera image of rice field. *J. Crop Sci. Biotechnol.* 14, 151–155. doi: 10.1007/s12892-011-0029-z

Li, Z., Chen, Z., Cheng, Q., Duan, F., Sui, R., Huang, X., et al. (2022). UAV-based hyperspectral and ensemble machine learning for predicting yield in winter wheat. *Agronomy* 12, 202. doi: 10.3390/agronomy12010202

Li, C., Ma, C., Cui, Y., Lu, G., and Wei, F. (2021). Uav hyperspectral remote sensing estimation of soybean yield based on physiological and ecological parameter and meteorological factor in China. *J. Indian Soc. Remote Sens.* 49, 873–886. doi: 10.1007/s12524-020-01269-3

Li, Z., Wang, J., Liu, C., Song, X., Xu, X., and Li, Z. (2018). "Predicting grain protein content in winter wheat using hyperspectral and meteorological factor," in *2018 7th International Conference on Agro-geoinformatics (Agro-geoinformatics)*, Hangzhou, China, 1–4. doi: 10.1109/Agro-Geoinformatics.2018.8476098

Liao, H., Tang, M., Luo, L., Li, C., Chiclana, F., and Zeng, X.-J. (2018). A bibliometric analysis and visualization of medical big data research. *Sustainability* 10, 166. doi: 10.3390/su10010166

Liu, Y., Nie, C., Zhang, Z., Wang, Z., Ming, B., Xue, J., et al. (2023). Evaluating how lodging affects maize yield estimation based on UAV observations. *Front. Plant Sci.* 13, 979103. doi: 10.3389/fpls.2022.979103

Liu, J., Zhu, Y., Tao, X., Chen, X., and Li, X. (2022). Rapid prediction of winter wheat yield and nitrogen use efficiency using consumer-grade unmanned aerial vehicles multispectral imagery. *Front. Plant Sci.* 13, 1032170. doi: 10.3389/fpls.2022.1032170

Longfei, Z., Ran, M., Xing, Y., Yigui, L., Zehua, H., Zhengang, L., et al. (2023). Improved yield prediction of ratoon rice using unmanned aerial vehicle-based multi-temporal feature method. *Rice Sci.* 30, 247–256. doi: 10.1016/j.rsci.2023.03.008

Luo, S., Jiang, X., Jiao, W., Yang, K., Li, Y., and Fang, S. (2022). Remotely sensed prediction of rice yield at different growth durations using UAV multispectral imagery. *Agriculture* 12, 1447. doi: 10.3390/agriculture12091447

Ma, J., Liu, B., Ji, L., Zhu, Z., Wu, Y., and Jiao, W. (2023). Field-scale yield prediction of winter wheat under different irrigation regimes based on dynamic fusion of multimodal UAV imagery. *Int. J. Appl. Earth Observ. Geoinform.* 118, 103292. doi: 10.1016/j.jag.2023.103292

Maes, W. H., and Steppe, K. (2019). Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends Plant Sci.* 24, 152–164. doi: 10.1016/j.tplants.2018.11.007

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., and Fritschi, F. B. (2020). Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sens. Environ.* 237, 111599. doi: 10.1016/j.rse.2019.111599

Marino, S., and Alvino, A. (2020). Agronomic traits analysis of ten winter wheat cultivars clustered by UAV-derived vegetation indices. *Remote Sens.* 12, 249. doi: 10.3390/rs12020249

McCarthy, C., Nyoni, Y., Kachamba, D. J., Banda, L. B., Moyo, B., Chisambi, C., et al. (2023). Can drones help smallholder farmers improve agriculture efficiencies and reduce food insecurity in sub-Saharan Africa? Local perceptions from Malawi. *Agriculture* 13, 1075. doi: 10.3390/agriculture13051075

Meiyan, S., Mengyuan, S., Qizhou, D., Xiaohong, Y., Baoguo, L., and Yuntao, M. (2022). Estimating the maize above-ground biomass by constructing the tridimensional concept model based on UAV-based digital and multi-spectral images. *Field Crops Res.* 282, 108491. doi: 10.1016/j.fcr.2022.108491

Melesse, A. M., Weng, Q., Thenkabail, P. S., and Senay, G. B. (2007). Remote sensing sensors and applications in environmental resources mapping and modelling. *Sensors* 7, 3209–3241. doi: 10.3390/s7123209

Mia, M. S., Tanabe, R., Habibi, L. N., Hashimoto, N., Homma, K., Maki, M., et al. (2023). Multimodal deep learning for rice yield prediction using UAV-based multispectral imagery and weather data. *Remote Sens.* 15, 2511. doi: 10.3390/rs15102511

Mohsan, S. A. H., Othman, N. Q. H., Li, Y., Alsharif, M. H., and Khan, M. A. (2023). Unmanned aerial vehicles (UAVs): Practical aspects, applications, open challenges, security issues, and future trends. *Intell. Serv. Robot.* 16, 109–137. doi: 10.1007/s11370-022-00452-4

Muchiri, G., and Kimathi, S. (2022). "A review of applications and potential applications of UAV," in *Proceedings of the Sustainable Research and Innovation Conference*, 280–283.

Müller, B., Hardt, L., Armbruster, A., Kiefer, K., and Reise, C. (2016). Yield predictions for photovoltaic power plants: empirical validation, recent advances and remaining uncertainties. *Prog. Photovol.: Res. Appl.* 24, 570–583. doi: 10.1002/pip.2616

Mustafa, G., Zheng, H., Khan, I. H., Tian, L., Jia, H., Li, G., et al. (2022a). Hyperspectral reflectance proxies to diagnose in-field fusarium head blight in wheat with machine learning. *Remote Sens.* 14, 2784. doi: 10.3390/rs14122784

Mustafa, G., Zheng, H., Khan, I. H., Zhu, J., Yang, T., Wang, A., et al. (2024). Enhancing fusarium head blight detection in wheat crops using hyperspectral indices and machine learning classifiers. *Comput. Electron. Agric.* 218, 108663. doi: 10.1016/j.compag.2024.108663

Mustafa, G., Zheng, H., Li, W., Yin, Y., Wang, Y., Zhou, M., et al. (2022b). Fusarium head blight monitoring in wheat ears using machine learning and multimodal data from asymptomatic to symptomatic periods. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1102341

Nabwire, S., Suh, H.-K., Kim, M. S., Baek, I., and Cho, B.-K. (2021). Application of artificial intelligence in phenomics. *Sensors* 21, 4363. doi: 10.3390/s21134363

Osco, L. P., Junior, J. M., Ramos, A. P. M., Furuya, D. E. G., Santana, D. C., Teodoro, L. P. R., et al. (2020). Leaf nitrogen concentration and plant height prediction for maize using UAV-based multispectral imagery and machine learning techniques. *Remote Sens.* 12, 3237. doi: 10.3390/rs12193237

Ostos-Garrido, F. J., De Castro, A. I., Torres-Sánchez, J., Pistón, F., and Peña, J. M. (2019). High-throughput phenotyping of bioethanol potential in cereals using UAV-based multi-spectral imagery. *Front. Plant Sci.* 10, 948. doi: 10.3389/fpls.2019.00948

Pádua, L., Vanko, J., Hruška, J., Adão, T., Sousa, J. J., Peres, E., et al. (2017). UAS, sensors, and data processing in agroforestry: A review towards practical applications. *Int. J. Remote Sens.* 38, 2349–2391. doi: 10.1080/01431161.2017.1297548

Panday, U. S., Shrestha, N., Maharjan, S., Pratihast, A. K., Shahnawaz,, Shrestha, K. L., et al. (2020). Correlating the plant height of wheat with above-ground biomass and crop yield using drone imagery and crop surface model, a case study from Nepal. *Drones* 4, 28. doi: 10.3390/drones4030028

Peng, X., Han, W., Ao, J., and Wang, Y. (2021). Assimilation of LAI derived from UAV multispectral data into the SAFY model to estimate maize yield. *Remote Sens.* 13, 1094. doi: 10.3390/rs13061094

Peng, Y., Li, Y., Dai, C., Fang, S., Gong, Y., Wu, X., et al. (2019). Remote prediction of yield based on LAI estimation in oilseed rape under different planting methods and nitrogen fertilizer applications. *Agric. For. Meteorol.* 271, 116–125. doi: 10.1016/j.agrformet.2019.02.032

Perros, N., Kalivas, D., and Giovos, R. (2021). Spatial analysis of agronomic data and uav imagery for rice yield estimation. *Agriculture* 11, 809. doi: 10.3390/agriculture11090809

Prey, L., Hanemann, A., Ramgraber, L., Seidl-Schulz, J., and Noack, P. O. (2022). UAV-Based estimation of grain yield for plant breeding: Applied strategies for optimizing the use of sensors, vegetation indices, growth stages, and machine learning algorithms. *Remote Sens.* 14, 6345. doi: 10.3390/rs14246345

Prey, L., Ramgraber, L., Seidl-Schulz, J., Hanemann, A., and Noack, P. O. (2023). The transferability of spectral grain yield prediction in wheat breeding across years and trial locations. *Sensors* 23, 4177. doi: 10.3390/s23084177

Puget Systems (2021). *Recommended Systems for Pix4D*. Available online at: https://www.pugetsystems.com/recommended/Recommended-Systems-for-Pix4D-207.

Qian, J., Law, R., and Wei, J. (2019). Knowledge mapping in travel website studies: A scientometric review. *Scandinavian J. Hospital. Tourism* 19, 192–209. doi: 10.1080/15022250.2018.1526113

Ramos, A. P. M., Osco, L. P., Furuya, D. E. G., Gonçalves, W. N., Santana, D. C., Teodoro, L. P. R., et al. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Comput. Electron. Agric.* 178, 105791. doi: 10.1016/j.compag.2020.105791

Raun, W. R., Solie, J. B., Johnson, G. V., Stone, M. L., Lukina, E. V., Thomason, W. E., et al. (2001). In-season prediction of potential grain yield in winter wheat using canopy reflectance. *Agron. J.* 93, 131–138. doi: 10.2134/agronj2001.931131x

Rembold, F., Atzberger, C., Savin, I., and Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. *Remote Sens.* 5, 1704–1733. doi: 10.3390/rs5041704

Roy Choudhury, M., Das, S., Christopher, J., Apan, A., Chapman, S., Menzies, N. W., et al. (2021). Improving biomass and grain yield prediction of wheat genotypes on sodic soil using integrated high-resolution multispectral, hyperspectral, 3D point cloud, and machine learning techniques. *Remote Sens.* 13, 3482. doi: 10.3390/rs13173482

Samaras, S., Diamantidou, E., Ataloglou, D., Sakellariou, N., Vafeiadis, A., Magoulianitis, V., et al. (2019). Deep learning on multi sensor data for counter UAV applications—A systematic review. *Sensors* 19, 4837. doi: 10.3390/s19224837

Santesteban, L., Di Gennaro, S., Herrero-Langreo, A., Miranda, C., Royo, J., and Matese, A. (2017). High-resolution UAV-based thermal imaging to estimate the instantaneous and seasonal variability of plant water status within a vineyard. *Agric. Water Manage.* 183, 49–59. doi: 10.1016/j.agwat.2016.08.026

Sapkota, S., and Paudyal, D. R. (2023). Growth monitoring and yield estimation of maize plant using unmanned aerial vehicle (UAV) in a hilly region. *Sensors* 23, 5432. doi: 10.3390/s23125432

Saravia, D., Salazar, W., Valqui-Valqui, L., Quille-Mamani, J., Porras-Jorge, R., Corredor, F.-A., et al. (2022). Yield predictions of four hybrids of maize (Zea mays) using multispectral images obtained from UAV in the Coast of Peru. *Agronomy* 12, 2630. doi: 10.3390/agronomy12112630

Sarkar, T. K., Ryu, C.-S., Kang, Y.-S., Kim, S.-H., Jeon, S.-R., Jang, S.-H., et al. (2018). Integrating UAV remote sensing with GIS for predicting rice grain protein. *J. Biosyst. Eng.* 43, 148–159. doi: 10.5307/JBE.2018.43.2.148

Shafiee, S., Lied, L. M., Burud, I., Dieseth, J. A., Alsheikh, M., and Lillemo, M. (2021). Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. *Comput. Electron. Agric.* 183, 106036. doi: 10.1016/j.compag.2021.106036

Shafiee, S., Mroz, T., Burud, I., and Lillemo, M. (2023). Evaluation of UAV multispectral cameras for yield and biomass prediction in wheat under different sun elevation angles and phenological stages. *Comput. Electron. Agric.* 210, 107874. doi: 10.1016/j.compag.2023.107874

Shanahan, J. F., Schepers, J. S., Francis, D. D., Varvel, G. E., Wilhelm, W. W., Tringe, J. M., et al. (2001). Use of remote-sensing imagery to estimate corn grain yield. *Agron. J.* 93, 583–589. doi: 10.2134/agronj2001.933583x

Shen, Y., Mercatoris, B., Cao, Z., Kwan, P., Guo, L., Yao, H., et al. (2022). Improving wheat yield prediction accuracy using LSTM-RF framework based on UAV thermal infrared and multispectral imagery. *Agriculture* 12, 892. doi: 10.3390/agriculture12060892

Song, Y., Wang, J., and Shan, B. (2021). Estimation of winter wheat yield from UAV-based multi-temporal imagery using crop allometric relationship and SAFY model. *Drones* 5, 78. doi: 10.3390/drones5030078

Song, Y., Wang, J., Shang, J., and Liao, C. (2020). Using UAV-based SOPC derived LAI and SAFY model for biomass and yield estimation of winter wheat. *Remote Sens.* 12, 2378. doi: 10.3390/rs12152378

Su, X., Wang, J., Ding, L., Lu, J., Zhang, J., Yao, X., et al. (2023). Grain yield prediction using multi-temporal UAV-based multispectral vegetation indices and endmember abundance in rice. *Field Crops Res.* 299, 108992. doi: 10.1016/j.fcr.2023.108992

Sunoj, S., Cho, J., Guinness, J., van Aardt, J., Czymmek, K. J., and Ketterings, Q. M. (2021). Corn grain yield prediction and mapping from Unmanned Aerial System (UAS) multispectral imagery. *Remote Sens.* 13, 3948. doi: 10.3390/rs13193948

Sziroczak, D., Rohacs, D., and Rohacs, J. (2022). Review of using small UAV based meteorological measurements for road weather management. *Prog. Aerospace Sci.* 134, 100859. doi: 10.1016/j.paerosci.2022.100859

Tao, H., Feng, H., Xu, L., Miao, M., Yang, G., Yang, X., et al. (2020). Estimation of the yield and plant height of winter wheat using UAV-based hyperspectral images. *Sensors* 20, 1231. doi: 10.3390/s20041231

Tavares, C. J., Ribeiro Junior, W. Q., Ramos, M. L. G., Pereira, L. F., Casari, R. A., das, C. N., et al. (2022). Water stress alters morphophysiological, grain quality and vegetation indices of soybean cultivars. *Plants* 11, 559. doi: 10.3390/plants11040559

Teodoro, P. E., Teodoro, L. P. R., Baio, F. H. R., da Silva Junior, C. A., dos Santos, R. G., Ramos, A. P. M., et al. (2021). Predicting days to maturity, plant height, and grain yield in soybean: A machine and deep learning approach using multispectral data. *Remote Sens.* 13, 4632. doi: 10.3390/rs13224632

Tian, Z., Zhang, Y., Liu, K., Li, Z., Li, M., Zhang, H., et al. (2022). UAV remote sensing prediction method of winter wheat yield based on the fused features of crop and soil. *Remote Sens.* 14, 5054. doi: 10.3390/rs14195054

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci.* 108, 20260–20264. doi: 10.1073/pnas.1116437108

Ubina, N. A., and Cheng, S.-C. (2022). A review of unmanned system technologies with its application to aquaculture farm monitoring and management. *Drones* 6, 12. doi: 10.3390/drones6010012

Van Eck, N., and Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *scientometrics* 84, 523–538. doi: 10.1007/s11192-009-0146-3

Vega, F. A., Ramirez, F. C., Saiz, M. P., and Rosúa, F. O. (2015). Multi-temporal imaging using an unmanned aerial vehicle for monitoring a sunflower crop. *Biosyst. Eng.* 132, 19–27. doi: 10.1016/j.biosystemseng.2015.01.008

Wahab, I., Hall, O., and Jirström, M. (2018). Remote sensing of yields: Application of uav imagery-derived ndvi for estimating maize vigor and yields in complex farming systems in sub-saharan africa. *Drones* 2, 28. doi: 10.3390/drones2030028

Wan, L., Cen, H., Zhu, J., Zhang, J., Zhu, Y., Sun, D., et al. (2020). Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer–a case study of small farmlands in the South of China. *Agric. For. Meteorol.* 291, 108096. doi: 10.1016/j.agrformet.2020.108096

Wang, D., Li, R., Zhu, B., Liu, T., Sun, C., and Guo, W. (2022a). Estimation of wheat plant height and biomass by combining uav imagery and elevation data. *Agriculture* 13, 9. doi: 10.3390/agriculture13010009

Wang, Z., Wang, S., Wang, H., Liu, L., Li, Z., Zhu, Y., et al. (2022b). Field-scale rice yield estimation based on UAV-based MiniSAR data with Ku band and modified water-cloud model of panicle layer at panicle stage. *Front. Plant Sci.* 13, 1001779. doi: 10.3389/fpls.2022.1001779

Wang, F., Yao, X., Xie, L., Zheng, J., and Xu, T. (2021a). Rice yield estimation based on vegetation index and florescence spectral information from UAV hyperspectral remote sensing. *Remote Sens.* 13, 3390. doi: 10.3390/rs13173390

Wang, F., Yi, Q., Hu, J., Xie, L., Yao, X., Xu, T., et al. (2021b). Combining spectral and textural information in UAV hyperspectral images to estimate rice grain yield. *Int. J. Appl. Earth Observ. Geoinform.* 102, 102397. doi: 10.1016/j.jag.2021.102397

Wolfert, S., Ge, L., Verdouw, C., and Bogaardt, M.-J. (2017). Big data in smart farming–a review. *Agric. Syst.* 153, 69–80. doi: 10.1016/j.agsy.2017.01.023

Yang, S., Hu, L., Wu, H., Ren, H., Qiao, H., Li, P., et al. (2021a). Integration of crop growth model and random forest for winter wheat yield estimation from UAV hyperspectral imagery. *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* 14, 6253–6269. doi: 10.1109/JSTARS.2021.3089203

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., et al. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. *Front. Plant Sci.* 8, 1111. doi: 10.3389/fpls.2017.01111

Yang, W., Nigon, T., Hao, Z., Paiao, G. D., Fernández, F. G., Mulla, D., et al. (2021b). Estimation of corn yield based on hyperspectral imagery and convolutional neural network. *Comput. Electron. Agric.* 184, 106092. doi: 10.1016/j.compag.2021.106092

Yang, J., Xing, M., Tan, Q., Shang, J., Song, Y., Ni, X., et al. (2023). Estimating effective leaf area index of winter wheat based on UAV point cloud data. *Drones* 7, 299. doi: 10.3390/drones7050299

Yang, B., Zhu, W., Rezaei, E. E., Li, J., Sun, Z., and Zhang, J. (2022). The optimal phenological phase of maize for yield prediction with high-frequency UAV remote sensing. *Remote Sens.* 14, 1559. doi: 10.3390/rs14071559

Yu, D., Zha, Y., Sun, Z., Li, J., Jin, X., Zhu, W., et al. (2023). Deep convolutional neural networks for estimating maize above-ground biomass using multi-source UAV images: A comparison with traditional machine learning algorithms. *Precis. Agric.* 24, 92–113. doi: 10.1007/s11119-022-09932-0

Yuan, N., Gong, Y., Fang, S., Liu, Y., Duan, B., Yang, K., et al. (2021). UAV remote sensing estimation of rice yield based on adaptive spectral endmembers and bilinear mixing model. *Remote Sens.* 13, 2190. doi: 10.3390/rs13112190

Yue, J., Feng, H., Li, Z., Zhou, C., and Xu, K. (2021). Mapping winter-wheat biomass and grain yield based on a crop model and UAV remote sensing. *Int. J. Remote Sens.* 42, 1577–1601. doi: 10.1080/01431161.2020.1823033

Yue, J., Yang, G., Tian, Q., Feng, H., Xu, K., and Zhou, C. (2019). Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices. *ISPRS J. Photogram. Remote Sens.* 150, 226–244. doi: 10.1016/j.isprsjprs.2019.02.022

Zeng, L., Peng, G., Meng, R., Man, J., Li, W., Xu, B., et al. (2021). Wheat yield prediction based on unmanned aerial vehicles-collected red–green–blue imagery. *Remote Sens.* 13, 2937. doi: 10.3390/rs13152937

Zhang, J., Liu, X., Liang, Y., Cao, Q., Tian, Y., Zhu, Y., et al. (2019). Using a portable active sensor to monitor growth parameters and predict grain yield of winter wheat. *Sensors* 19, 1108. doi: 10.3390/s19051108

Zhang, M., Zhou, J., Sudduth, K. A., and Kitchen, N. R. (2020). Estimation of maize yield and effects of variable-rate nitrogen application using UAV-based RGB imagery. *Biosyst. Eng.* 189, 24–35. doi: 10.1016/j.biosystemseng.2019.11.001

Zhao, X., Yuan, Y., Song, M., Ding, Y., Lin, F., Liang, D., et al. (2019). Use of unmanned aerial vehicle imagery and deep learning unet to extract rice lodging. *Sensors* 19, 3859. doi: 10.3390/s19183859

Zhao, B., Zhang, J., Yang, C., Zhou, G., Ding, Y., Shi, Y., et al. (2018). Rapeseed seedling stand counting and seeding performance evaluation at two early growth stages based on unmanned aerial vehicle imagery. *Front. Plant Sci.* 9, 1362. doi: 10.3389/fpls.2018.01362

Zhaosheng, Y., Tao, L., Tianle, Y., Chengxin, J., and Chengming, S. (2022). Rapid detection of wheat ears in orthophotos from unmanned aerial vehicles in fields based on YOLOX. *Front. Plant Sci.* 13, 851245. doi: 10.3389/fpls.2022.851245

Zheng, H., Cheng, T., Zhou, M., Li, D., Yao, X., Tian, Y., et al. (2019). Improved estimation of rice aboveground biomass combining textural and spectral analysis of UAV imagery. *Precis. Agric.* 20, 611–629. doi: 10.1007/s11119-018-9600-7

Zheng, H., Ji, W., Wang, W., Lu, J., Li, D., Guo, C., et al. (2022). Transferability of models for predicting rice grain yield from unmanned aerial vehicle (UAV) multispectral imagery across years, cultivars and sensors. *Drones* 6, 423. doi: 10.3390/drones6120423

Zhou, J., Beche, E., Vieira, C. C., Yungbluth, D., Zhou, J., Scaboo, A., et al. (2022). Improve soybean variety selection accuracy using UAV-based high-throughput phenotyping technology. *Front. Plant Sci.* 12, 768742. doi: 10.3389/fpls.2021.768742

Zhou, X., Kono, Y., Win, A., Matsui, T., and Tanaka, T. S. (2021b). Predicting within-field variability in grain yield and protein content of winter wheat using UAV-based multispectral imagery and machine learning approaches. *Plant Product. Sci.* 24, 137–151. doi: 10.1080/1343943X.2020.1819165

Zhou, X., Zheng, H., Xu, X., He, J., Ge, X., Yao, X., et al. (2017). Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS J. Photogram. Remote Sens.* 130, 246–255. doi: 10.1016/j.isprsjprs.2017.05.003

Zhou, J., Zhou, J., Ye, H., Ali, M. L., Chen, P., and Nguyen, H. T. (2021a). Yield estimation of soybean breeding lines under drought stress using unmanned aerial vehicle-based imagery and convolutional neural network. *Biosyst. Eng.* 204, 90–103. doi: 10.1016/j.biosystemseng.2021.01.017

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 19 August 2024
DOI 10.3389/fpls.2024.1459968

Check for updates

OPEN ACCESS

EDITED BY
Liujun Xiao,
Nanjing Agricultural University, China

REVIEWED BY
Chengming Sun,
Yangzhou University, China
Yahui Guo,
Central China Normal University, China
Shunfu Xiao,
China Agricultural University, China

*CORRESPONDENCE
Shijuan Li
lishijuan@caas.cn

RECEIVED 05 July 2024
ACCEPTED 29 July 2024
PUBLISHED 19 August 2024

CITATION
Sun S, Zhu Y, Liu S, Chen Y, Zhang Y and Li S (2024) An integrated method for phenotypic analysis of wheat based on multi-view image sequences: from seedling to grain filling stages.
Front. Plant Sci. 15:1459968.
doi: 10.3389/fpls.2024.1459968

COPYRIGHT
© 2024 Sun, Zhu, Liu, Chen, Zhang and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An integrated method for phenotypic analysis of wheat based on multi-view image sequences: from seedling to grain filling stages

Shengxuan Sun[1], Yeping Zhu[1], Shengping Liu[1], Yongkuai Chen[2], Yihan Zhang[3] and Shijuan Li[1]*

[1]Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing, China, [2]Institute of Digital Agriculture, Fujian Academy of Agricultural Sciences, Fuzhou, Fujian, China, [3]College of Letters & Science, University of California, Davis, Davis, CA, United States

Wheat exhibits complex characteristics during its growth, such as extensive tillering, slender and soft leaves, and severe organ cross-obscuration, posing a considerable challenge in full-cycle phenotypic monitoring. To address this, this study presents a synthesized method based on SFM-MVS (Structure-from-Motion, Multi-View Stereo) processing for handling and segmenting wheat point clouds, covering the entire growth cycle from seedling to grain filling stages. First, a multi-view image acquisition platform was constructed to capture image sequences of wheat plants, and dense point clouds were generated using SFM-MVS technology. High-quality dense point clouds were produced by implementing improved Euclidean clustering combined with centroids, color filtering, and statistical filtering methods. Subsequently, the segmentation of wheat plant stems and leaves was performed using the region growth segmentation algorithm. Although segmentation performance was suboptimal during the tillering, jointing, and booting stages due to the glut leaves and severe overlap, there was a salient improvement in wheat leaf segmentation efficiency over the entire growth cycle. Finally, phenotypic parameters were analyzed across different growth stages, comparing automated measurements of plant height, leaf length, and leaf width with actual measurements. The results demonstrated coefficients of determination (R^2) of 0.9979, 0.9977, and 0.995; root mean square errors (RMSE) of 1.0773 cm, 0.2612 cm, and 0.0335 cm; and relative root mean square errors (RRMSE) of 2.1858%, 1.7483%, and 2.8462%, respectively. These results validate the reliability and accuracy of our proposed workflow in processing wheat point clouds and automatically extracting plant height, leaf length, and leaf width, indicating that our 3D reconstructed wheat model achieves high precision and can quickly, accurately, and non-destructively extract phenotypic parameters. Additionally, plant height, convex hull volume, plant surface area, and Crown area were extracted, providing a detailed analysis of dynamic changes in wheat throughout its growth cycle. ANOVA was conducted across different cultivars, accurately revealing significant differences

at various growth stages. This study proposes a convenient, rapid, and quantitative analysis method, offering crucial technical support for wheat plant phenotypic analysis and growth dynamics monitoring, applicable for precise full-cycle phenotypic monitoring of wheat.

KEYWORDS

wheat plant, multi-view stereo reconstruction, phenotype analysis, point cloud processing, growth dynamics analysis

1 Introduction

With the rapid growth of the global population and the intensification of environmental challenges, global agriculture faces serious challenges of ensuring food safety, necessitating the breeding of high-yield, stable crop genotypes that can adapt to climate change to secure future food and fuel supplies (Furbank and Tester, 2011). As one of the three major global food crops, wheat plays a crucial role in global food safety (Erenstein et al., 2022). Accurate acquisition of wheat phenotypic traits is of great practical importance for crop growth assessment, yield estimation, strain breeding and variety improvement (Fiorani and Schurr, 2013; Araus and Cairns, 2014). However, traditional manual measurements are inefficient and destructive in modern plant sciences, making it difficult to capture small changes in crop growth, thereby impacting breeding efficiency and progress (Das Choudhury et al., 2019; Liu et al., 2023). Therefore, there is an urgent call to develop advanced phenotypic data acquisition techniques (Yang et al., 2020; Li et al., 2022) to obtain plant phenotypic parameters in a non-destructive, rapid, efficient, and precise manner (Xie and Yang, 2020).

To this end, researchers have widely applied various high-tech sensors to collect plant morphological phenotypic information, including RGB cameras (Wu et al., 2022b; Wei et al., 2024; Yang et al., 2024), depth cameras (Teng et al., 2021; Ma et al., 2022; Liu et al., 2023), LiDAR (Sun et al., 2018; Forero et al., 2022; Liao et al., 2022), and three-dimensional (3D) scanners (Chenxi et al., 2022; Wu et al., 2022a), etc. Depth cameras offer high precision and real-time capabilities; however, they are constrained by range, environmental factors, typically lower resolution, complex data processing, and higher costs. Similarly, LiDAR and 3D scanners provide high precision and accuracy in capturing detailed 3D point clouds, but they are limited by the need for high computational resources, elevated costs, and sensitivity to environmental conditions (Rongsheng et al., 2021; Yu et al., 2024). In contrast, RGB cameras are widely used due to the advantages of low cost, ease of use, reliable imaging, and strong environmental adaptability (Paproki et al., 2012; Wu et al., 2022b). However, 2D image data from RGB cameras often face challenges in addressing the complicated shielding between plant organs, and the accuracy of morphological parameter extraction is frequently limited by shooting angles. To overcome these limitations, researchers have developed the image sequence-based SFM-MVS (Structure-from-Motion, Multi-View Stereo) reconstruction technique, which generates a point cloud through multi-view image reconstruction, capturing rich and abundant 3D information of the plant and considerably improving measurement accuracy (Yang et al., 2024). This technique has been successfully applied to crop phenotyping, enabling the construction of high-precision 3D models and accurate extraction of phenotypic parameters for various crops such as corn (Wu et al., 2020; Li et al., 2022), cotton (Hao et al., 2024), soybean (He et al., 2023), sugar beet (Xiao et al., 2020, 2021), and more.

Compared to other plants, current 3D reconstruction techniques for wheat and rice are mainly applicable to early-stage plants with relatively simple structures or few tillers. The generalization and robustness of these techniques in dealing with the entire growth cycle of wheat still require improvement (Gu et al., 2024). In recent years, substantial progress has been achieved and accomplished in wheat phenotyping research. For example, Fu et al. (Fu et al., 2021) used Mask R-CNN technology to extract leaf length and plant height from 2D images of wheat seedlings, with coefficients of determination (R^2) of 0.87 and 0.98, respectively. Zheng et al. (Chenxi et al., 2022) manually labeled key points using a 3D digitizer to obtain 3D data of wheat and gleaned phenotypic parameters during standing, jointing, and heading stages, with R^2 for plant height, leaf length, and leaf width being 0.98, 0.87, and 0.75, respectively.

Additionally, studies based on SFM algorithms for multi-view image reconstruction have gained advanced developments. For example, Duan et al. (Duan et al., 2016) successfully reconstructed a 3D model of wheat at the seedling stage using the SFM algorithm, achieving an R^2 of 0.98 for leaf length. The MVS-Pheno V2 platform, developed by Wu et al (Wu et al., 2022b), utilized the SFM algorithm to focus on analyzing the shoots of low plants and acquired wheat point clouds from the beginning of returning to green stage to booting stage with notable success, obtaining R^2 for plant height, leaf length, and leaf width reaching 0.9991, 0.9949, and 0.9693, respectively. Combining with multi-view images and new virtual design method, Gu et al. (Gu et al., 2024) reconstructed a wheat plant model at heading stage, and attained plant height, crown width, plant leaf area, and coverage parameters, with an average R^2 of 0.78. Wheat exhibits complicated

changes throughout its growth cycle (Liu et al., 2023), such as dense tillers, thin and soft leaves, and cross-obscuration between organs, making phenotypic monitoring throughout its full cycle a challenging and arduous task. Although current research has yielded some results, existing methods still require advancement to continuously track wheat growth dynamics and accurately extract phenotypic parameters.

In view of this, this study designed a comprehensive SFM-MVS-based methodological workflow to process and segment wheat point clouds and automatically extract phenotypic parameters, covering the entire growth period (from seedling to grain-filling stages) for three varieties. First, we propose a new method for point cloud preprocessing that improves algorithm efficiency by using centroid-based segmentation to enhance the Euclidean clustering algorithm. Meanwhile, in the preprocessing workflow, noise is successfully removed through scale reduction, coordinate system correction, and combining color and statistical filtering algorithms, resulting in high-quality, 1:1 restored wheat point cloud models.

To enhance the efficiency of stem and leaf segmentation, a region growing algorithm was employed to automatically segment wheat stems and leaves, achieving notable success and improved segmentation efficiency to some extent. Important phenotypic parameters such as wheat plant height, leaf length, and leaf width were also extracted and compared with manual measurements for an in-depth assessment of the model reconstruction's accuracy. Ultimately, this study analyzed changes in growth dynamics throughout the growth cycle by extracting phenotypic parameters such as plant height, convex volume, plant surface area, and crown area, and ANOVA was performed according to different varieties in order to accurately demonstrate and differentiate the significant differences among different varieties of wheat at each growth stage.

2 Materials and methods

In this study, the 3D morphology of wheat plants with multiple life cycles was reconstructed using the SFM-MVS (Structure-from-Motion, Multi-View Stereo) algorithm, selecting three wheat cultivars with different plant characteristics as subjects. Main steps are introduced as follows (Figure 1):(1)Sample field-grown wheat plants and transplant them into pots without damaging their three-dimensional morphological structure; (2) Acquire multi-view image sequences of the plants using the image acquisition platform; (3) Reconstruct the three-dimensional point cloud using the SFM-MVS algorithm; (4) Restore the real-world scale of the point cloud by measuring the width of the marker stickers and rotate the point cloud to the real-world coordinate system using the RANSAC algorithm; (5) Use improved Euclidean clustering, color filtering, and statistical filtering to remove noise; (6) Perform stem-leaf segmentation using the region growing algorithm; (7) Down-sample the leaf point cloud and fit the leaf midrib with a local polynomial function; (8) Extract plant height, leaf length, and leaf width automatically to compare with the measured data to assess the accuracy of the wheat model; (9) Obtain crown width, convex volume, and plant surface area, and combine with ANOVA based on species effect to reveal differences and growth dynamics among different varieties and periods. Detailed descriptions of the specific experiments and the technologies and methods used are provided in the following sections.

2.1 Experimental environment

Three winter wheat cultivars with distinct plant characteristics were used in this study: FuMai1 (FM), PinHong1 (PH), and PinBai1 (PB). FM, a common commercial variety, has a compact plant type, moderate tillers, and average uniformity. PH and PB, intermediate varieties bred by the Fujian Academy of Agricultural Sciences, differ in their plant types; PH has a loose plant type with more tillers, while PB has a moderately compact plant type with fewer tillers. The experiment was conducted at the National Digital Plantation Innovation Center test field of the Institute of Agricultural Information, Chinese Academy of Agricultural Sciences (26°07′N, 119°20′E). Wheat was sown on November 27, 2023, at a density of 10 kg per mu. Each variety was planted in two plots, totaling six plots, with each plot divided into three rows spaced 30 cm apart. Adequate water and fertilizer were provided throughout the growth period. Plant samples were collected at the seedling, tillering, jointing, booting, heading, and grain filling stages, with four replicates per variety. The wheat samples were quickly transplanted into pots, and adequate water was added to prevent leaf wilting. Image sequences were collected after the wheat morphology stabilized.

An automated multi-view image acquisition platform, consisting of a rotating imaging arm, a lifting platform, supplementary lights, and a black background cloth, was self-constructed for the entire growth period of wheat (Figure 1A). Three RGB industrial cameras, each with 20 million effective pixels, were mounted on a rotating imaging arm, spaced 25 cm apart. The height of the plants was adjusted using a lifting platform to accommodate the cameras. The rotating arm with cameras rotated around the plants, capturing image sequences at intervals of 7° to 10° over a 360° range, with the camera and plant positions referenced, as shown in Figure 1A. After imaging, the plant height of the sampled wheat was measured immediately, and 2-3 leaves were randomly selected for measuring leaf length and width, with the leaves marked for subsequent verification.

2.2 SFM-MVS based 3D reconstruction and point cloud preprocessing

In this paper, Agisoft Metashape (version 2.1.0, Agisoft LLC, St. Petersburg, Russia), a software integrating SFM (structure from motion) and MVS (multi-view stereo) algorithms, was used for 3D point cloud reconstruction of wheat image sequences. Image-based 3D reconstruction converts a set of 2D images into a 3D model (Aharchi and Ait Kbir, 2020), utilizing SFM and MVS techniques, and is considered a powerful method (Kholil et al., 2021) capable of generating high-quality 3D models from 2D images.

SFM-MVS works as follows: first, the image sequence is analyzed by SFM to obtain the camera pose and create a sparse

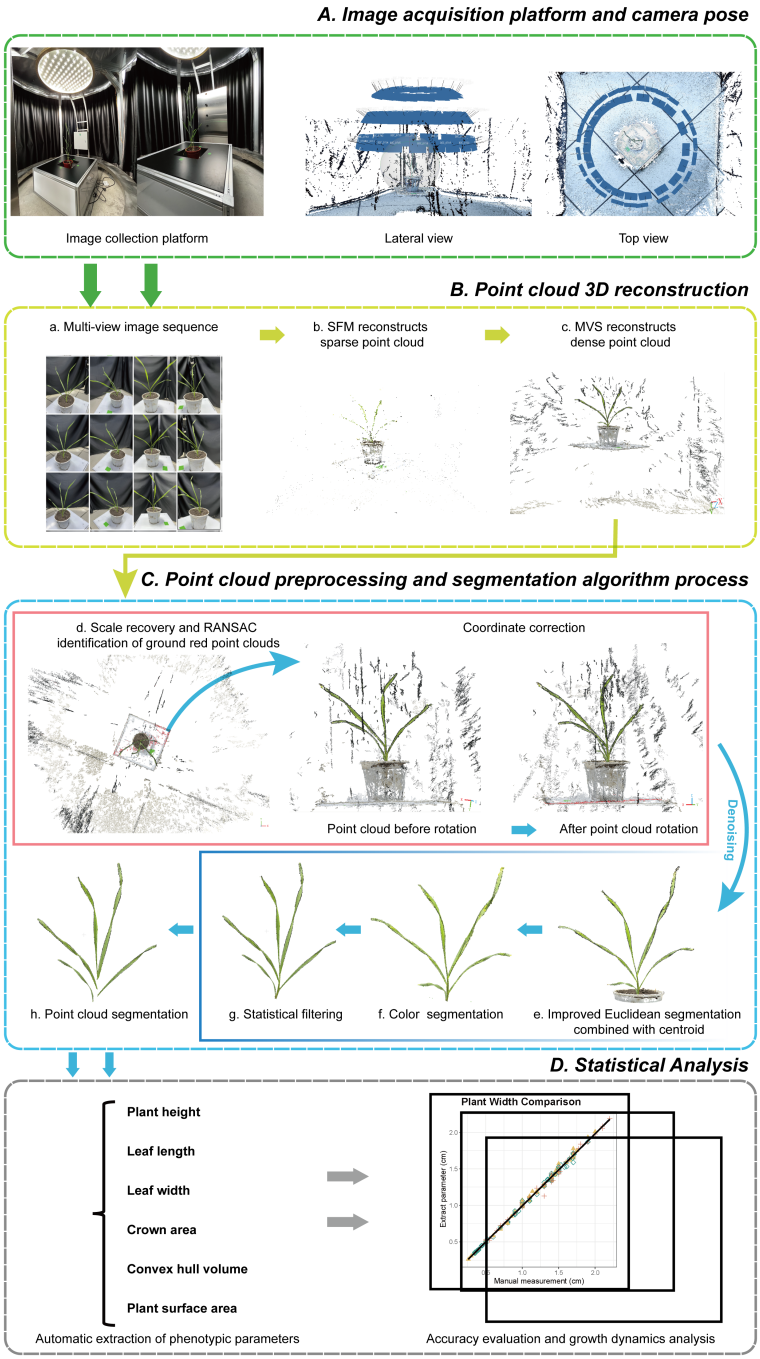


FIGURE 1
Overall process of phenotypic parameter extraction and analysis of wheat. **(A)** Image acquisition platform and camera pose. **(B)** Point cloud 3D reconstruction using SFM-MVS (a-c). **(C)** Preprocessing of point cloud data, including scale recovery, coordinate correction (d), noise removal (e-g), and segmentation (h). **(D)** Statistical analysis.

3D point cloud. This step involves feature detection, matching, and bundle adjustment to estimate the camera pose and the sparse 3D point cloud (Westoby et al., 2012). Subsequently, MVS generates a dense point cloud by estimating depth information for each pixel, utilizing multiple viewpoints to improve accuracy (Furukawa et al., 2015). This process includes techniques such as stereo matching for calculating pixel differences and generating depth maps, which are then fused to create a coherent representation of the scene geometry, resulting in high-quality 3D reconstruction.

During the point cloud reconstruction process, the generated point cloud usually differs in scale and coordinate system from the real world. To accurately restore the wheat point clouds to a 1:1 scale, two key steps were taken. First, the scaling factor was obtained by repeatedly measuring the width of marker stickers in the point

cloud and comparing it to the real sticker width, and then adjusting the scale of the point to match the real-world dimensions. Second, the ground is detected using the RANSAC (Random Sample Consensus) algorithm (Fischler and Bolles, 1981), and the rotation matrix of the point cloud is calculated based on the detection results. By rotating the point cloud to the real-world coordinate system, it was ensured that the plants were aligned along the Z-axis from top to bottom (Figure 1C.d).

Additionally, during the wheat image data acquisition process, numerous noise points may exist in the generated point cloud due to environmental influences and equipment accuracy. External factors such as view obscuration and obstacles can also result in unrelated discrete points and other object point clouds (Figures 1B, C, Figure 1C.d). To eliminate these irrelevant points and improve subsequent point cloud processing speed, an improved Euclidean clustering algorithm combined with centroid calculation was proposed to segment the main plant point cloud (Figure 1C.e). The RGB frequency distribution of the segmented point cloud was calculated (Figure 2), categorizing the point cloud into three clusters: soil, pot, and plant, from left to right in the R-B distribution cluster. A boundary threshold of G-R=5 and G-B=30 was set to remove most unrelated points connected to soil and pots (Figure 1C.f). Finally, statistical filtering was applied to remove outliers from the plant point cloud (Figure 1C.g). Based on above steps, a high-quality wheat point cloud was acquired, with the preprocessing algorithm workflow used in this study shown in Figure 1C.

2.3 Point cloud scaling and coordinate system correction

To accurately restore the wheat point clouds to a 1:1 scale, a scaling factor S was calculated and applied to match the real-world scale. This is done by taking the average of three measurements of the width of the marker sticker in the point cloud to obtain W_{virtual} and comparing it to the real sticker width W_{real} (3cm). The scaling factor S is determined using Equation 1.

$$S = \frac{W_{\text{real}}}{W_{\text{virtual}}} \tag{1}$$

The original coordinates $(X_{\text{old}}, Y_{\text{old}}, Z_{\text{old}})$ of each point were scaled according to Equation 2.

$$(X_{\text{scaled}}, Y_{\text{scaled}}, Z_{\text{scaled}}) = (X_{\text{old}} \times S, Y_{\text{old}} \times S, Z_{\text{old}} \times S) \tag{2}$$

Next, coordinate system correction was performed by detecting the ground using the RANSAC algorithm to obtain the normal vector $\mathbf{n}_1$ of the fitted plane (ground). The normal vector $\mathbf{n}_2$ of the reference plane was set as $(0, 0, 1)$, representing the Z-axis direction. The rotation matrix R, was calculated to align the fitted plane with the reference plane using Equation 3.

$$R = \frac{\mathbf{n}_2 \times \mathbf{n}_1}{\| \mathbf{n}_2 \times \mathbf{n}_1 \|} \tag{3}$$

Each point in the point cloud, represented in homogeneous coordinates $(x, y, z, 1)$, was transformed using the rotation matrix R to achieve coordinate system correction, ensuring alignment with the reference plane. The transformation process is outlined in Equation 4.

$$(X_{\text{rotated}}, Y_{\text{rotated}}, Z_{\text{rotated}}) = (X_{\text{scaled}}, Y_{\text{scaled}}, Z_{\text{scaled}}) \times R \tag{4}$$

This resulted in a rotated point cloud, with the plant oriented along the Z-axis and the ground aligned with the XOY plane, ensuring accurate dimensions and coordinates for the reconstructed wheat point cloud model, providing a reliable foundation for subsequent analysis.

2.4 Improved Euclidean clustering algorithm combined with centroid

When processing the rotated point cloud, a considerable portion of points, such as those from the ground, lifting platform, and pots, are unrelated to the plant and constitute over 90% of the overall point cloud. Direct application of Euclidean clustering to the entire point cloud leads to extensive computational demands, prolonged processing time, and high computer performance requirements. Therefore, an improved Euclidean clustering algorithm, combined with centroid calculation, is proposed to

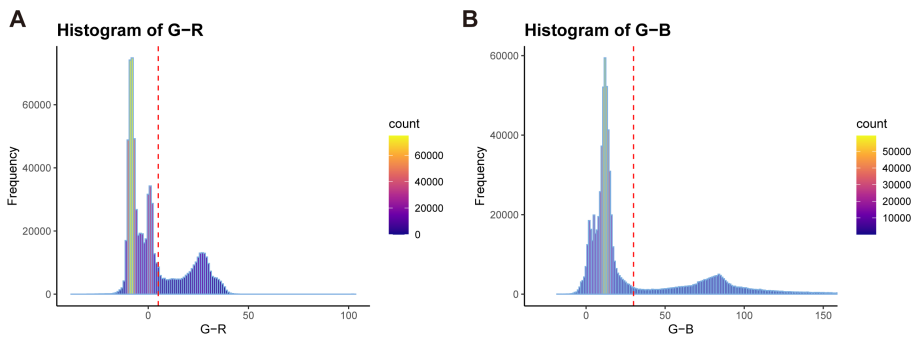


FIGURE 2
RGB color distinction frequency distribution of wheat point cloud. **(A)** G-R distribution with red dotted lines indicating G-R=5. **(B)** G-B distribution with red dotted lines indicating G-B=30.

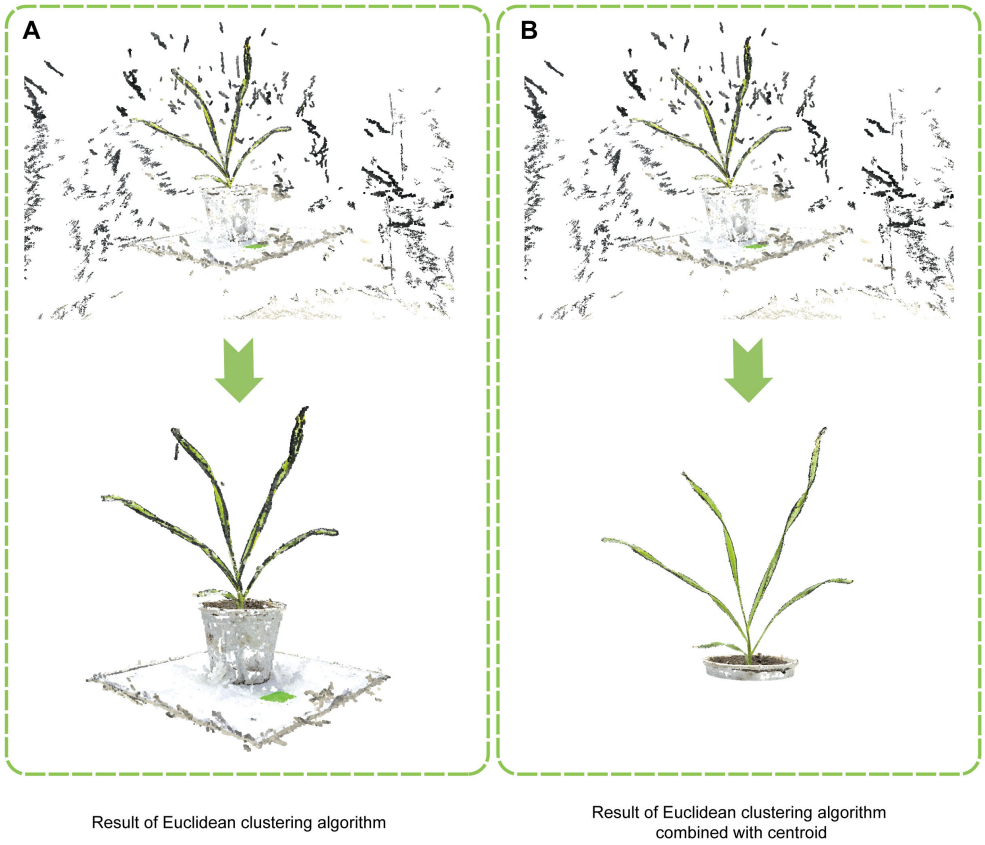


FIGURE 3
Comparison of wheat point cloud segmentation methods. **(A)** Result of the traditional Euclidean clustering algorithm. **(B)** Result of the improved Euclidean clustering algorithm combined with centroid.

markedly enhance segmentation efficiency. Traditional Euclidean clustering (Sun et al., 2020; Miao et al., 2023) is based on Euclidean distance metrics, where the core idea is to determine distances between points by distinguishing their proximity to neighbors, thereby clustering the spatial point cloud. To accelerate the Euclidean clustering algorithm, KD-Tree nearest neighbor search is often employed as a crucial preprocessing step. The specific steps are as follows: randomly selecting an initial point P, finding k nearest neighborhood points of P through a KD-Tree, calculating the Euclidean distance between each neighborhood point and P, and clustering points with distance smaller than the presetting threshold (ε) into the set Q. This process is iterated until the number of elements in Q stabilizes.

The improved Euclidean clustering algorithm combined with centroid calculation follows these steps:

(1) Perform Voxel Down-sampling: Apply voxel down-sampling (Sun et al., 2022) on the rotated point cloud to reduce density and computational load.

(2) Detect Ground and Calculate Centroids: Use the RANSAC algorithm to detect and mark the ground point cloud set and calculate its centroid $(x_{ground}, y_{ground}, z_{ground})$. Use color filtering to isolate the majority of the plant point cloud set and calculate its centroid $(x_{plant}, y_{plant}, z_{plant})$. (3) Determine the cut-off threshold: The cut-off threshold $Z_{threshold}$ is calculated using Equation 5.

$$Z_{threshold} = z_{ground} + \alpha(z_{plant} - z_{ground}), \qquad 0 \le \alpha \le 1 \qquad (5)$$

This threshold effectively isolates plant points from non-plant points, ensuring accurate separation.

(4) Preliminary Segmentation: Remove points below the $Z_{threshold}$ to eliminate the ground, lifting platform, and most pot points.

(5) Secondary Segmentation: Use the Euclidean clustering algorithm for secondary segmentation of the point clouds after preliminary segmentation to obtain the accurate plant point cloud set.

Compared to the original algorithm (Figure 3A) , the improved Euclidean clustering algorithm incorporated centroid calculation and Euclidean clustering in a two-step segmentation process, and the implementation of adaptive determination of the cut-off threshold allowed for preliminary segmentation, reducing the volume of point cloud data (Figure 3B). After that, Euclidean clustering was used for fine segmentation, effectively removing irrelevant points and markedly enhancing segmentation efficiency and accuracy. These improvements enabled the efficient removal of

most non-plant points, facilitating more effective Euclidean segmentation of the rotated point cloud.

2.5 Region growing segmentation algorithm

In this study, the region growing segmentation algorithm (Vo et al., 2015; Li et al., 2018) is employed to automatically segment wheat stems and leaves, drastically reducing the labor-intensive manual segmentation process and greatly improving the efficiency of obtaining leaf phenotypes. The region growing segmentation algorithm operates by comparing the angles between normals and curvatures of points, merging points that are sufficiently close under smoothness constraints, thereby combining similar points into regions for effective point cloud segmentation. The algorithm's workflow is as follows:

Step 1: Calculate normals and curvatures for all points, then sort the points based on their curvature values.

Step 2: Select the point with the smallest curvature as the initial seed point. Compare the normal angles and curvature differences of neighboring points. Points meeting the predefined normal angle and curvature difference thresholds are added to the seed point's region, while points meeting only the normal angle threshold are classified separately. Finally, all points were traversed to segment the point cloud into multiple regions with similar attributes.

2.6 3D phenotype extraction of wheat

Using pre-processed wheat point clouds, 3D phenotypic parameters of wheat plants were extracted, including plant height, leaf length, leaf width, crown area, plant surface area, and convex volume. Among these, plant height, leaf length, and leaf width during various growth stages were used to quantitatively assess the accuracy of the reconstructed point clouds.

Plant height serves as a crucial metric for evaluating the scaling precision of the reconstructed point clouds (Wu et al., 2022b). Crown area, plant surface area, and convex volume are essential indicators for monitoring wheat growth dynamics (Che et al., 2020; Paturkar et al., 2020). The selection of a suitable calculation method is crucial for obtaining accurate values of wheat phenotypic parameters. In this study, the following definitions and methods were used to measure key phenotypic traits:

Plant height H is defined as the vertical distance from the base of the stem at the soil interface to the highest point of the plant, represented by the height difference between the maximum and minimum Z-axis values of the plant point cloud (Figure 4A), given by Equation 6.

$$H = Z_{\max} - Z_{\min} \tag{6}$$

Crown area C is defined as the product of the maximum horizontal distance of the wheat canopy projected onto the ground and its perpendicular vertical distance, which corresponds to the base area of the bounding box in Figure 4A. The relevant formula is provided in Equation 7.

$$C = L_{\max} \times W_{\max} \tag{7}$$

Plant surface area refers to the total surface area of the wheat stems and leaves, calculated by reconstructing the surface area from the point cloud. The calculation process is as follows (Figure 4E): Voxel down-sampling of point clouds was applied (Voxel Size=0.01) to the point cloud to reduce computational load and ensure uniformity. Subsequently, the triangular mesh reconstruction was then performed using the Ball Pivoting (Bernardini et al., 1999). Finally, we calculated the sum of the areas of all triangular facets after reconstruction as the plant surface area. The left side of Figure 4D illustrates the originally colored point cloud and the rendered plant after surface reconstruction. Convex volume is estimated by calculating the convex hull of the plant point cloud (Wu et al., 2022b), which forms the outermost convex polygon or polyhedron enclosing the point cloud (Figure 4B).

Leaf length refers to the maximum surface distance from the leaf base to the leaf tip, and leaf width is the maximum surface width perpendicular to the leaf length. The automatic acquisition of leaf length is achieved by fitting the leaf midrib as follows (Figure 4C): (1) Apply voxel down-sampling to the segmented individual leaf point cloud to reduce computational load and ensure uniformity. (2) Conduct principal component analysis (PCA) on the sampled leaf point cloud coordinates to obtain the direction of the principal axis, then rotate the leaf point cloud so that the principal axis aligns with the X-axis, and also determining the leaf base and tip locations. (3) Use a local polynomial regression algorithm to fit the leaf midrib (Duan et al., 2016) and calculate its length. Additionally, the maximum width perpendicular to leaf length is chosen by the interactive selected point measurement method (Li et al., 2022), representing the leaf width (Figure 4C).

2.7 Statistical accuracy evaluation indicators

To evaluate the accuracy of point cloud parameter extraction and model precision, we calculated several statistical indicators including the Coefficient of Determination (R^2), Root Mean Square Error ($RMSE$), and Relative Root Mean Square Error (RRMSE). These are defined in Equations 8-10.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{8}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{9}$$

$$RRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\frac{1}{n}\sum_{i=1}^{n} y_i} \tag{10}$$

Where y_i is the observed value (manual measurement), and $\hat{y}_i$ is the predicted value (model extracted value).

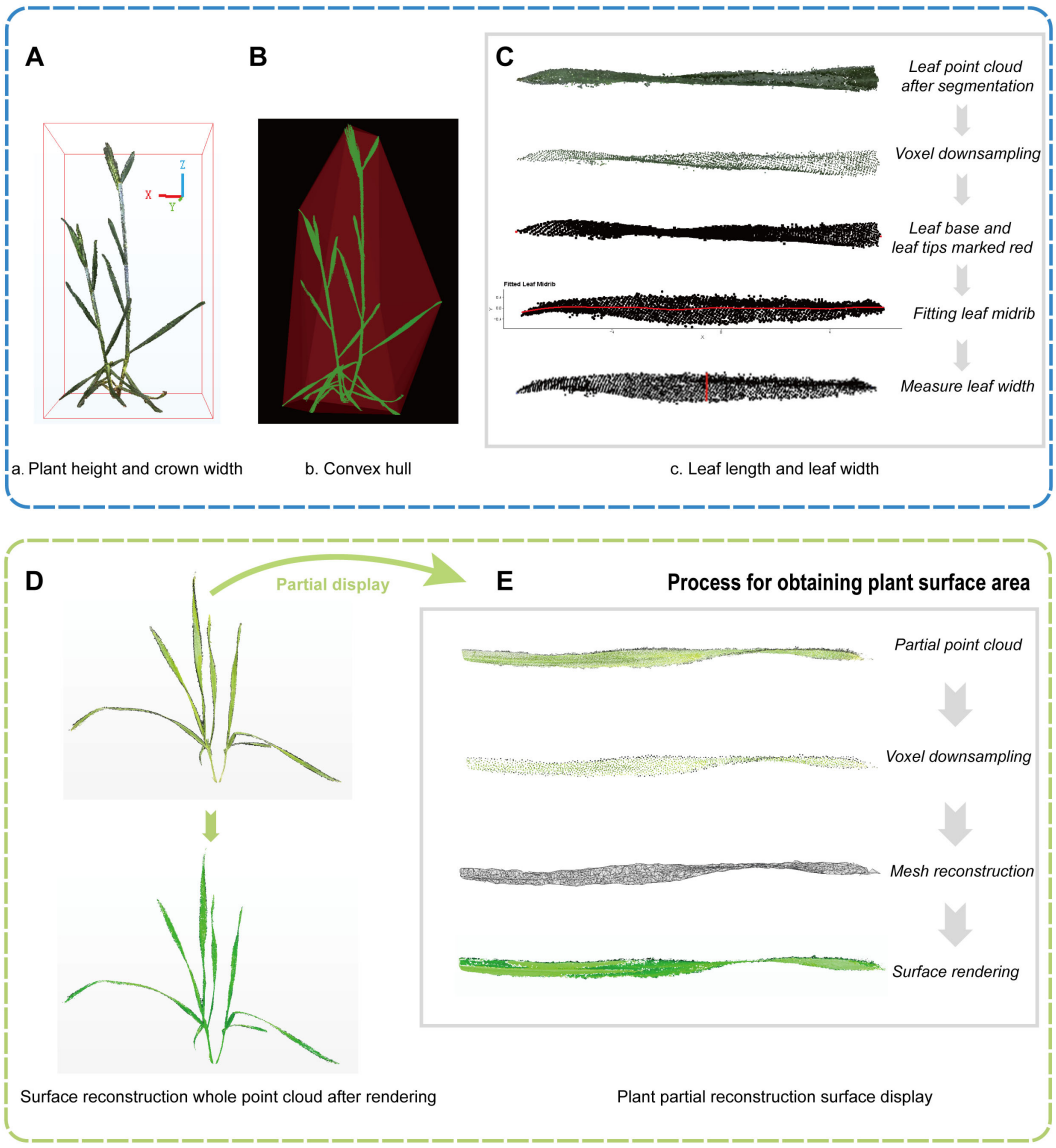


FIGURE 4
Automated extraction of 3D phenotypes in wheat. **(A)** Plant height and crown width. **(B)** Convex hull. **(C)** Leaf length and width measurement, including steps from leaf point cloud segmentation to measuring leaf width. **(D)** Surface reconstruction of the wheat plant. **(E)** Detailed view of the reconstructed surface.

3 Results

3.1 Reconstruction and preprocessing of wheat point clouds

According to the preprocessing workflow of point clouds mentioned in 2.2, the point cloud data of three wheat species were meticulously processed throughout their entire growth periods, including seedling, tillering, jointing, booting, heading, and grain filling stages. This comprehensive process included scale restoration, coordinate system correction, improved Euclidean clustering segmentation, color filtering, and statistical filtering. Subsequently, the generated plant point clouds underwent triangular surface reconstruction and rendering, ensuring accurate representation of plant characteristics at each stage. The reconstruction results of the three wheat species at various stages are shown in Figure 5, providing a clear visualization of the 3D structures and morphological changes of the wheat plants throughout their growth periods.

3.2 Region growing segmentation results

The region growing segmentation algorithm performed well in segmenting wheat leaves with simple and non-overlapping structures. Specifically, during the seedling stage, the algorithm accurately separated each leaf due to the small number of leaves and simple structure, as shown in Figure 6A, consistent with the findings of Duan et al (Duan et al., 2016; Li et al., 2022). Even during the heading and grain filling stages, despite an

10.3389/fpls.2024.1459968

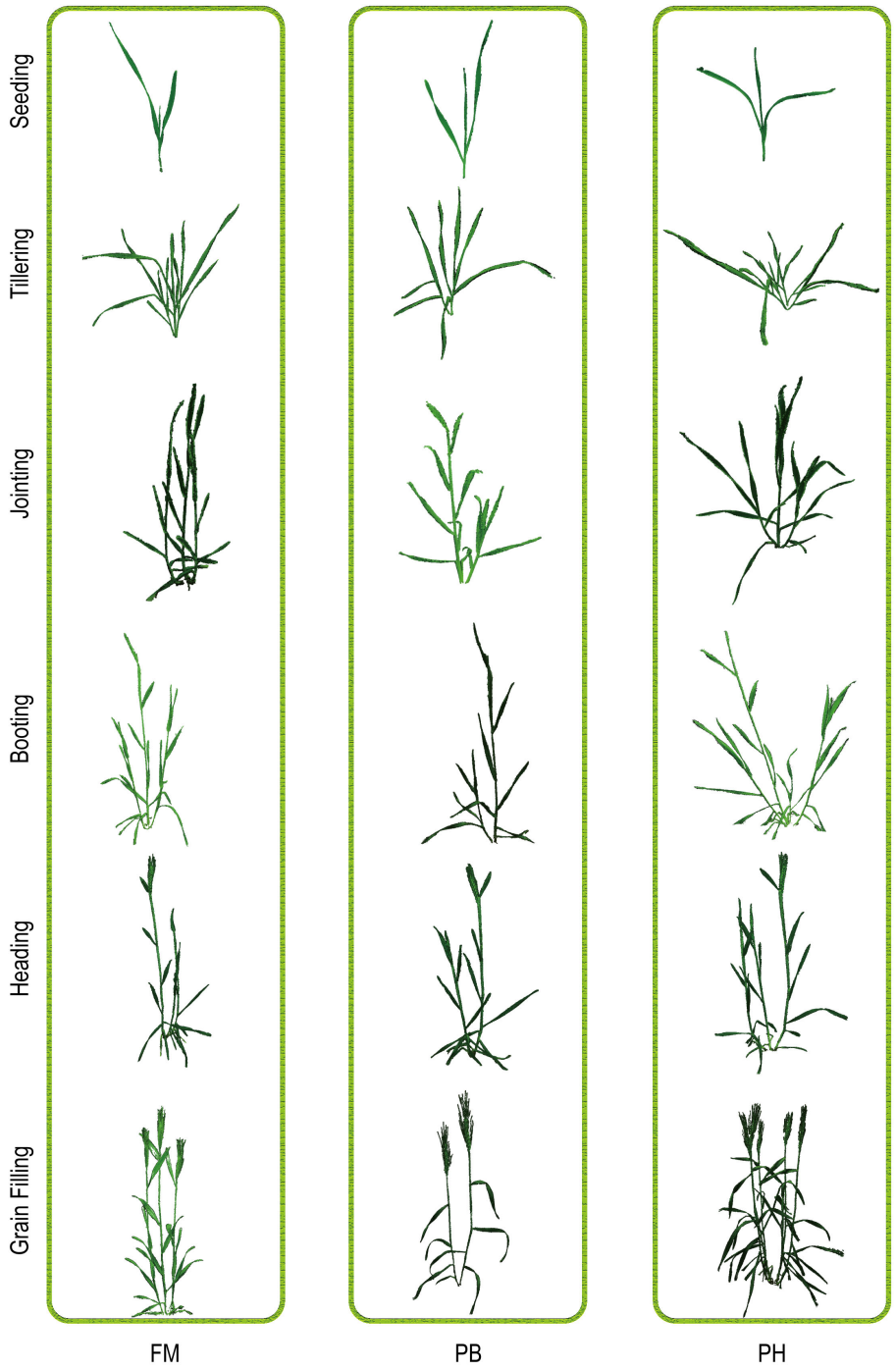


FIGURE 5
Visualization of point clouds for three wheat cultivars (FM, PH, PB) across different growth stages: seedling, tillering, jointing, booting, heading, and grain filling.

increase in leaf number, the algorithm remained effective due to the pronounced leaf extension (Figure 6B). However, during the tillering and jointing stages (Figures 6C, D), segmentation performance was compromised by high leaf density, severe overlapping, and leaf curling. The algorithm could not fully segment all leaves, achieving only partial segmentation of complete leaves.

Manual segmentation was employed to supplement these unsegmented parts, resulting in more complete and accurate segmentation outcomes. Although manual intervention was required, the overall efficiency of wheat leaf segmentation was considerably improved. In summary, the region growing segmentation algorithm exhibits limitations when dealing with complex overlapping leaf structures, particularly during specific

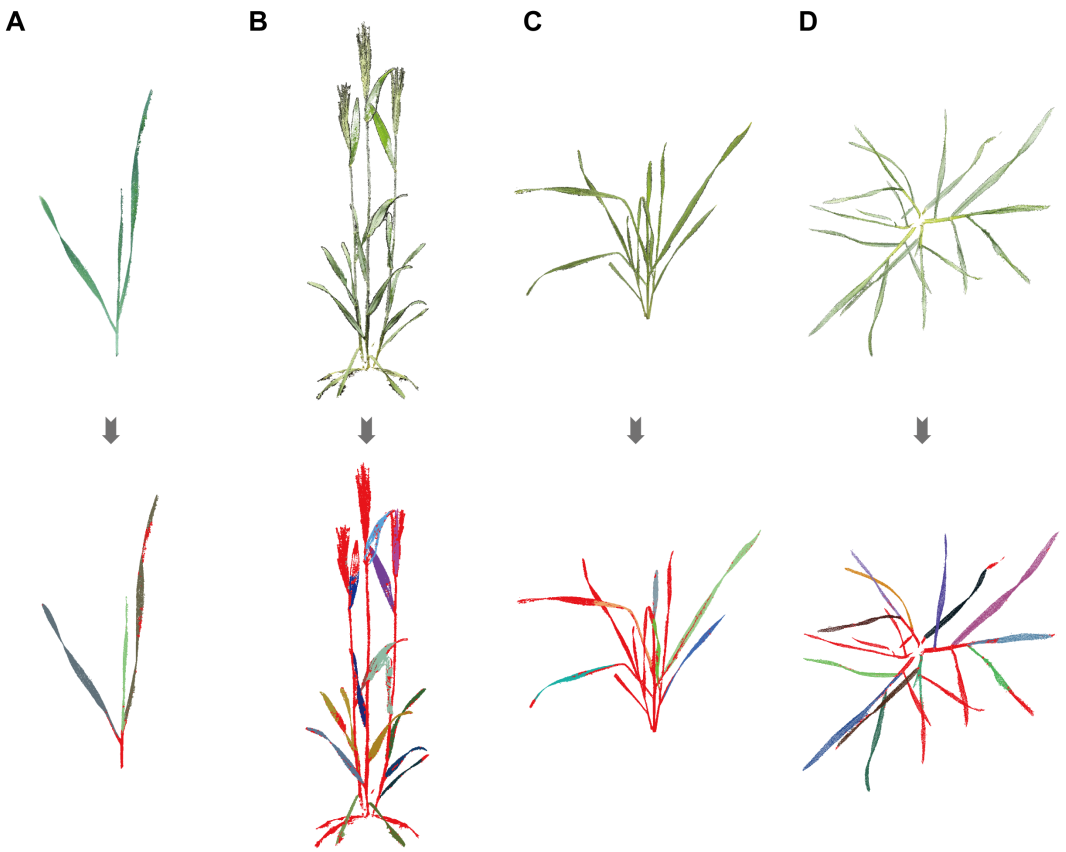


FIGURE 6
Region growing segmentation results of wheat point clouds: **(A)** Seedling, **(B)** Grain filling, **(C)** Tillering, **(D)** Jointing. Leaves are marked with different colors, and red point clouds represent unclassified points.

growth stages. Nonetheless, combining this algorithm with manual segmentation can remarkably enhance segmentation efficiency and accuracy, laying a solid foundation for subsequent leaf phenotypic extraction.

3.3 Accuracy evaluation of 3D wheat point clouds

Using the phenotypic extraction methods described in section 2.6, we automatically extracted and compared 72 sets of plant height, 160 sets of leaf length, and 160 sets of leaf width with manually measured values (Figure 7) to quantitatively assess the accuracy of point cloud parameter extraction and model precision. The results demonstrated a significant linear relationship between the automatically extracted plant height, leaf length, and leaf width from the wheat point clouds and the manual measurements. The coefficients of determination (R^2) were 0.9979, 0.9977, and 0.995, respectively; the root mean square errors (*RMSE*) were 1.0773 cm, 0.2612 cm, and 0.0335 cm, respectively; and the relative root mean square errors (*RRMSE*) were 2.1858%, 1.7483%, and 2.8462%, respectively.

By comparing point cloud measurements with manual measurements, the proposed algorithm demonstrated high accuracy in multi-view imaging reconstruction and phenotypic extraction of wheat. This also validated, to a certain extent, the practicality and stability of using 3D imaging to extract crop phenotypic parameters, enabling rapid and non-destructive extraction of plant phenotypic values.

3.4 Evaluation of wheat growth dynamics and variance analysis

The average plant height, crown area, convex volume, and plant surface area of different wheat cultivars at various growth stages were meticulously extracted and calculated to reflect the growth dynamics of each variety (Figure 8). Meanwhile, a quantitative analysis of differences among wheat species in term of plant height, canopy area, convex hull area and plant surface area was carried out through one-way analysis of variance (ANOVA) (Figure 9).

Regarding plant height, there were no significant differences among the three wheat cultivars across these growth stages (Figure 9A, $p > 0.05$). Plant height increased progressively throughout the growth cycle, with the most notable growth occurring from the jointing stage to the booting stage (Figure 8A). In contrast, growth from the heading stage to the grain filling stage was less pronounced.

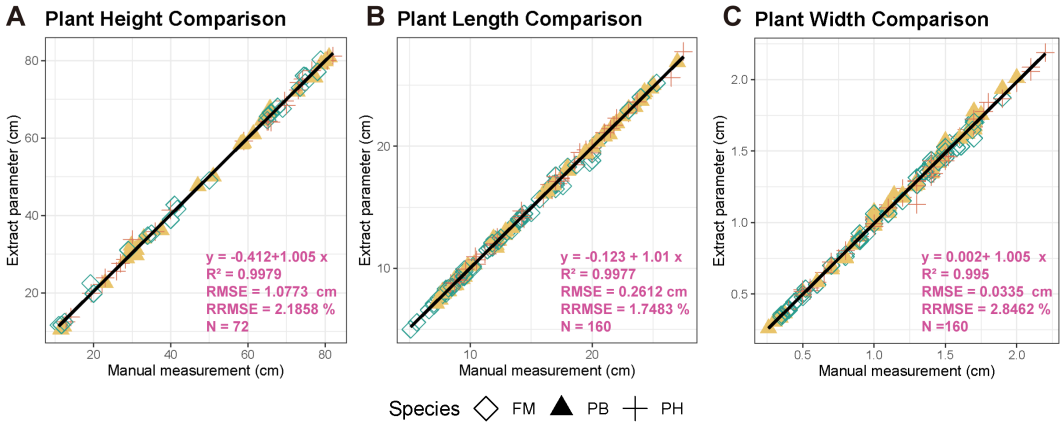


FIGURE 7
Accuracy evaluation of extracted wheat point cloud parameters. **(A)** Plant height comparison. **(B)** Plant length comparison. **(C)** Plant width comparison.

For crown area, significant differences were observed between PH and both FM and PB ($p \leq 0.01$), with PH demonstrating a larger crown area, while FM and PB showed no significant differences (Figure 9B, $p > 0.05$). The crown area followed a trend of initially increasing, then decreasing, and subsequently increasing again throughout the growth cycle (Figure 8B). The turning points for PH and PB were at the jointing and heading stages, with the maximum crown area reached at the jointing stage. Conversely, FM's turning points were at the booting and heading stages, with the peak at the booting stage.

When examining convex volume, significant differences were noted ($p \leq 0.05$), with PH having a larger convex volume compared to both FM and PB, while FM and PB showed no significant differences (Figure 9C, $p > 0.05$). The convex volume for each

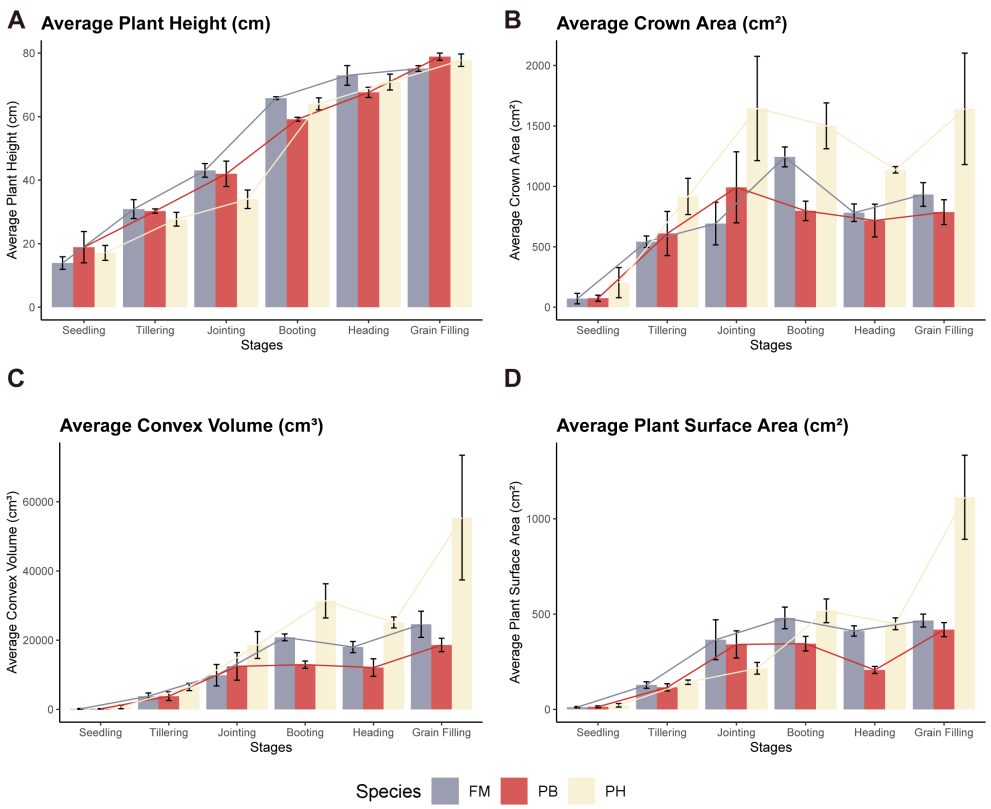


FIGURE 8
Dynamic growth changes in wheat cultivars. **(A)** Average plant height, **(B)** Average crown area, **(C)** Average convex volume, **(D)** Average plant surface area across different growth stages.

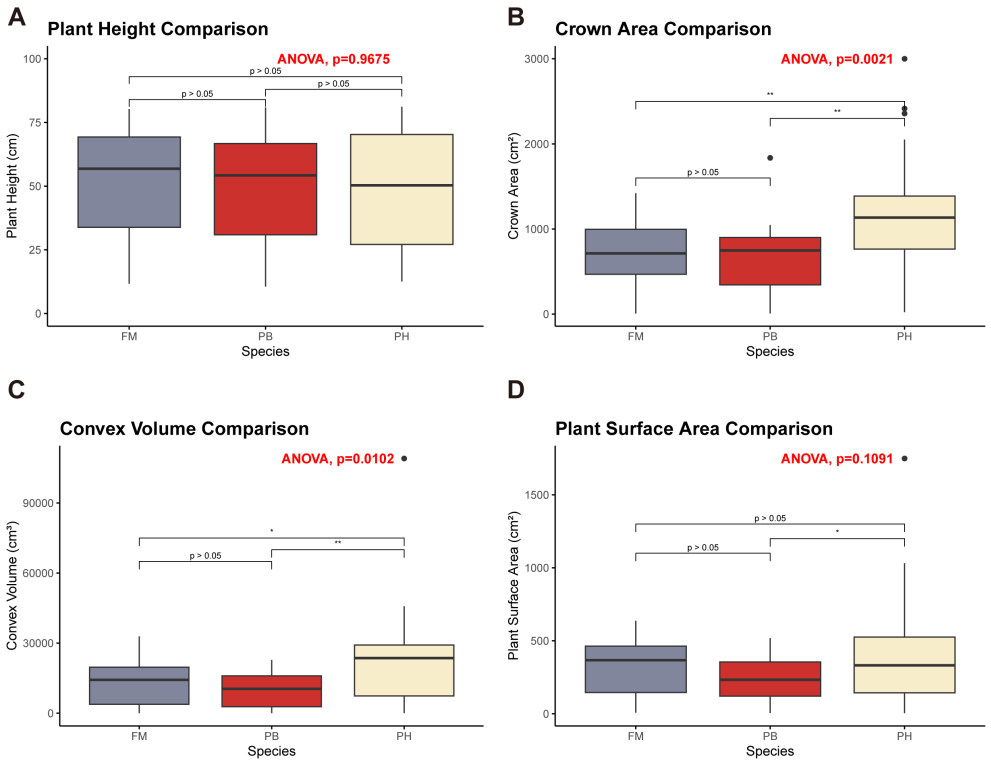


FIGURE 9
ANOVA results comparing different wheat cultivars. **(A)** Plant height comparison. **(B)** Crown area comparison. **(C)** Convex volume comparison. **(D)** Plant surface area comparison. Asterisks indicate levels of statistical significance: * $p \leq 0.05$, ** $p \leq 0.01$.

wheat variety exhibited a pattern of increasing, then decreasing, and increasing once more throughout the growth cycle, with turning points at the booting and heading stages, reaching a maximum at the grain filling stage (Figure 8C).

In terms of plant surface area, significant differences were identified between PH and PB ($p \leq 0.05$), with PH having a larger plant surface area (Figure 9D). No significant differences were observed among the other combinations. The plant surface area of each wheat variety showed an initial increase, followed by a decrease, and another increase throughout the growth cycle, with turning points at the booting and heading stages (Figure 8D). PH and PB had the largest plant surface area during the grain filling stage, whereas FM reached its largest plant surface area at both the booting and grain filling stages, with minimal difference between the two stages.

Consequently, the proposed methodology effectively captures phenotypic differences across different growth stages and wheat cultivars, enabling dynamic monitoring of wheat growth status.

4 Discussion

With considerations to complicated features (e.g. multiple tillers, slender leaves, and organ cross-obscuration) during wheat growth, a set of method and procedure based on multi-view geometric technology was proposed in this study to process and segment wheat point clouds and automatically extract phenotypic parameters. The accuracy of phenotypic parameters for three wheat cultivars, which exhibited distinct characteristic differences throughout the growth period, was analyzed, and their growth dynamics were compared. The results demonstrate that this methodology is highly accurate and consistent, effectively monitoring phenotypic changes across the entire life cycle of wheat and the phenotypic differences among different varieties. Those discoveries have profound implications for wheat breeding and plant architecture development as well. Additionally, the study highlights the substantial potential of multi-view imaging and 3D reconstruction techniques for monitoring wheat growth dynamics and extracting high-precision phenotypic traits. Based on these findings, a detailed discussion is provided to explore the implications, limitations, and future research directions.

4.1 Significance of full-cycle monitoring of wheat phenotypes based on multi-view geometry

In the field of botany, traditional methods of plant phenotyping are destructive, time-consuming, inefficient, and costly, making non-destructive, rapid, and accurate acquisition of plant

phenotypes a current research focus (Xie and Yang, 2020; Yang et al., 2020). Compared to 2D images, 3D reconstruction-based phenotypic extraction can express richer and more detailed 3D information (Yang et al., 2024). With the development of 3D reconstruction technology, remarkable progress has been made in crop 3D reconstruction studies using various sensors (Gibbs et al., 2016). Among these, using RGB cameras to capture multi-view images and performing 3D reconstruction based on SFM-MVS is widely applied by virtue of its low cost, strong environmental adaptability, and the capability to generate colored point clouds. This method has been successfully applied to extract high-precision phenotypic parameters for crops such as maize (Wu et al., 2020; Li et al., 2022), cotton (Hao et al., 2024), soybean (He et al., 2023), and sugar beet (Xiao et al., 2020, 2021), and others.

Compared to other crops, wheat's growth characteristics, including dense tillers, slender leaves, and overlapping organs, make full-cycle phenotypic monitoring more complex and challenging. Previous studies on wheat phenotypic parameter extraction have used different methods. One method involves directly extracting phenotypic parameters from 2D images, such as using Mask R-CNN (Fu et al., 2021) to extract leaf length and plant height of seedling-stage wheat, with R^2 of 0.87 and 0.98, respectively. Another method involves high-precision instruments, such as LiDAR (Sun, Li et al., 2018; Forero et al., 2022; Liao et al., 2022) and 3D digitizers (Chenxi et al., 2022; Wu et al., 2022a). Zheng et al. (Chenxi et al., 2022) manually annotated key points to obtain 3D data of wheat, extracting phenotypes at different stages with R^2 of 0.98, 0.87, and 0.75 for plant height, leaf length, and leaf width, respectively. This study focuses on a simpler, cost-effective image-based SFM reconstruction method by integrating low-cost RGB cameras into our phenotyping platform. Compared to LiDAR and digitizers, this method not only captures the plant's morphological structure but also its color characteristics, reflecting growth conditions and ensuring crop health. Our method extracts richer and more abundant information than 2D images, achieving higher accuracy and consistency in extracting plant height, leaf length, and leaf width, with R^2 of 0.9979, 0.9977, and 0.995, respectively. Next, we focused more on the application of Structure from Motion (SFM) in wheat reconstruction. Although previous studies have made notable progress in specific stages such as the seedling stage (Duan et al., 2016) (with an R^2 of 0.98 for leaf length), the greening to booting stages (Wu et al., 2022b) (R^2 of plant height, leaf length and leaf width: 0.9991, 0.9949 and 0.9693), and the heading stage (Gu et al., 2024) (average R^2 of plant height, crown width, total leaf area and coverage: 0.78), they have yet to comprehensively characterize the growth dynamics of wheat throughout the entire growth period.

Compared to previous studies, this study used more data spanning the entire growth cycle from seedling to grain filling. By combining various algorithms for point cloud processing and phenotypic extraction, more precise and efficient wheat phenotypic extraction was achieved, comprehensively monitoring growth changes and significant differences throughout the growth period. This is of practical importance for wheat breeding and plant architecture development.

4.2 Accuracy assessment and error analysis

Through quantitative accuracy evaluation (Figure 7), a highly consistent linear relationship was identified between the wheat plant parameters extracted from the 3D point cloud model and those obtained through manual measurements. This result not only validates the reliability and completeness of the proposed algorithm for processing wheat point clouds but also confirms the accuracy of the automatic methods for obtaining plant height, leaf length, and leaf width through relatively low RMSE and RRMSE values. Compared to previous studies on wheat phenotyping, this study demonstrates a high level of accuracy, with R^2 values for all phenotypes exceeding 0.995. However, RMSE still can be improved to some extent (RMSE of plant height, leaf length and leaf width values 1.0773cm, 0.2612cm and 0.0335cm, respectively). With comprehensive and thorough considerations to experimental process, processing flow of the algorithm and previous research results, the differences in wheat phenotypic parameters may stem from several factors: 1) measurement errors in the actual data; 2) potential reconstruction deficiencies in the leaf edge regions during the point cloud reconstruction process (Duan et al., 2016); 3) possible measurement errors during the point cloud scale recovery process; 4) the potential filtering out of some plant points during the point cloud preprocessing workflow; and 5) errors in the phenotypic extraction algorithm. Optimizing these aspects is crucial for improving accuracy and represents a key direction for future work.

Furthermore, during the experiments, various quantities and distributions of images were tested to improve the resolution and quality of the wheat 3D point cloud reconstruction. The results indicated that an excessive number of images (>200) did not enhance the reconstruction resolution and sometimes resulted in poorer outcomes; conversely, an insufficient number of images (<50) led to fewer matching points, making it impossible to reconstruct the wheat plants. Considering the different characteristics of the wheat growth cycle, we recommend capturing 60-80 images during periods with fewer tillers, simpler structure and less cross-shading, and 100-140 images during periods with more tillers, more complicated structures and more cross-shading, aiming to get better reconstruction results.

4.3 Improvement of Euclidean clustering segmentation

Noteworthy advancements were achieved in the improvement of Euclidean clustering segmentation in this study. During the preprocessing of point clouds, the initial point cloud was found to contain a large number of irrelevant points, such as ground, pots, and lift platforms, accounting for more than 90% of the point cloud. Directly applying the traditional Euclidean clustering algorithm (Li et al., 2022) posed challenges, including high computational demands, time consumption, and substantial computer performance requirements. Hence, an improved Euclidean clustering segmentation algorithm was proposed, which remarkably enhanced segmentation efficiency and reduced algorithm runtime by incorporating centroid

calculations. This improvement facilitated faster and more efficient processing of large-scale wheat point cloud data.

4.4 Growth dynamics monitoring and prospects

In assessing growth dynamics, ANOVA was introduced alongside traditional crop growth dynamics characterization to demonstrate significant differences among different wheat cultivars at various growth stages (Figures 8, 9). For instance, PH exhibited greater crown area, higher convex hull volume and larger plant surface area than the rest two cultivars. This analysis highlights the structural differences among wheat cultivars. Moreover, trends in plant height, crown area, convex volume, and plant surface area were observed throughout the growth cycle, closely linked to physiological changes during development. Non-destructive and high-throughput phenotypic analysis of wheat plants enables early monitoring of crop structural morphology and growth conditions, providing valuable insights for wheat breeding and variety development. Our method allows for precise, automated, and high-throughput phenotyping of key growth metrics, offering breeders accurate and detailed growth data. This method can also be combined with genomic data to explore the associations between phenotypes and genotypes, thereby enhancing the selection process for superior genotypes, accelerating breeding cycles, and improving trait selection. In the realm of precision agriculture, this method can be integrated with UAV systems and ground-based sensing platforms to extract crop phenotypes and monitor crop health and development (Varela et al., 2021). Experts can use this growth and development data to provide timely interventions and prescriptions based on specific crop needs, optimizing resource use and increasing yield (Araus et al., 2022). The ability to quantify growth dynamics and trait variation also provides valuable data for crop modeling, precision agriculture, and decision support systems. Future research could focus on integrating phenotypic analysis with plant structure-function models, coupling multiple models to construct crop digital twins (Nasirahmadi and Hensel, 2022; Purcell and Neubauer, 2023), and combining crop prescription decisions to offer more precise decision support for optimizing agricultural production.

4.5 Limitations and prospects

The effectiveness of region growing segmentation was suboptimal during certain growth stages of wheat. Although the introduction of region growing segmentation algorithms improved segmentation efficiency and achieved some success during seedling and heading stages, overall segmentation performance was still not very satisfying during periods with extensive tillers, cross-obscuration between organs, and greater curling angles (tillering, jointing, and booting stages). The algorithm could only partially segment complete wheat leaves; moreover, achieving satisfactory segmentation results required iterative adjustments of the point cloud segmentation thresholds, which was a time-consuming process. These findings exhibit congruity with current advancements in traditional segmentation algorithms for plants. Traditional segmentation algorithms perform well on plants with few or no tillers, such as seedling wheat (Duan et al., 2016), barley (Wahabzada et al., 2015), seedling maize (Li et al., 2022), etc. However, achieving fully automated segmentation for structurally complex crops like wheat and rice remains a major challenge. Future research will focus on addressing this issue by exploring the use of neural network models (e.g., PointNet (Qi et al., 2017a), PointNet++ (Qi et al., 2017b), PointCNN (Li et al., 2018), KPConv (Kernel Point Convolution) (Thomas et al., 2019) for point cloud segmentation. Building large-scale wheat point cloud datasets for model training and optimization will help develop more effective and targeted segmentation algorithms, aiming for fully automated segmentation and improved efficiency in wheat phenotyping segmentation.

5 Conclusions

Considering the complex characteristics of wheat growth (e.g., extensive tillers, slender leaves, and organ cross-obscuration), this study proposed and validated an algorithm based on multi-view geometric technology. This method effectively processes point cloud data of different wheat plants at various growth stages and automatically extracts phenotypic parameters, enabling the monitoring of dynamic growth changes throughout the entire growth period of wheat. Image sequences were captured using a constructed image acquisition platform and reconstructed into wheat point clouds using the SFM-MVS algorithm. Later, wheat point clouds were extracted using a series of point cloud preprocessing algorithms. Specifically, the improved Euclidean clustering segmentation algorithm increased segmentation efficiency, reduced point cloud processing time, and effectively segmented the main parts of the plants. By comparing the extracted wheat plant height, leaf length, and leaf width with measurement results, an average R^2 of 0.9969 was achieved, indicating that this point cloud processing and phenotypic extraction method is highly accurate and efficient, making it suitable for wheat phenotypic analysis. Although RMSE still can be improved to some extent, some optimization directions are proposed to further improve accuracy and efficiency. Furthermore, through dynamic growth monitoring and ANOVA, the study demonstrated the structural growth changes and differences among different wheat varieties at various growth stages, providing valuable references for wheat plant breeding and variety development. We have recognized that the region growing segmentation algorithm performs well when wheat leaf structures are simple and non-overlapping, but its performance diminishes during periods of extensive tillers, leaf occlusion, and greater curling angles. However, future improvement directions have been proposed to achieve fully automated segmentation and enhance the efficiency of wheat phenotypic analysis.

In summary, this study proposes an effective method for monitoring the dynamic growth and phenotypic characteristics of wheat, offering insights into the relationship between plant structure and function, and optimizing agricultural production. With continued research and technological advancements, this method is expected to play a greater role in plant science and agriculture, contributing to precise and sustainable agricultural development.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. YPZ: Investigation, Project administration, Resources, Software, Validation, Writing – review & editing. SPL: Project administration, Resources, Software, Supervision, Writing – review & editing. YC: Investigation, Resources, Software, Supervision, Validation, Writing – review & editing. YHZ: Formal analysis, Investigation, Validation, Visualization, Writing – review & editing. SJL: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the National Key Research and Development Program of China (2022YFD2001000, 2022YFD2001002).

Acknowledgments

Sincere appreciation is extended to the authors of the R packages utilized in this study for making their code available for data analysis and visualization. Furthermore, the Fujian Academy of Agricultural Sciences is acknowledged for providing the experimental site and data collection equipment. Technical support from China Agricultural University is also gratefully acknowledged.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aharchi, M., and Ait Kbir, M. (2020). "A Review on 3D Reconstruction Techniques from 2D Images," In: Ben Ahmed, M., Boudhir, A., Santos, D., El Aroussi, M., and Karas, I. *Innovations in Smart Cities Applications Edition 3. SCA 2019. Lecture Notes in Intelligent Transportation and Infrastructure*, (Springer, Cham). doi: 10.1007/978-3-030-37629-1_37

Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/j.tplants.2013.09.008

Araus, J. L., Kefauver, S. C., Vergara-Díaz, O., Gracia-Romero, A., Rezzouk, F. Z., Segarra, J., et al. (2022). Crop phenotyping in a context of global change: What to measure and how to do it. *J. Integr. Plant Biol.* 64, 592–618. doi: 10.1111/jipb.13191

Bernardini, F., Mittleman, J., Rushmeier, H., Silva, C., and Taubin, G. (1999). The ball-pivoting algorithm for surface reconstruction. *IEEE Trans. Visualization Comput. Graphics* 5, 349–359. doi: 10.1109/2945.817351

Che, Y., Wang, Q., Xie, Z., Zhou, L., Li, S., Hui, F., et al. (2020). Estimation of maize plant height and leaf area index dynamics using an unmanned aerial vehicle with oblique and nadir photography. *Ann. Bot.* 126, 765–773. doi: 10.1093/aob/mcaa097

Chenxi, Z., Weiliang, W., Xianju, L., Xinyu, G., and Chunjiang, Z. (2022). Phenotypic traits extraction of wheat plants using 3D digitization. *Smart Agric.* 4, 150. doi: 10.12133/j.smartag.SA202203009

Das Choudhury, S., Samal, A., and Awada, T. (2019). Leveraging image analysis for high-throughput plant phenotyping. *Front. Plant Sci.* 10, 508. doi: 10.3389/fpls.2019.00508

Duan, T., Chapman, S. C., Holland, E., Rebetzke, G. J., Guo, Y., and Zheng, B. (2016). Dynamic quantification of canopy structure to characterize early plant vigour in wheat genotypes. *J. Exp. Bot.* 67, 4523–4534. doi: 10.1093/jxb/erw227

Erenstein, O., Jaleta, M., Mottaleb, K. A., Sonder, K., Donovan, J., and Braun, H.-J. (2022). Global Trends in Wheat Production, Consumption and Trade. In; Reynolds, M. P., and Braun, H. J. (eds) *Wheat Improvement* (Springer, Cham), 47–66. doi: 10.1007/978-3-030-90673-3_4

Fiorani, F., and Schurr, U. (2013). Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 64, 267–291. doi: 10.1146/annurev-arplant-050312-120137

Fischler, M. A., and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 381–395. doi: 10.1145/358669.358692

Forero, M. G., Murcia, H. F., Méndez, D., and Betancourt-Lozano, J. (2022). LiDAR platform for acquisition of 3D plant phenotyping database. *Plants* 11, 2199. doi: 10.3390/plants11172199

Fu, J.-b., Shi, J.-w., Zhang, J., and Yang, W.-n. (2021). Extraction of leaf length and plant height from potted wheat based on mask R-CNN. *J. Jilin Agric. Univ.* 43, 163–170.

Furbank, R. T., and Tester, M. (2011). Phenomics–technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644. doi: 10.1016/j.tplants.2011.09.005

Furukawa, Y., Hernández, C., et al. (2015). *Multi-View Stereo: A Tutorial. Foundations and Trends® in Computer Graphics and Vision*, 9, (1-2), 1–148. doi: 10.1561/0600000052

Gibbs, J. A., Pound, M., French, A. P., Wells, D. M., Murchie, E., and Pridmore, T. (2016). Approaches to three-dimensional reconstruction of plant shoot topology and geometry. *Funct. Plant Biol.* 44, 62–75. doi: 10.1071/FP16167

Gu, W., Wen, W., Wu, S., Zheng, C., Lu, X., Chang, W., et al. (2024). 3D reconstruction of wheat plants by integrating point cloud data and virtual design optimization. *Agriculture* 14, 391. doi: 10.3390/agriculture14030391

Hao, H., Wu, S., Li, Y., Wen, W., Fan, J., Zhang, Y., et al. (2024). Automatic acquisition, analysis and wilting measurement of cotton 3D phenotype based on point cloud. *Biosyst. Eng.* 239, 173–189. doi: 10.1016/j.biosystemseng.2024.02.010

He, W., Ye, Z., Li, M., Yan, Y., Lu, W., and Xing, G. (2023). Extraction of soybean plant trait parameters based on SfM-MVS algorithm combined with GRNN. *Front. Plant Sci.* 14, 1181322. doi: 10.3389/fpls.2023.1181322

Kholil, M., Ismanto, I., and Fu'Ad, M. (2021). "3D reconstruction using structure from motion (SFM) algorithm and multi view stereo (MVS) based on computer vision," in *IOP Conference Series: Materials Science and Engineering* (Bristol: IOP Publishing). doi: 10.1088/1757-899X/1073/1/012066

Li, Y., Bu, R., Sun, M., Wu, W., Di, X., and Chen, B. (2018). Pointcnn: Convolution on x-transformed points. *Adv. Neural Inf. Process. Syst.* 31. doi: arXiv:1801.07791

Li, D., Cao, Y., Tang, X.-s., Yan, S., and Cai, X. (2018). Leaf segmentation on dense plant point clouds with facet region growing. *Sensors* 18, 3625. doi: 10.3390/s18113625

Li, Y., Liu, J., Zhang, B., Wang, Y., Yao, J., Zhang, X., et al. (2022). Three-dimensional reconstruction and phenotype measurement of maize seedlings based on multi-view image sequences. *Front. Plant Sci.* 13, 974339. doi: 10.3389/fpls.2022.974339

Liao, L., Cao, L., Xie, Y., Luo, J., and Wang, G. (2022). Phenotypic traits extraction and genetic characteristics assessment of eucalyptus trials based on UAV-borne liDAR and RGB images. *Remote Sens.* 14, 765. doi: 10.3390/rs14030765

Liu, H., Xin, C., Lai, M., He, H., Wang, Y., Wang, M., et al. (2023). RepC-MVSNet: A reparameterized self-supervised 3D reconstruction algorithm for wheat 3D reconstruction. *Agronomy* 13, 1975. doi: 10.3390/agronomy13081975

Liu, Y., Yuan, H., Zhao, X., Fan, C., and Cheng, M. (2023). Fast reconstruction method of three-dimension model based on dual RGB-D cameras for peanut plant. *Plant Methods* 19, 17. doi: 10.1186/s13007-023-00998-z

Ma, X., Wei, B., Guan, H., and Yu, S. (2022). A method of calculating phenotypic traits for soybean canopies based on three-dimensional point cloud. *Ecol. Inf.* 68, 101524. doi: 10.1016/j.ecoinf.2021.101524

Miao, Y., Li, S., Wang, L., Li, H., Qiu, R., and Zhang, M. (2023). A single plant segmentation method of maize point cloud based on Euclidean clustering and K-means clustering. *Comput. Electron. Agric.* 210, 107951. doi: 10.1016/j.compag.2023.107951

Nasirahmadi, A., and Hensel, O. (2022). Toward the next generation of digitalization in agriculture based on digital twin paradigm. *Sensors* 22, 498. doi: 10.3390/s22020498

Paproki, A., Sirault, X., Berry, S., Furbank, R., and Fripp, J. (2012). A novel mesh processing based technique for 3D plant analysis. *BMC Plant Biol.* 12, 63. doi: 10.1186/1471-2229-12-63

Paturkar, A., Gupta, G. S., and Bailey, D. (2020). Non-destructive and cost-effective 3D plant growth monitoring system in outdoor conditions. *Multimed. Tools Appl.* 79, 34955–34971. doi: 10.1007/s11042-020-08854-1

Purcell, W., and Neubauer, T. (2023). Digital twins in agriculture: A state-of-the-art review. *Smart Agric. Technol.* 3, 100094. doi: 10.1016/j.atech.2022.100094

Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017). "Pointnet: Deep Learning on Point Sets for 3D Classification and Segmentation," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA, pp. 77–85. doi: 10.1109/CVPR.2017.16

Rongsheng, Z., Shuai, L., Yongzhe, S., Yangyang, C., Kai, S., Yixin, G., et al. (2021). Research advances and prospects of crop 3D reconstruction technology. *Smart Agric.* 3, 94. doi: 10.12133/j.smartag.2021.3.3.202102-SA002

Sun, Z., Li, Z., and Liu, Y. (2020). "An Improved Lidar Data Segmentation Algorithm Based on Euclidean Clustering," in R. Wang, Z. Chen, W. Zhang and Q. Zhu (eds) *Proceedings of the 11th International Conference on Modelling, Identification and Control (ICMIC2019)* (Lecture Notes in Electrical Engineering: Springer, Singapore), vol 582. doi: 10.1007/978-981-15-0474-7_105

Sun, S., Li, C., Paterson, A. H., Jiang, Y., Xu, R., Robertson, J. S., et al. (2018). In-field high throughput phenotyping and cotton plant growth analysis using LiDAR. *Front. Plant Sci.* 9, 16. doi: 10.3389/fpls.2018.00016

Sun, C., Sun, M., and Chen, H.-T. (2022). "Direct Voxel Grid Optimization: Super-fast Convergence for Radiance Fields Reconstruction," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, pp. 5449–5459. doi: 10.1109/CVPR52688.2022.00538

Teng, X., Zhou, G., Wu, Y., Huang, C., Dong, W., and Xu, S. (2021). Three-dimensional reconstruction method of rapeseed plants in the whole growth period using RGB-D camera. *Sensors* 21, 4628. doi: 10.3390/s21144628

Thomas, H., Qi, C. R., Deschaud, J.-E., Marcotegui, B., Goulette, F., and Guibas, L. (2019). "Kpconv: Flexible and deformable convolution for point clouds," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), pp. 6410–6419 doi: 1109/ICCV.2019.00651

Varela, S., Pederson, T., Bernacchi, C. J., and Leakey, A. D. (2021). Understanding growth dynamics and yield prediction of sorghum using high temporal resolution UAV imagery time series and machine learning. *Remote Sens.* 13, 1763. doi: 10.3390/rs13091763

Vo, A. V., Linh, T. H., Laefer, D. F., and Bertolotto, M. (2015). Octree-based region growing for point cloud segmentation. *Isprs J. Photogramm. Remote Sens.* 104, 88–100. doi: 10.1016/j.isprsjprs.2015.01.011

Wahabzada, M., Paulus, S., Kersting, K., and Mahlein, A. K. (2015). Automated interpretation of 3D laserscanned point clouds for plant organ segmentation. *BMC Bioinf.* 16, 248. doi: 10.1186/s12859-015-0665-2

Wei, K., Liu, S., Chen, Q., Huang, S., Zhong, M., Zhang, J., et al. (2024). Fast Multi-View 3D reconstruction of seedlings based on automatic viewpoint planning. *Comput. Electron. Agric.* 218, 108708. doi: 10.1016/j.compag.2024.108708

Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., and Reynolds, J. M. (2012). 'Structure-from-Motion'photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology* 179, 300–314. doi: 10.1016/j.geomorph.2012.08.021

Wu, S., Wang, J., Zhao, Y., Wen, W., Zhang, Y., Lu, X., et al. (2022a). Characterization and genetic dissection of maize ear leaf midrib acquired by 3D digital technology. *Front. Plant Sci.* 13, 1063056. doi: 10.3389/fpls.2022.1063056

Wu, S., Wen, W., Gou, W., Lu, X., Zhang, W., Zheng, C., et al. (2022b). A miniaturized phenotyping platform for individual plants using multi-view stereo 3D reconstruction. *Front. Plant Sci.* 13, 897746. doi: 10.3389/fpls.2022.897746

Wu, S., Wen, W., Wang, Y., Fan, J., Wang, C., Gou, W., et al. (2020). MVS-pheno: A portable and low-cost phenotyping platform for maize shoots using multiview stereo 3D reconstruction. *Plant Phenom.* 2020, 1848437. doi: 10.34133/2020/1848437

Xiao, S., Chai, H., Shao, K., Shen, M., Wang, Q., Wang, R., et al. (2020). Image-based dynamic quantification of aboveground structure of sugar beet in field. *Remote Sens.* 12, 269. doi: 10.3390/rs12020269

Xiao, S., Chai, H., Wang, Q., Shao, K., Meng, L., Wang, R., et al. (2021). Estimating economic benefit of sugar beet based on three-dimensional computer vision: a case study in Inner Mongolia, China. *Eur. J. Agron.* 130, 126378. doi: 10.1016/j.eja.2021.126378

Xie, C., and Yang, C. (2020). A review on plant high-throughput phenotyping traits using UAV-based sensors. *Comput. Electron. Agric.* 178, 105731. doi: 10.1016/j.compag.2020.105731

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., et al. (2020). Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. *Mol. Plant* 13, 187–214. doi: 10.1016/j.molp.2020.01.008

Yang, D., Yang, H., Liu, D., and Wang, X. (2024). Research on automatic 3D reconstruction of plant phenotype based on Multi-View images. *Comput. Electron. Agric.* 220, 108866. doi: 10.1016/j.compag.2024.108866

Yu, S., Liu, X., Tan, Q., Wang, Z., and Zhang, B. (2024). Sensors, systems and algorithms of 3D reconstruction for smart agriculture and precision farming: A review. *Comput. Electron. Agric.* 224, 109229. doi: 10.1016/j.compag.2024.109229

frontiers | Frontiers in Plant Science

TYPE Data Report
PUBLISHED 12 September 2024
DOI 10.3389/fpls.2024.1414324

Check for updates

OPEN ACCESS

EDITED BY
Ting Sun,
China Jiliang University, China

REVIEWED BY
Alfred Holzwarth,
Max Planck Institute for Chemical Energy Conversion, Germany
Antonio Ferrante,
Sant'Anna School of Advanced Studies, Italy

*CORRESPONDENCE
Yu Tan
tanyu@cau.edu.cn
Fang Ji
jifang@cau.edu.cn

RECEIVED 08 April 2024
ACCEPTED 29 July 2024
PUBLISHED 12 September 2024

CITATION
Jiang Y, Tan Y, Ji F, Su D, Wang S, Zhang L and Zhou Q (2024) CFIHL: a variety of chlorophyll a fluorescence transient image datasets of hydroponic lettuce. *Front. Plant Sci.* 15:1414324. doi: 10.3389/fpls.2024.1414324

COPYRIGHT
© 2024 Jiang, Tan, Ji, Su, Wang, Zhang and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

CFIHL: a variety of chlorophyll a fluorescence transient image datasets of hydroponic lettuce

Yiyu Jiang[1], Yu Tan[1]*, Fang Ji[2]*, Daobilige Su[1], Shuo Wangc[1], Lina Zhang[1] and Qing Zhou[2]

[1]College of Engineering, China Agricultural University, Beijing, China, [2]College of Water Resources and Civil Engineering, China Agricultural University, Beijing, China

KEYWORDS

chlorophyll a fluorescence, hydroponic lettuce, parameterized image, image fusion, semantic segmentation

1 Introduction

Artificial light plant factories ensure year-round stable vegetable production, and vertical multi-cultivation racks can boost vegetable yields (Toyoki et al., 2020). Seedling diagnosis technology is crucial for early identification and removal of weak seedlings, aiming to cut costs in plant factories. This technology relies on seedling biological data to predict growth and effectively manage underperforming seedlings.

Chlorophyll a fluorescence (ChlF) is a long-wavelength signal emitted by plants following light absorption during photosynthesis, directly associated with photosynthetic efficiency (Huipeng, 2023). ChlF is transient and diminishes over time after excitation. Cameras can capture different parameterized images of ChlF depending on excitation wavelength, intensity, and mode (Yao et al., 2018). Parameters like photosynthetic quantum yield (Φ_{PSII}), photochemical quenching (qP), and Photosystem II electron transport rate (ETR(II)) are vital for assessing plant productivity. Higher qP, Φ_{PSII}, and ETR(II) in lettuce indicate greater light utilization efficiency and higher plant yield. Therefore, ChlF imaging is widely used for monitoring photosynthetic performance in plants (Thoren et al., 2010). ChlF imaging technology can assess heterogeneous distribution of photosynthetic activity Lichtenthaler et al. (2013), Xia et al. (2018). This includes early detection of herbicide application, nutrient deficiency, drought stress, photorespiratory mutants, improved photosynthesis, freezing tolerance, insect herbivory, leaf fungal infection, and ozone damage (Lawson and Vialet-Chabrand, 2018). Therefore, it is necessary to collect ChlF images to realize the study of plant growth stress and plant yield estimation.

At present, there are few plant chlorophyll fluorescence datasets, with most focusing on stress identification in Arabidopsis crops. To enable the chlorophyll fluorescence system to track photosynthesis in a single leaf, Ruiz et al. (2024) proposed a chlorophyll fluorescence dataset for Arabidopsis plants at different developmental stages and used an improved Mask R-CNN network to segment Arabidopsis leaves, accurately tracking changes in leaf photosynthesis. The Computer Vision Problems in Plant Phenotyping (CVPPP) dataset

consists of 783 RGB images of wild Arabidopsis and tobacco plants, used for testing algorithms such as plant segmentation, leaf segmentation, and leaf counting (Minervini et al., 2016). Additionally, Rousseau et al. (2013) collected 49 fluorescence images of Arabidopsis inoculated with sterile water, naming it the 'Real-Fluo-Healthy' dataset. However, this dataset only contains images of healthy Arabidopsis and cannot provide a basis for disease diagnosis. Therefore, Rousseau et al. further used Arabidopsis plants inoculated with the DC3000 bacterial strain as the research object, performing chlorophyll fluorescence imaging on each culture dish every two days to obtain TIFF images of Fo, Fm, and Fv/Fm for judging the status of diseased Arabidopsis. In addition, Pavicic et al. (2021) proposed an annotated dataset of diseased Arabidopsis fluorescence images and a threshold segmentation algorithm to segment the normal and diseased leaf areas in the images for disease diagnosis. Currently, there is a lack of similar chlorophyll fluorescence datasets for hydroponic lettuce research, which mainly uses RGB image data to evaluate lettuce growth status. Islam et al. (2024) used an improved Mask R-CNN method to detect and segment lettuce seedlings from the background of the seedling growth tray and estimate seedling growth. Abidi et al. (2024) collected an RGB image dataset of nutrient deficiencies in hydroponic lettuce and proposed a deep learning image processing method to classify these deficiencies. However, RGB images can only capture superficial color and texture information, making it difficult to obtain internal physiological parameters of lettuce leaves and accurately judge lettuce growth status. Therefore, this paper collects a ChlF dataset of hydroponic plants and tests region of interest extraction algorithms for fluorescence images, providing data for accurately estimating lettuce seedling status using ChlF technology.

Hydroponically grown lettuce often uses sponges for anchorage, making it more challenging to extract the fluorescent area compared to plants grown in substrates, due to the common presence of microbes or green algae within the cultivation sponges (Weizhong, 2023). During the capture of ChlF systems, the fluorescence from microbes or green algae present in the cultivation sponges is also captured, resulting in minimal differences between them in grayscale images taken under blue light, affecting the extraction of areas of interest in the ChlF images of hydroponic lettuce seedlings. In microbial fluorescence images, such as that of Chlamydomonas reinhardtii, exposure to excitation light produces a lower level of Non-Photochemical Quenching (NPQ). In the chloroplasts of algae like Chlorella or Chlamydomonas, effective carbohydrate metabolism decomposition leads to a much reduced redox balance compared to plant chloroplasts (Alric, 2010). Differences exist in the fluorescence parameters under ChlF induction kinetics between lettuce and green algae (Papageorgiou et al., 2007). However, the threshold segmentation algorithm cannot determine the category of the area. In practice, the grayscale of the sponge block of green algae in the initial fluorescence image is very similar to that of the seedling canopy, leading to mis-segmentation of areas with similar grayscales. Given the excellent feature extraction and generalization performance of deep learning algorithms in image recognition, a deep learning semantic segmentation algorithm is considered for extracting the seedling canopy area from a single fluorescence image. Testing has shown that fusing multiple fluorescence images can improve the extraction accuracy of the seedling canopy area, effectively reduce background noise interference, and more accurately extract the fluorescence values. Image fusion, by integrating information from multiple sources with complementary datasets to enhance a single image (Qu et al., 2024), makes the generated images more detailed and reliable through the combination of key information from multiple source images (Karim et al., 2023). At this stage, it has been proposed to introduce accuracy of advanced visual tasks such as detection, recognition, and segmentation into the design of the loss function during the fusion phase (Liu et al., 2023; Ma et al., 2023), guiding the fusion process from a decision-making level to achieve better segmentation results (James and Dasarathy, 2014).

This article introduces the Hydroponic Lettuce Seedling ChlF Dataset. The dataset comprises various transient images of ChlF from different cultivation states, with annotations marking the positions of the entire lettuce canopy projection. The aim is to accelerate the development of algorithms for extracting ChlF areas in hydroponically grown lettuce seedlings. The data for this article is publicly released at https://github.com/Yiyu-Jiang/CFIHL. For user convenience, the original.igr data files (which can be exported as chlorophyll images, tables, documents) are provided, along with the corresponding annotation JSON files.

The primary contributions of this dataset are summarized as follows:

(1) The CF-imager has provided a ChlF image set of the whole hydroponic lettuce seedlings, with the dataset being approximately 1.3 G in size. The dataset includes more than 300 sets of images, each set containing ten transient ChlF images. The dataset offers a variety of physiological images of plants and annotations for individual seedlings.

(2) Compared to the existing lettuce seedling datasets, this paper has conducted tests with multiple image algorithms, comparing threshold segmentation and deep learning algorithms, as well as the segmentation effects of images before and after fusion. The fused image can extract the ChlF value of lettuce more accurately, and can further accurately obtain parameters such as photochemical quantum yield, thereby reflecting the photosynthetic utilization efficiency and lettuce yield.

2 Materials and methods

2.1 Data acquisition

Lettuce was cultivated in the artificial light plant factory of China Agricultural University in 2022 and 2023. During the cultivation of hydroponic lettuce, we used an LED lamp (WR-16W) with an R ratio of 1.2, a light intensity of 200 $\mu molm^{-2}s^{-1}$, and a photoperiod of 16 hd^{-1}. The temperature during the light period was 22 ± 1°C, with a relative humidity of 70 ± 5% and a CO_2 concentration of 800 ± 50$\mu molmol^{-1}$. During the dark period, the temperature was 18 ± 1°C, the relative humidity was 65 ± 10%, and

the CO_2 concentration was not controlled. The seedling raising period for hydroponic lettuce was 20 days. The nutrient solution used was the Yamazaki lettuce formula Yan (2019), with an EC of 1.0-1.2 $mScm^{-1}$ and a pH adjusted to 6.0-6.5 mol/L.

In this paper, the seedlings grown for about 21 days were placed in the CF-Imager(Technologica, UK) instrument, then photographed, as shown in Figure 1A. The CF-Imager system uses a high-performance AVT Stingray SXGA+ low-noise line scan CCD camera with a resolution of 1392 × 1040. When collecting ChlF images, the camera needs to respond quickly, which sacrifices some resolution. Therefore, the image resolution was set to 696 × 519. The imaging frequency is 50 FPS, and the readout noise is less than 12e. This ChlF imaging system used 470 nm blue light as the excitation light. Based on the principles of slow fluorescence kinetics, the detection light was set to 5 $\mu molm^{-2}s^{-1}$, the actinic light to 200 $\mu molm^{-2}s^{-1}$, and the saturating pulse light to 6000 $\mu molm^{-2}s^{-1}$ with a duration of 800 ms, establishing the excitation process (Daquan, 2002).

The measurement principle is illustrated in Figure 1B. After 30 minutes in the dark, a fully dark-adapted leaf was exposed to measuring light. Within 1-2 minutes, as the fluorescence level stabilized, the minimum initial fluorescence Fo image was obtained. Then, a saturating pulse light was applied and turned off after one pulse to obtain the maximum fluorescence image Fm. Therefore, the fluorescence parameter image Fv/Fm (Fv=Fm-Fo) can be calculated, representing the potential photochemical efficiency of PSII.

Secondly, the actinic light of 200 $\mu molm^{-2}s^{-1}$ was turned on to induce leaf photosynthesis. After a few minutes, the leaf's photosynthetic rate reached a steady state, and the steady-state fluorescence(Fs) image was obtained. At this point, a pulse of saturating light was applied and then turned off to obtain the maximum fluorescence image under light Fm'. The actual quantum efficiency of PSII(Φ_{PSII}) and the non-photochemical quenching coefficient (NPQ), defined as (Fm -Fm')/Fm', were then calculated.

Using the method described above to stimulate fluorescence in vegetables, various ChlF transient images were obtained, as shown in Figure 1C, all with a resolution of 696 × 519. The ChlF images are essentially grayscale images with a single channel. For intuitive description, this paper maps the grayscale values of some ChlF images to the RGB color space (Oxborough, 2004), rendering them as pseudocolor images.

The main ChlF parameters and descriptions are shown in Table 1:

2.2 Data structure and annotation

The CF-Imager can capture 10 types of images: the initial grayscale image (Igi), Fo, Fm, Fv/Fm, Fm', Fs, NPQ, Fv'/Fm',Φ_{PSII} and qP. Detailed information about the dataset is shown in Figure 2A. To help users differentiate the fluorescence images of seedlings in various states, the raw fluorescence data are stored in separate folders in chronological order, totaling approximately 270 sets of image data. Additionally, to enrich the dataset and investigate the changes in ChlF parameters under different growth conditions of lettuce. This dataset collected images under three distinct environments: nutrient deficiency (only ultrapure water, no nutrients), extremely low light intensity (only 10 - 20 $\mu molm^{-2}s^{-1}$ ambient light), and abnormal EC and pH (EC is about 900 mS/cm, pH is about 7.5 mol/L).

Image segmentation aims to accurately extract the ChlF parameters of the seedling canopy. Image annotation is a prerequisite for training deep learning-based image segmentation models. It provides the training model with prior knowledge of the seedling canopy area within the image. This prior knowledge helps establish the main parameters of the model, enabling the recognition of similar images. This study uses LabelImg (https://github.com/tzutalin/labelimg) to annotate the seedling canopy area in the fluorescence images of the dataset. The seedling canopy area is labeled as 'lettuce', generating a label mask image and a JSON file for further neural network model training, as shown in Figure 2B.

2.3 Image fusion algorithm and semantic segmentation algorithm test

2.3.1 Algorithm introduction

This paper tests extraction algorithms for various lettuce canopy projection areas using the constructed ChlF transient image dataset. It eliminates fluorescence interference caused by

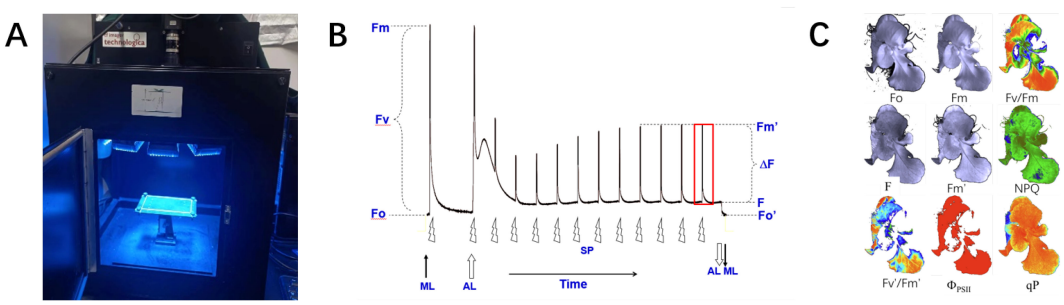


FIGURE 1
Details of data acquisition. **(A)** ChlF imaging equipment. **(B)** Schematic diagram of ChlF measurement principle. **(C)** ChlF result images.

TABLE 1 Commonly abbreviated and described ChlF parameters.

Parameter	Formula	Description
PSII	\	Photosystem II: it is a photosynthetic unit in the thylakoid membrane that contains two light-harvesting complexes and a light reaction center.
Fo	\	Minimum initial fluorescence
Fm	\	Maximum fluorescence
Fm'	\	Maximum fluorescence under light
Fs	\	Steady state fluorescence
NPQ	Fm/Fm'-1	Non-photochemical quenching: estimates the rate constant for heat loss from PSII.
Fv/Fm	(Fm-Fo)/Fm	Maximum quantum efficiency of PSII photochemistry.
Fv'/Fm'	(Fm'-Fo')/Fm'	Maximum efficiency of PSII photochemistry in the light, if all centers were open.
Φ_{PSII}	(Fm'-F')/Fm'	PSII operating efficiency: the quantum efficiency of PSII electron transport in the light.
qP	(Fm'-Fs)/Fm'-Fo'	Photochemical quenching: relates PSII maximum efficiency to operating efficiency. Non-linearly related to proportion of PSII centers that are open. See qL.

This list is only used to identify the most common parameters, as detailed in the review (Murchie and Lawson, 2013).

microorganisms or green algae in the background sponge and explores effective methods to accurately obtain ChlF parameters of the seedling canopy. This provides a basis for further evaluation of lettuce growth status.

First, a comparison between threshold segmentation and deep learning image segmentation algorithms was conducted on a single ChlF image. Early research on image segmentation mainly focused on threshold segmentation methods, where one or several threshold values divide the grayscale histogram into several classes. It was believed that the grayscale value of the seedling canopy was consistent and distinct from the background area. The existing fluorescence systems mainly use threshold segmentation (Lawson and Vialet-Chabrand, 2018). However, threshold segmentation algorithms cannot accurately categorize different areas. In practice, the grayscale of green algae on the sponge is very similar to that of the seedling canopy in the initial fluorescence grayscale image, leading to mis-segmentation. Therefore, this paper introduces a more adaptable deep learning segmentation algorithm. Deep learning has been widely used in image recognition due to its excellent feature extraction and generalization capabilities, achieving good segmentation results (Li et al., 2021). Among them, the DeepLab series is one of the most popular image segmentation models (Minaee et al., 2022). Therefore, this paper selects the DeepLabv3+ network (Chen et al., 2018), which has the best segmentation performance in the DeepLab series, as the deep learning semantic segmentation test algorithm.

Then, the traditional fusion algorithm and the SeAFusion algorithm based on advanced visual tasks were tested on multiple images, and the deep learning segmentation algorithm was used to segment and compare the images before and after fusion. In addition to single image segmentation, this paper also tested multiple image fusion methods. Image fusion technology can be divided into several types, namely remote sensing image fusion, multi-exposure image fusion, and visible light and infrared image fusion. Among these, visible light and infrared image fusion is one of the most commonly used types. The main purpose of visible light and infrared image fusion is to extract the detail information from visible light and the contrast information from infrared image,

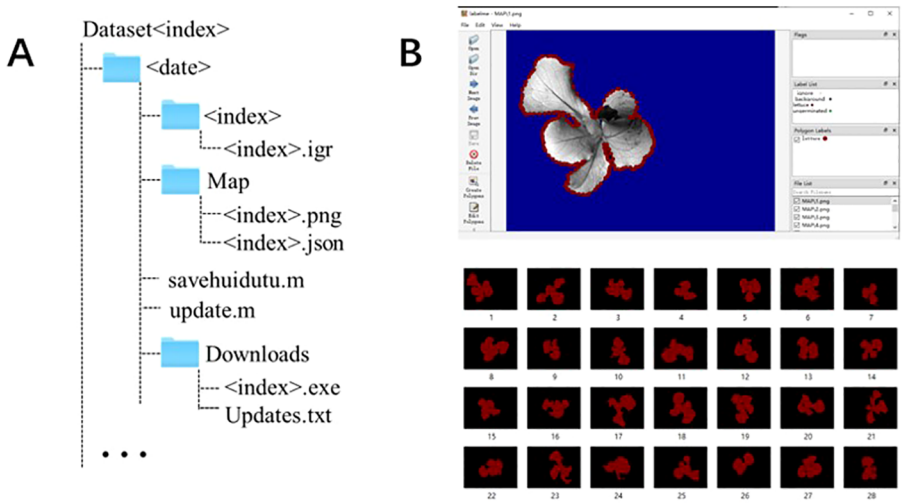


FIGURE 2
Dataset format and image annotation. **(A)** Folder structure for dataset. The term <date> refers to the time of the acquisition of a dataset, while the term <index> identifies each piece of data. **(B)** Example of image annotation.

which is similar to the goal of fluorescence image fusion. Therefore, this paper selected four recently released image fusion algorithms CNN, IFEVIP, LatLRR, MST-SR (Zhang et al., 2020).

However, these fusion algorithms primarily focus on the visual quality and statistical indicators of the fused image, while ignoring the needs of advanced visual tasks (i.e. classification, localization, detection and segmentation). Therefore, this paper tested the image fusion algorithm named SeAFusion (Tang et al., 2022) combined with segmentation tasks. The main concept of the SeAFusion algorithm is to generate a fused image by passing the source image through a fusion network. Then inputting the fused image into a segmentation network to obtain the segmentation result, where the segmentation network used in the SeAFusion algorithm is DeepLabv3+.

2.3.2 Image segmentation evaluation indicators

This study uses Precision, Recall and intersection over union (IoU) as evaluation indicators for training target detection models. The range of these three evaluation metrics is [0, 1], the detailed descriptions are provided in Table 2. In Table 2, 'TP' represents true positives, indicating the number of samples correctly classified as positive. 'FP' represents false positives, indicating the number of samples incorrectly classified as positive. 'FN' represents false negatives, indicating the number of samples incorrectly classified as negative.

TABLE 2 Image segmentation algorithm evaluation indicators.

Parameter	Formula	Describe
Precision	$\frac{TP}{TP+FP} \times 100\,\%$	Precision indicates the percentage of correctly identified pixels among the total number of identified pixels.
Recall	$\frac{TP}{TP+FN} \times 100\,\%$	Recall indicates the percentage of correctly identified pixels that meet the requirements to the total number of pixels in the test set.
IoU	$\frac{TP-FP}{TP+FN} \times 100\,\%$	IoU (Intersection over Union) is the ratio of the number of elements in the intersection of two sets to the number of elements in their union.

2.3.3 Implementation details

All models in this study were trained on a single GPU. The hardware configuration for training and testing the network models includes an i9-12900K CPU, NVIDIA GeForce RTX 3070 GPU, and 32GB DDR4 RAM. The dataset was randomly divided into training, test, and validation sets in an 8:1:1 ratio. The input size of each fluorescence image is 696 × 519 pixels. The DeepLabv3+ algorithm (Yu et al., 2024) sets the number of iterations (Epochs) to 300 times, with an initial learning rate of 0.001 and a batch size of 4. SGD is used as the optimizer, with a weight decay of 0.0001 and a momentum of 0.9. The training weights are verified and saved every 5 iterations.

3 Result

In this paper, the automatic threshold segmentation method (Zhao et al., 2023) and the deep learning DeepLabv3+ network are used for the initial grayscale image (Igi), photochemical quantum yield (Φ_{PSII}) and maximum fluorescence (Fm) images. The segmentation results are shown in Figure 3A. Visualizing the segmentation results allows us to intuitively assess the accuracy of each algorithm in segmenting the edges of lettuce seedlings. The green areas in the figure indicate non-canopy areas mistakenly identified as canopy areas, while the magenta areas indicate canopy areas not recognized by the segmentation algorithm.

The image segmentation results indicator results are shown in Table 3. Among them, the threshold segmentation algorithm has the best segmentation effect for the Fm parameterized image with an IoU of 84.89%, but there is still a 15% error with the actual canopy. The deep learning DeepLabv3+ network has higher segmentation accuracy in most images than the threshold

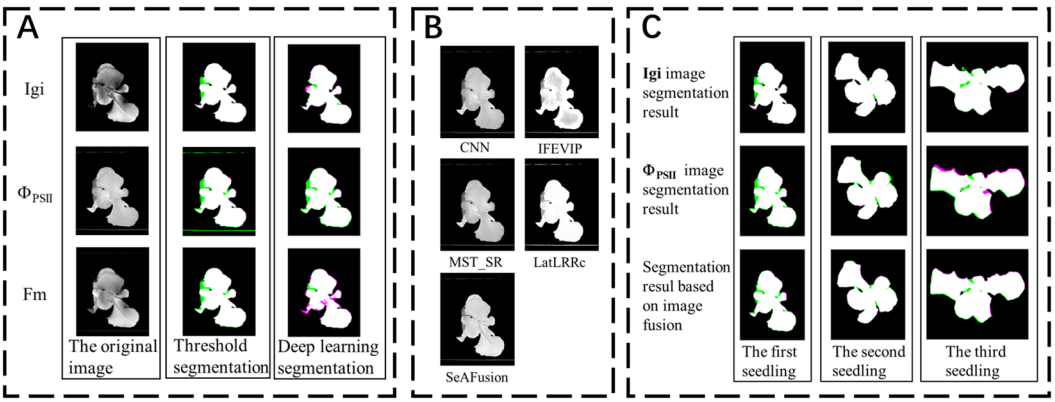


FIGURE 3
(A) Comparison of segmentation results of different fluorescence images using different segmentation algorithms. **(B)** Results of fusion algorithms using 5 other infrared and visible light fusion algorithms. **(C)** Deep learning image segmentation results before and after fusion.

TABLE 3 Comparison of image segmentation results.

Algorithm		Precision/%	Recall/%	IoU/%
Threshold segmentation	Igi	88.74	99.36	83.02
	Φ_{PSII}	86.00	88.95	77.69
	Fm	86.82	97.45	84.89
Deeplabv3 + algorithm	Igi	89.57	97.13	87.27
	Φ_{PSII}	83.80	93.36	80.93
	Fm	85.81	96.94	83.55

segmentation method, and the best segmentation effect is for the Igi image, with an IoU of 87.27%, an increase of about 2%. Automatic threshold segmentation can roughly distinguish the seedling canopy from the background. However, the presence of green algae can cause the background threshold to be similar to that of the seedling, leading to incorrect segmentation. In contrast, deep learning methods automatically extract features and are wellsuited for segmenting complex scenes with similar pixels. Consequently, deep learning has shown improved segmentation effectiveness compared to simple grayscale threshold segmentation.

This study uses the DeepLabv3+ algorithm to segment the seedling canopy area of fluorescence images. The Precision, Recall, and IoU are slightly improved compared to threshold segmentation, but the IoU is still below 90%. Therefore, it is necessary to consider other segmentation methods to improve the accuracy of seedling canopy segmentation. Considering the complementarity between fluorescence images, this study first fuses the ChlF transient image set. Then inputs the fused image into the segmentation network for canopy segmentation.

The fusion results of Igi and Φ_{PSII} for the same seedling using four popular fusion algorithms (Zhang et al., 2020) and the SeAFusion algorithm are shown in Figure 3B. The segmentation visualization results of the DeepLabv3+ network before and after the fusion of three seedlings are shown in Figure 3C. This paper uses indicators such as entropy (EN), standard deviation (SD), spatial frequency (SF) and sum of difference correlations (SCD) to characterize the effect of image fusion (Zhang et al., 2019). The higher EN and SD indicate that the method can better preserve the source image. These images contain a large amount of information and have high pixel contrast (Liu et al., 2023).The comparison results of each fusion algorithm are shown in Table 4. Among them, the fusion image texture information obtained by the CNN, IFEVIP, LatLRR, and MST_SR algorithms has a greater loss. The fused image obtained by the SeAFusion algorithm not only ensures the contrast between the seedling canopy and the sponge block, but also fully retains the seedling texture. Combined with the visualization results in Figure 3B, the SeAFusion algorithm in Table 4 has higher AG, SD, and EN parameter values and clearer edges compared to other fusion algorithms from a visual perspective.

Finally, to verify that the fused image segmentation effect is better, this paper uses the Deeplabv3+ algorithm to compare the segmentation results of the fused images of the two fusion methods (Igi is fused with Φ_{PSII} and Fm images respectively), and the results are shown in Table 5. Compared with the segmentation results in Table 3, the fusion Precision, Recall and IoU ratio parameters are increased by 0.25%, 0.35% and 0.56% respectively, which is helpful for the accurate extraction of fluorescent areas of interest such as the lettuce canopy. Among them, the fluorescent image segmentation results of three seedlings are shown in Figure 3C. There are many wrongly segmented areas (green and magenta areas) in the seedling canopy segmentation using the Deeplabv3+ algorithm without image fusion, while the wrongly segmented areas (green and magenta areas) are significantly reduced after fusing Igi and Φ_{PSII}. This is because the SeAFusion algorithm transmits the semantic information required for high-level visual tasks (segmentation) back to the fusion network, thereby enabling the fusion network to effectively retain the semantic information in the source image.

TABLE 4 Fusion algorithm comparison.

Indicators / Algorithm	AG	SF	SD	SCD	Q^{abf}	EN
CNN	1.63	16.37	6.16	NaN[1]	0.98	1.35
IFEVIP	1.76	21.22	6.17	1.70	0.67	1.23
LatLRR	1.93	25.52	6.17	1.72	0.40	0.93
MST-SR	1.63	16.37	6.16	NaN[1]	0.98	1.35
SeAFusion	3.11	0.054	9.23	1.62	0.49	3.76

[1]NaN (Not a Number) is a value in computer science representing an undefined or unrepresentable value.

TABLE 5 Comparison of image segmentation results.

Indicators / Image collection	Precision/%	Recall/%	IoU/%
Igi and Φ_{PSII} images are fusioned	95.35	96.17	91.86
Igi and Fm images are fusioned	95.60	96.52	92.42

4 Conclusion

This paper presents a ChlF dataset of hydroponic lettuce seedlings, featuring diverse transient images under various cultural conditions. Additionally, the study compares the effectiveness of the threshold segmentation algorithm with the deep learning-based Deeplabv3+ algorithm for extracting the seedling canopy. Owing to its superior feature extraction capability, the Deeplabv3+ algorithm outperforms the threshold segmentation method in various fluorescence parameterized images. However, its IoU score remains below 90%. Consequently, various image fusion algorithms were tested, and the fused images were semantically segmented using the Deeplabv3+ algorithm. The results indicate that the fused images provide better segmentation of the lettuce canopy. The dataset introduced in this paper supports the evaluation of lettuce growth status based on ChlF data and offers methods for accurately extracting fluorescence image parameters of lettuce.

5 Discussion

Inspired by the Arabidopsis dataset, this paper proposes a new hydroponic lettuce seedling ChlF dataset to address the lack of similar datasets. Additionally, the paper deeply analyzes the issue of extracting plant canopy areas from ChlF images in the dataset. The ChlF system captures both lettuce and green algae fluorescence simultaneously, with the initial fluorescence image showing similar values for the seedling canopy and the background. Therefore, this paper uses the deep learning Deeplabv3+ algorithm to leverage its powerful feature extraction capabilities to improve the segmentation of the seedling canopy.

However, due to limited information in a single image, the paper also considers the fluorescence value differences between the seedling canopy and the background, adopting canopy segmentation after image fusion. Further testing of multiple image fusion algorithms revealed that the fused images provided better segmentation of the lettuce canopy. Thus, using complementary information from different fluorescence images improves the segmentation of the seedling canopy's edge contours.

There are some limitations in this study. For example, only some image fusion methods were tested during image fusion, and all fusion combinations were not enumerated due to the limited length of the article. Only two fluorescence parameter images were considered as input, and multiple image inputs were not considered. The fusion algorithm tested in this study is designed for merging infrared and visible light. It is recommended to design a fusion method suitable for the fusion of multiple fluorescence images in the future to help improve the acquisition of ChlF values. In addition, for the current dataset, future research should further expand the sample range and explore applications in different objects and scenarios such as plant mesophyll and vein segmentation, and segmentation of different types of cells in fluorescence images, in order to prove that the fusion of different fluorescence images has a wide range of applications and is worthy of in-depth research.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/Yiyu-Jiang/CFIHL.

Author contributions

YJ: Writing – original draft, Writing – review & editing. YT: Writing – original draft, Writing – review & editing, Data curation, Formal analysis. SW: Methodology, Writing – review & editing. DS: Data curation, Formal analysis, Writing – original draft. LZ: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft. FJ: Data curation, Funding acquisition, Writing – review & editing. QZ: Project administration, Funding acquisition, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was financially supported by the Key Research and Development Project of Shandong Province (Grant No. 2022CXGC020708).

Acknowledgments

We would like to thank the editor and reviewers for their valuable input, time, and suggestions to improve the quality of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1414324/full#supplementary-material

References

Abidi, M. H., Chintakindi, S., Rehman, A. U., and Mohammed, M. K. (2024). Elucidation of intelligent classification framework for hydroponic lettuce deficiency using enhanced optimization strategy and ensemble multi-dilated adaptive networks. *IEEE Access* 12, 58406–58426. doi: 10.1109/ACCESS.2024.3392482

Alric, J. (2010). Cyclic electron flow around photosystem i in unicellular green algae. *Photosynthesis Res.* 106, 47–56. doi: 10.1007/s11120-010-9566-4

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 834–848. doi: 10.1109/TPAMI.2017.2699184

Daquan, X. (2002). *Photosynthetic efficiency* (Shanghai: Shanghai Scientific and Technical Publishers).

Huipeng, M. (2023). *Multi-scale Monitoring of Drought Stress in Forests Based on Sif And Multi-source Remote Sensing Indices* (Nanjing, China: Nanjing Forestry University).

Islam, S., Reza, M. N., Chowdhury, M., Ahmed, S., Lee, K.-H., Ali, M., et al. (2024). Detection and segmentation of lettuce seedlings from seedling-growing tray imagery using an improved mask r-cnn method. *Smart Agric. Technol.* 8, 100455. doi: 10.1016/j.atech.2024.100455

James, A. P., and Dasarathy, B. V. (2014). Medical image fusion: A survey of the state of the art. *Inf. Fusion* 19, 4–19. doi: 10.1016/j.inffus.2013.12.002

Karim, S., Tong, G., Li, J., Qadir, A., Farooq, U., and Yu, Y. (2023). Current advances and future perspectives of image fusion: A comprehensive review. *Inf. Fusion* 90, 185–217. doi: 10.1016/j.inffus.2022.09.019

Lawson, T., and Vialet-Chabrand, S. (2018). Chlorophyll fluorescence imaging. *Photosynthesis: Methods Protoc.* 1770, 121–140. doi: 10.1007/978-1-4939-7786-4_8

Li, L., Zhang, S., and Wang, B. (2021). Plant disease detection and classification by deep learning-a review. *IEEE Access* 9, 56683–56698. doi: 10.1109/ACCESS.2021.3069646

Lichtenthaler, H. K., Langsdorf, G., and Buschmann, C. (2013). Uptake of diuron and concomitant loss of photosynthetic activity in leaves as visualized by imaging the red chlorophyll fluorescence. *Photosynthesis Res.* 116, 355–361. doi: 10.1007/s11120-013-9842-1

Liu, J., Liu, Z., Wu, G., Ma, L., Liu, R., Zhong, W., et al. (2023). "Multi-interactive feature learning and a full-time multi-modality benchmar for image fusion and segmentation," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. (Paris, France: IEEE), 8081–8090. doi: 10.1109/ICCV51070.2023.00745

Ma, J., Tang, L., Zhang, H., and Xu, H. (2023). Deep learningbased image fusion: a survey. *J. Image Graphics* 28, 3–36. doi: 10.11834/jig.220422

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., and Terzopoulos, D. (2022). Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 3523–3542. doi: 10.1109/TPAMI.2021.3059968

Minervini, M., Fischbach, A., Scharr, H., and Tsaftaris, S. A. (2016). Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognition Lett.* 81, 80–89. doi: 10.1016/j.patrec.2015.10.013

Murchie, E. H., and Lawson, T. (2013). Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications. *J. Exp. Bot.* 64, 3983–3998. doi: 10.1093/jxb/ert208

Oxborough, K. (2004). Imaging of chlorophyll a fluorescence: theoretical and practical aspects of an emerging technique for the monitoring of photosynthetic performance. *J. Exp. Bot.* 55, 1195–1205. doi: 10.1093/jxb/erh145

Papageorgiou, G. C., Tsimilli-Michael, M., and Stamatakis, K. (2007). The fast and slow kinetics of chlorophyll a fluorescence induction in plants, algae and cyanobacteria: a viewpoint. *Photosynthesis Res.* 94, 275–290. doi: 10.1007/s11120-007-9193-x

Pavicic, M., Overmyer, K., Rehman, A. U., Jones, P., Jacobson, D., and Himanen, K. (2021). Image-based methods to score fungal pathogen symptom progression and severity in excised arabidopsis leaves. *Plants* 10, 158. doi: 10.3390/plants10010158

Qu, L., Liu, S., Wang, M., Li, S., Yin, S., and Song, Z. (2024). Trans2fuse: Empowering image fusion through self-supervised learning and multi-modal transformations via transformer networks. *Expert Syst. Appl.* 236, 121363. doi: 10.1016/j.eswa.2023.121363

Rousseau, C., Belin, E., Bove, E., Rousseau, D., Fabre, F., Berruyer, R., et al. (2013). High throughput quantitative phenotyping of plant resistance using chlorophyll fluorescence image analysis. *Plant Methods* 9, 17. doi: 10.1186/1746-4811-9-17

Ruiz, F. J., Nguyen, T. P., Peller, J., Aranzana, M. J., Polder, G., and Aarts, M. G. M. (2024). Letra: a leaf tracking workflow based on convolutional neural networks and intersection over union. *Plant Methods* 20, 11. doi: 10.1186/s13007-024-01138-x

Tang, L., Yuan, J., and Ma, J. (2022). Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network. *Inf. Fusion* 82, 28–42. doi: 10.1016/j.inffus.2021.12.004

Thoren, D., Thoren, P., and Schmidhalter, U. (2010). Influence of ambient light and temperature on laser-induced chlorophyll fluorescence measurements. *Eur. J. Agron.* 32, 169–176. doi: 10.1016/j.eja.2009.10.003

Toyoki, K., Niu, G., and Takagaki, M. (2020). *Plant Factory, -an Indoor Vertical Farming System for Efficient Quality Foor Production, 2nd* (Chiba, Japan: Elsevier).

Weizhong, Y. (2023). Development of Biosensor for Microbial Quantitative Monitoring in Plant Factory Based on Microfluidics. (Beijing, China: China Agricultural University).

Xia, Q., Ji, X., Jiang, Y., Guo, Y., and Tan, J. (2018). Modelling and simulation of chlorophyll fluorescence from photosystem ii as affected by temperature. *IET Syst. Biol.* 12, 304–310. doi: 10.1049/iet-syb.2018.5030

Yan, Z. (2019). *Effect of LED Lighting Environment on Quality Control and Light Energy Use Efficiency of Hydroponic Lettuce* (Beijing, China: China Agricultural University).

Yao, J., Sun, D., Cen, H., Xu, H., Weng, H., Yuan, F., et al. (2018). Phenotyping of arabidopsis drought stress response using kinetic chlorophyll fluorescence and multicolor fluorescence imaging. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00603

Yu, G., Wang, Y., Gan, S., Xu, H., Chen, Y., and Lei, W. (2024). Extracting the navigation lines of crop-free ridges using improved deeplabv3+. *Trans. Chin. Soc. Agric. Eng. (Transactions CSAE)* 40, 168–175. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). doi: 10.11975/j.issn.1002-6819.202401211

Zhang, X., Ye, P., and Xiao, G. (2020). "Vifb: A visible and infrared image fusion benchmark (IEEE; CVF; IEEE Comp Soc)," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. (Seattle, WA, USA: IEEE), 468–478. doi: 10.1109/CVPRW50498.2020.00060

Zhang, X., Zhang, J., Kou, J., and Xu, X. (2019). "Research status of image fusion: A review," in *2019 4th International Conference on Control, Robotics and Cybernetics (CRC)*. (Tokyo, Japan: IEEE), 179–183. doi: 10.1109/CRC.2019.00044

Zhao, R., Tang, W., An, L., Qiao, L., Wang, N., Sun, H., et al. (2023). Solar-induced chlorophyll fluorescence extraction based on heterogeneous light distribution for improving *in-situ* chlorophyll content estimation. *Comput. Electron. Agric.* 215, 108405. doi: 10.1016/j.compag.2023.108405

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 10 October 2024
DOI 10.3389/fpls.2024.1435042

Check for updates

OPEN ACCESS

EDITED BY
Lei Shu,
Nanjing Agricultural University, China

REVIEWED BY
Yongliang Qiao,
University of Adelaide, Australia
Jun Ni,
Nanjing Agricultural University, China
Jaya Krishna Mandivarapu,
Microsoft, United States

*CORRESPONDENCE
Baohua Yang
ybh@ahau.edu.cn

RECEIVED 19 May 2024
ACCEPTED 24 September 2024
PUBLISHED 10 October 2024

CITATION
Yang B, Chen R, Gao Z and Zhi H (2024) FIDMT-GhostNet: a lightweight density estimation model for wheat ear counting. *Front. Plant Sci.* 15:1435042. doi: 10.3389/fpls.2024.1435042

COPYRIGHT
© 2024 Yang, Chen, Gao and Zhi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

FIDMT-GhostNet: a lightweight density estimation model for wheat ear counting

Baohua Yang*, Runchao Chen, Zhiwei Gao and Hongbo Zhi

School of Information and Artificial Intelligence, Anhui Agricultural University, Hefei, China

Wheat is one of the important food crops in the world, and the stability and growth of wheat production have a decisive impact on global food security and economic prosperity. Wheat counting is of great significance for agricultural management, yield prediction and resource allocation. Research shows that the wheat ear counting method based on deep learning has achieved remarkable results and the model accuracy is high. However, the complex background of wheat fields, dense wheat ears, small wheat ear targets, and different sizes of wheat ears make the accurate positioning and counting of wheat ears still face great challenges. To this end, an automatic positioning and counting method of wheat ears based on FIDMT-GhostNet (focal inverse distance transform maps - GhostNet) is proposed. Firstly, a lightweight wheat ear counting network using GhostNet as the feature extraction network is proposed, aiming to obtain multi-scale wheat ear features. Secondly, in view of the difficulty in counting caused by dense wheat ears, the point annotation-based network FIDMT (focal inverse distance transform maps) is introduced as a baseline network to improve counting accuracy. Furthermore, to address the problem of less feature information caused by the small ear of wheat target, a dense upsampling convolution module is introduced to improve the resolution of the image and extract more detailed information. Finally, to overcome background noise or wheat ear interference, a local maximum value detection strategy is designed to realize automatic processing of wheat ear counting. To verify the effectiveness of the FIDMT-GhostNet model, the constructed wheat image data sets including WEC, WEDD and GWHD were used for training and testing. Experimental results show that the accuracy of the wheat ear counting model reaches 0.9145, and the model parameters reach 8.42M, indicating that the model FIDMT-GhostNet proposed in this study has good performance.

KEYWORDS

FIDMT, GhostNet, counting, convolutional neural network, wheat

1 Introduction

Wheat is a widely planted crop in the world, and its yield and quality are directly related to global food security and the development of agricultural economy (Atamanyuk et al.,

2023). Wheat ears, as the reproductive organs of wheat, are the key to reflecting the growth status, yield potential and quality characteristics of wheat. The precise positioning and accurate counting of wheat ears is an important link in agricultural production, which not only involves real-time monitoring of wheat growth status, but also is a core indicator for predicting wheat yield. Through high-precision positioning and counting technology, agricultural scientists and researchers can more accurately grasp the wheat growth, reproductive ability, and yield potential of wheat, thereby providing scientific basis for agricultural management, breeding optimization, and yield prediction. In addition, the positioning and counting of wheat ears are also of great significance to wheat breeding work (Hasan et al., 2018). By comparing the number and distribution of wheat ears under different varieties or different treatment conditions, breeders can screen out more advantageous germplasm resources and develop wheat varieties that are more suitable for the local growth environment and market demand (Simão et al., 2023).

Traditional wheat counting methods are highly dependent on the experience and visual judgment, resulting in inaccurate and unrepeatable counting results. With the widespread application and rapid development of artificial intelligence, deep learning has made significant progress and breakthroughs in the field of wheat ear counting research. Compared with traditional machine learning-based methods, its performance has been significantly improved. Deep learning models can handle more complex data patterns and optimize their performance through training on large amounts of data (Zou et al., 2023). In terms of wheat ear counting, deep learning is mainly based on three methods, target detection, image segmentation and density map estimation.

Methods based on target detection mainly use deep learning models to identify and locate wheat ears in images. Many scholars have trained detection models to identify the characteristics of wheat ears and mark the position of each wheat ear in the image. Madec et al. (2019) used the Faster R-CNN model to detect and count wheat ears on high-resolution wheat images. In fact, the advantage of this method is that it can accurately identify and locate each wheat ear. Li and Wu (2022) proposed an improved YOLOv5 algorithm based on shallow features. By adding four times downsampling to the feature pyramid to capture the characteristics of micro-wheat ears, an attention mechanism was added to the network to realize the detection and counting of wheat ears. Yang et al. (2021) used YOLOv4, which added a channel attention mechanism and a spatial attention mechanism, to suppress irrelevant background information, increase the expression ability of wheat ear features, and detect and count wheat ears in different data sets. Wang et al. (2021b) proposed an improved EfficientDet-D0 target detection model for wheat ear counting. Research shows regardless of the single-stage or two-stage detection network, it can achieve good results in wheat ear detection and counting. However, when wheat planting density is high and leaves are severely blocked, the target detection algorithm may cause false detections or missed detections, resulting in inaccurate counting results. Therefore, it is necessary to build a more robust model.

The counting method of wheat ears based on image segmentation is to segment the wheat ears in the image from the background or other objects. This method usually uses a deep learning model to learn pixel-level features in the image to accurately segment wheat ears and background. Through the segmented image, the number of wheat ears can be calculated. For example, Zhang et al. (2022a) used the semantic segmentation model Wheat-Net to segment and count wheat ears under complex backgrounds in the field. Wang et al. (2021a) proposed a semantic segmentation regression network SSRNet to count wheat ears in remote sensing images. Ma et al. (2020) proposed a model based on semantic segmentation, EarSegNet, which effectively improved the segmentation accuracy and efficiency of winter wheat ears. Xu et al. (2023) proposed the CBAM-HRNet model for wheat grain segmentation and counting. Misra et al. (2020) proposed a segmentation model SpikeSegNet to implement counting of wheat ears. Dandrifosse et al. (2022) developed the DeepMAC segmentation model to achieve wheat ear segmentation and density counting. Although the above segmentation models can improve the accuracy of wheat ear counting to a certain extent, due to the influence of environmental factors such as lighting, occlusion, and shadow, some wheat ears may be mistakenly segmented or missed, thus affecting the accuracy of counting. However, due to the diversity of wheat planting environments and growing conditions, it is difficult to ensure the generalization ability of the segmentation algorithm in different scenarios.

The method based on density estimation uses a deep learning model to estimate the density of wheat ears in the image. It does not require precise identification and positioning of each wheat ear, but estimates the overall distribution of wheat ears in the image. In addition, density map-based methods have greater flexibility and generalization capabilities and can be flexibly applied to counting tasks of different scales and scenarios. For example, Bao et al. (2020) used the CSRNet model based on density estimation and suitable for target counting in dense scenes to count wheat ears. Sun et al. (2021) optimized the CSRNet model to count wheat ears on the global wheat data set. Wu et al. (2023) proposed the DM-Net model to estimate and count the density of wheat ears. Ma et al. (2022) proposed a transfer learning model EarDensityNet based on a fully convolutional neural network to count wheat ears. Lu et al. (2022) proposed a TasselNetV3 model based on density estimation to implement wheat counting. Since the density map can reflect the spatial distribution of objects, it can better handle the occlusion problem between objects. Research shows that density map-based methods have significant advantages in the field of wheat counting and can overcome the limitations of traditional methods and improve counting accuracy and efficiency. Although the above-mentioned counting model based on density estimation has achieved certain results, it still has shortcomings in terms of training time, parameter redundancy, processing high density, and adapting to complex scenes. Therefore, wheat counting based on density plot regression still faces great challenges.

To this end, a wheat ear density estimation model based on lightweight convolutional neural network is proposed in this study, which can better alleviate the above-mentioned problems. The purpose of this study is: (1) to solve the problem of different size

of wheat ears, GhostNet is proposed as the feature extraction network, which can generate multi-scale feature maps through its unique Ghost module; (2) to improve the counting accuracy caused by dense wheat ears, FIDMT based on point annotation is introduced as the baseline network; (3) to obtain more characteristic information of small target wheat ears, a dense upsampling convolution module is proposed; and (4) to overcome the background noise or interference of wheat ears, a local maximum detection strategy is adopted to realize automated processing of wheat ear counting.

2 Materials and methods

2.1 Data collection

2.1.1 Data sources

Wheat images were collected the National Modern Agriculture Demonstration Zone (31°29′26″N, 117°13′46″E) located in Guohe Town, China. The shots were taken on May 7, 2021 and May 17, 2021 using the rear main camera (48 million pixels) of Huawei's nova5pro mobile phone. When collecting each image, the camera lens was kept at a distance of about 30cm directly above the wheat ears. A total of 500 top-view images of wheat with a resolution of 3024 × 3024 pixels were collected and record them as wheat ear counting (WEC) data set, which contains many samples with different light intensities, different densities, and different periods. Among them, there are 313 images of wheat in the filling stage and 187 in the mature stage.

To ensure the diversity and representativeness of wheat ears in the data set and enhance the universality and robustness of the model, this article added 500 images from the global wheat detection data set (GWHD, http://www.global-wheat.com) with a resolution of 1024 × 1024 pixels. This data set contains wheat data from multiple countries and different growth stages. In addition, a total of 500 clear wheat images from 10 countries in the GWHD dataset were selected, including Australia, Canada, China, France, Japan, Mexico, Sudan, Switzerland, the United Kingdom, and the United States. Wheat data from the wheat ears detection dataset (WEDD) was also added, and a Sony ILCE-6000 digital camera was used to maintain a distance of 2.9 meters from the ground for overhead photography. To reduce the number of parameters and increase the computing speed in the subsequent model training process, the wheat images with an original size of 6000 × 4000 pixels in the data set were cropped, and 200 wheat images with a resolution of 3072 × 3072 pixels were obtained.

Figure 1 shows some examples of wheat data sets, including three data sets, WEC, WEDD and GWHD. Among them, Figure 1A represents WEC data set, Figure 1B represents WEDD data set, and Figure 1C represents GWHD data set. There are certain differences in the shape, density, occlusion, and period of the wheat ears, which can intuitively reflect the diversity of wheat ears in the wheat data set constructed in this study.

2.1.2 Data annotation

After the wheat data is selected, MATLAB is used to label the wheat ear points on the images in the above data set, as shown in Figure 2A. And Figure 2B shows the.mat format tag that saves the wheat ear position information. This annotation technology not only maintains the clarity of the image, but also more accurately captures the spatial position information of the wheat ears in the image. To further reduce the computational burden and annotation workload, all wheat images in the data set are uniformly adjusted to a high resolution of 1024 × 1024 pixels before annotation, which not only ensures the accuracy of annotation, but also improves processing efficiency.

2.1.3 Dataset construction

The reasonable division of the wheat data set is a critical step for the subsequent model training and evaluation process. The density estimation model will be trained and tested based on the wheat data set divided in Table 1. The ratio of the training set, verification set, and test set is set to 7:2:1, and the detailed numbers are 840, 240, and 120 images respectively.

2.2 GhostNet

The GhostNet is a lightweight network framework based on the Ghost module, which is mainly composed of simple linear operations and standard convolutions (Han et al., 2019). Specifically, the Ghost module generates some intrinsic feature maps through ordinary convolutions, and then generates more feature maps from these feature maps through a series of simple linear operations. These new feature maps are combined with the original feature maps to form the final feature map output, as shown in Figure 3. This structure enables GhostNet to greatly reduce the size and number of parameters of the model while ensuring model performance, making the model more lightweight. The GhostBottleneck module is an important part of the GhostNet model, which is mainly composed of two Ghost modules. Each Ghost module is responsible for generating feature maps and adding features and channels through linear operations. This structure enables the GhostBottleneck module to further reduce the computational complexity of the model while maintaining feature extraction capabilities.

2.3 Upsampling method

2.3.1 Interpolation algorithm

Bilinear Interpolation Upsampling is a widely used technique in image processing to enlarge the size of an image to a higher resolution (Soh et al., 2024). The core idea is to perform linear interpolation in two directions (usually horizontal and vertical) to estimate and fill the values of the newly added pixels in the enlarged image. Compared with other more complex upsampling methods,

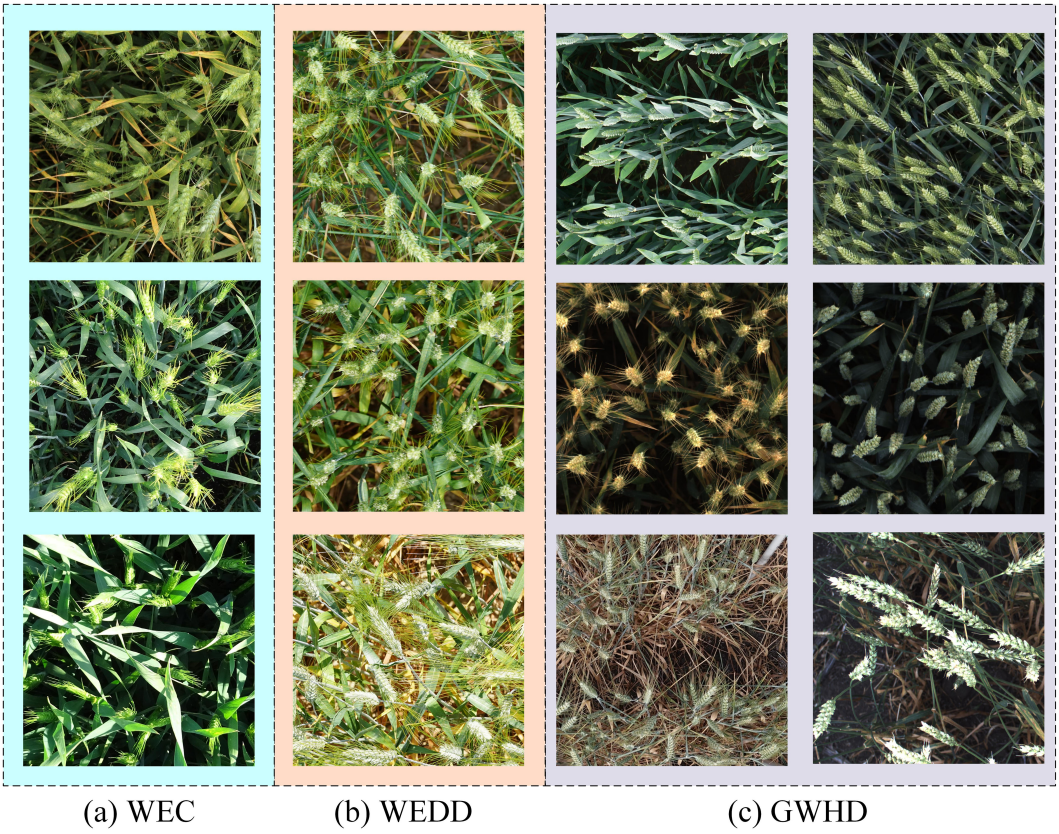


FIGURE 1
Examples of wheat image: **(A)** WEC, **(B)** WEDD, and **(C)** GWHD.

the calculation of bilinear interpolation is relatively simple and easy to implement.

2.3.2 Transposed convolution

Transposed Convolution (Faisal et al., 2023), as a commonly used upsampling method in deep learning, increases the size of feature maps by simulating the inverse process of convolution layer. Transposed convolution can effectively help the model recover



annPoint
49x2 double

	1	2	3
4	299.5000	71.5000	
5	129.5000	297.5000	
6	121.5000	469.5000	
7	261.5000	337.5000	
8	403.5000	319.5000	
9	347.5000	461.5000	
10	335.5000	609.5000	
11	331.5000	721.5000	
12	483.5000	687.5000	
13	493.5000	799.5000	
14	79.5000	703.5000	
15	9.5000	705.5000	
16	21.5000	793.5000	
17	115.5000	879.5000	
18	13.5000	975.5000	

(a) Point labeling example
(b) Label data

FIGURE 2
Example of wheat ear labeling: **(A)** Point labeling example, **(B)** Label data.

TABLE 1 Dataset details for this study.

Data set	Number of images	Training set	Validation set	Test set	Total number of wheat ears
WEC	500	350	100	50	12721
GWHD	500	350	100	50	22380
WEDD	200	140	40	20	11308
	1200	840	240	120	46409

high-resolution images or feature maps from low-resolution features. However, the computational complexity of transposed convolution is high, and the demand for computing resources and storage space is also large, which needs special consideration when designing large or complex deep learning models.

2.3.3 Dense upsampling convolution

Dense Upsampling Convolution (DUC) is an upsampling technique that compensates for the loss in length and width by dividing the entire feature map and operating on the channel dimension. It is mainly used to increase the resolution of an image or feature map to restore it to its original size or a larger size, while trying to preserve and restore the details in the image (Zhou et al., 2022).

2.4 Density estimation model

2.4.1 MCNN

MCNN (Multi-column Convolutional Neural Network) is a three-column convolutional neural network model composed of three different convolution kernel sizes (Zhang et al., 2016). It completes the density estimation and counting tasks of the target by mapping the input image into a predicted density map. Wang et al. (2019) first applied it to wheat ear counting, using the feature map generated by the multi-column convolutional neural network for feature fusion, reducing the dimensionality of the fused features, and then outputting the predicted wheat ear density map and counting the wheat ears. The detailed network structure of MCNN is shown in Figure 4.

2.4.2 CSRNet

CSRNet (Congested Scene Recogrition Network) is a target counting model suitable for dense scenes (Li et al., 2018). Its network structure mainly consists of a front-end network and a back-end network, which are VGG-16 with all fully connected layers removed, and 6 consecutive dilated convolution operations with the dilation rate set to 2. Finally, after a 1×1 ordinary convolution operation, the predicted density map is output and the targets are counted. The CSRNet model structure is shown in Figure 5.

2.4.3 FIDTM

FIDTM is a convolutional neural network model for density estimation and counting of dense targets (Liang et al., 2022). It is designed to solve the problem of accurately counting and locating targets in extremely dense scenes. When dealing with wheat ear counting tasks, the FIDTM model shows its unique advantages, as shown in Figure 6. To locate and count wheat ears more accurately, the original FIDTM model will be improved. On the one hand, the Ghost module is introduced. We replace the original convolutional

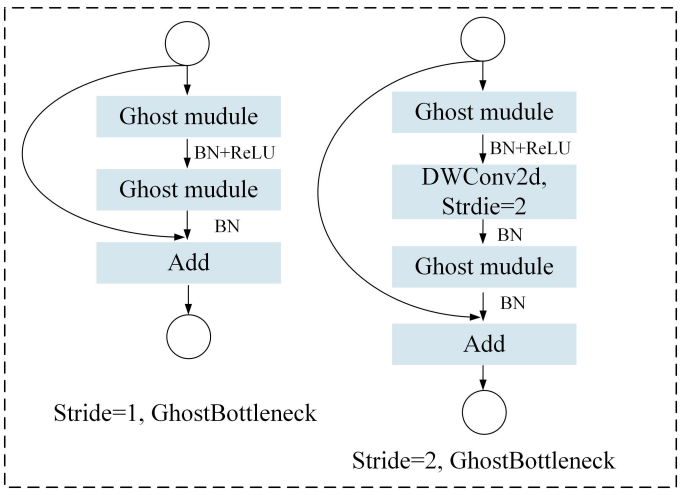


FIGURE 3
GhostNet model.

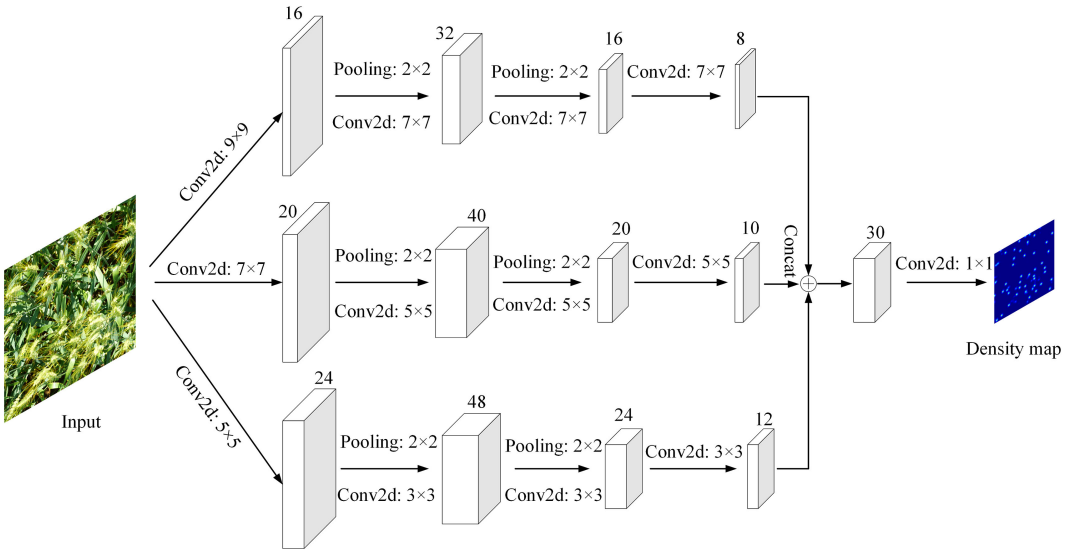


FIGURE 4
Network structure of MCNN model.

layer of the FIDTM model with the Ghost module to reduce the computational complexity and parameter amount of the model while maintaining or improving the performance of the model. On the other hand, in view of the fact that the FIDTM model may lose the detailed information of wheat ears during the upsampling process, the Dense Upsampling Convolution (DUC) module is used to obtain more wheat ear features. Dense upsampling mainly divides the entire feature map into multiple parts with the same size as the input feature map to make up for the loss in length and width by increasing the channel dimension.

2.5 Positioning based on LMDS

The local maximum detection strategy (LMDS) is an important component in the wheat ear counting task, which is used to effectively extract the center point of each wheat ear. In wheat ear counting, the local maximum detection strategy works by identifying possible wheat ear regions through image segmentation or object detection algorithms. Then, within these wheat ear regions, the maximum values of local pixel intensity or feature values are found, which usually correspond to the center

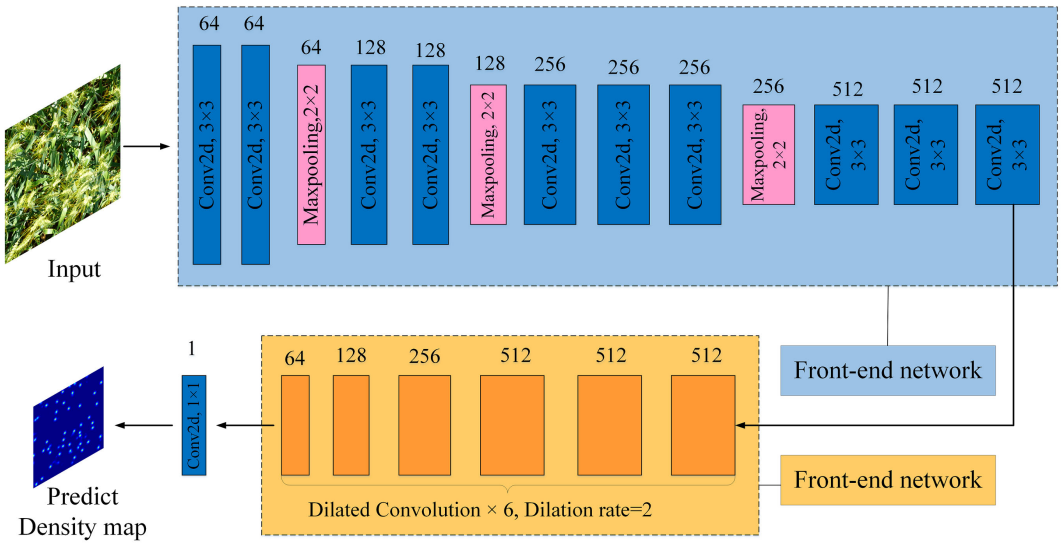


FIGURE 5
Network structure of CSRNet model.

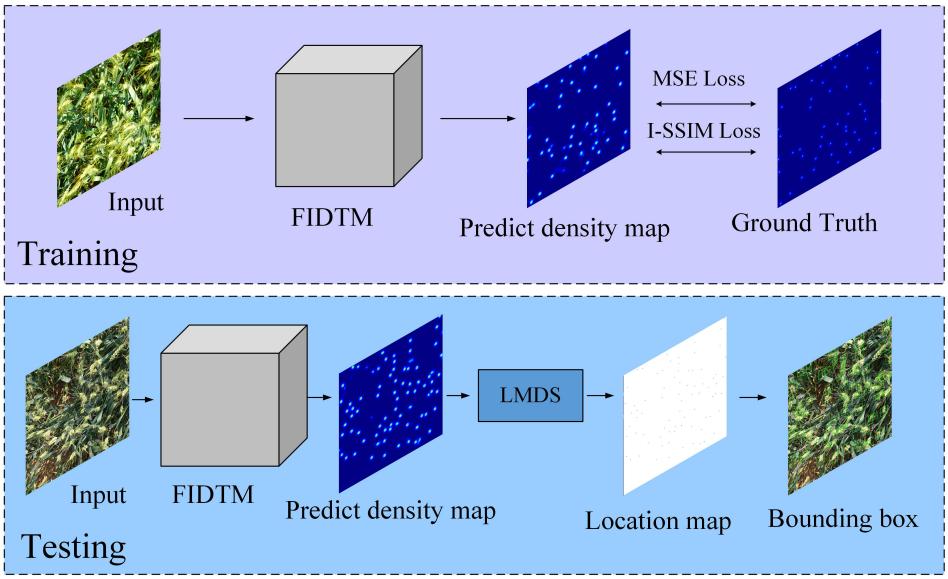


FIGURE 6
Wheat ear counting process based on FIDTM model.

points of the wheat ears. Finally, these center points are extracted and the number of wheat ears in the image was accurately counted.

In this study, the point coordinates and K nearest neighbor distance of each wheat ear can be obtained by using the LMDS algorithm, and the approximate size of the wheat ear detection frame is defined as shown in Equation 1.

$$D_{(x,y)\in P} = \min \begin{cases} \overline{d} = f \times \frac{1}{k} \sum_{1}^{k} d_{(x,y)}^{k} \\ \min\left(img_w, img_h\right) \times 0.05 \end{cases} \quad (1)$$

Where $D_{(x,y)}$ represents the size of the wheat ear detection box located at (x, y), P represents the set of predicted wheat ear positions, $\overline{d}$ represents the average distance between the wheat ear point set $P_{(x,y)}$ and uses a scalar factor f to limit the size. In images with sparse wheat ears, $\overline{d}$ is generally much larger than the actual wheat ear size, so a threshold needs to be selected related to the image size to limit the wheat ear size.

2.6 Density map based on FIDT mapping

To accurately locate wheat ears in images in relatively dense areas, the new label of Focal Inverse Distance Transform (FIDT) of the target positioning task is used, that is, the position of each wheat ear is represented by the nearest neighbor distance information. The generation principle of the FIDT graph is based on the conversion graph of Euclidean distance (Sun et al., 2019). The conversion graph of Euclidean distance is defined as shown in Equation 2:

$$P(x, y) = \min \sqrt{(x - x')^2 + (y - y')^2} \quad (2)$$

In the formula, $\left(x', y'\right) \subset$ B and B represents the set containing all wheat ear annotation information. For any pixel (x, y) in the image, $P(x, y)$ represents the shortest distance between each pixel in the image and its nearest wheat ear center point.

At the same time, to make it easier to distinguish the foreground target from the background, the focus inverse distance transformation map is used in the FIDTM model, which is defined as:

$$I = \frac{1}{P(x, y)^{(\alpha \times P(x,y) + \beta)} + C} \quad (3)$$

where I represents the FIDT diagram, α and β are set to 0.02 and 0.75 respectively in the FIDTM model.

2.7 Count of wheat ear density map based on FIDMT-GhostNet

2.7.1 Overall technical route

The flow chart of wheat ear counting based on the FIDMT-GhostNet model is shown in Figure 7. Firstly, the wheat image is used as the input of the model, and then the feature map is obtained after 1 standard convolution and 16 consecutive GhostBottleneck modules, which is upsampled through the DUC algorithm to make the resolution of the feature map consistent with the input wheat ear image. Secondly, the wheat ear density map was predicted based on a standard convolution with a convolution kernel size of 1 × 1 and an output channel. Finally, the counting and positioning of wheat ears are realized based on the density map and LMDS algorithm. The wheat ears were located and detection frames were generated by predicting the local maximum of the density map. Among them, the target positioning information output by the FIDMT-GhostNet model is based on the pixel level, that is, only the position coordinates of the center point of the wheat ear are determined. The final result of the detection requires obtaining the

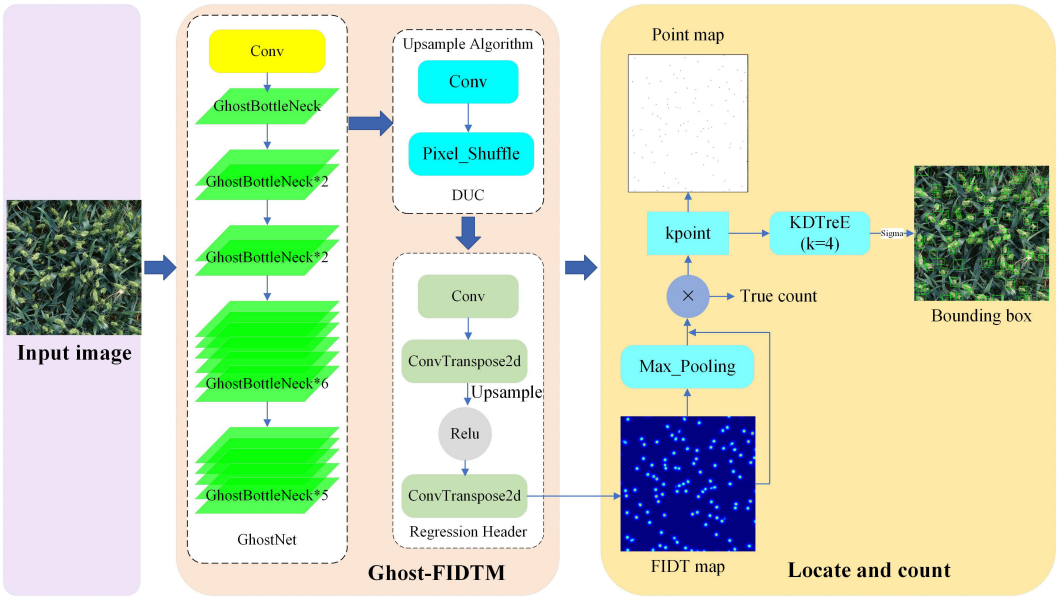


FIGURE 7
The flow chart of wheat ear counting based on the FIDMT-GhostNet.

wheat ear area in the image, that is, locating the target frame of each wheat ear. Therefore, the target frame of each wheat ear can be calculated using LMDS.

The FIDMT-GhostNet model innovatively introduces a continuous stacking structure of 16 GhostBottleneck modules when building its backbone feature extraction network. This strategy greatly enhances the model's ability to perceive details and texture information of wheat ear images in a multi-level feature space, thereby significantly improving the accuracy of wheat ear counting. It is worth mentioning that despite the large number of modules, the efficiency of the GhostBottleneck module ensures that the model consumes reasonable computing resources, allowing FIDMT-GhostNet to maintain high performance while also achieving high computing efficiency. This design not only reflects the ingenuity of the model design, but also demonstrates the wisdom of seeking a balance between calculation accuracy and efficiency in the wheat ear counting task.

2.7.2 Design of loss function

The loss function is used to calculate the error between the model output and the true result. The Euclidean distance loss L_e and the density consistency loss L_c are used as the loss L of the overall network, that is, the total loss L used by the network is the weighted superposition of L_e and L_c. The calculation formula is shown in Equation 4 and 5.

$$L = L_e + L_c \tag{4}$$

$$L_e = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left|x_i^{GT} - x_i^{P}\right|^2} \tag{5}$$

In the formula, N is the number of wheat ear samples, x_i^{GT} and x_i^{P} are respectively ground truth and the estimated wheat ear count corresponding to the ith wheat ear sample.

SSIM (Structural Similarity Index Measure) is a metric used to measure the structural similarity between two images, including the brightness, contrast, and structural information of the image. SSIM values range from -1 to 1, where 1 means the two images are identical, 0 means there is no similarity, and -1 means the two images are completely different. To comprehensively evaluate the performance of the FIDMT-GhostNet model in target counting and localization tasks, SSIM is used to evaluate the similarity between the predicted wheat ear density map and the ground-truth density map to evaluate the performance of the model.

$$SSIM(E,G) = \frac{(2\mu_E\mu_G + \chi_1)(2\sigma_{EG} + \chi_2)}{(\mu_E^2 + \mu_G^2 + \chi_1)(\sigma_E^2 + \sigma_G^2 + \chi_2)} \tag{6}$$

$$L_c = \frac{1}{N}\sum_{n=1}^{N}(1 - SSIM(DM(I_n), \omega) \tag{7}$$

Among them, N is the total number of training samples, I_n is the nth image input to the model, $DM(I_n)$ is the predicted FIDT image obtained by inputting the nth image into the model, and ω is the FIDT image obtained by the focal inverse distance transform of the nth image. μ_E and σ_E^2 are the mean and variance of the predicted density map respectively, μ_G and σ_G^2 represent the mean and variance of the Ground Truth respectively. σ_{EG} represents the covariance between the predicted density map and Ground Truth, χ_1 =0.0001, χ_2=0.0009. The value range of SSIM is between -1 and 1. The larger the value, the more similar the two images are.

10.3389/fpls.2024.1435042

TABLE 2 Details of experimental hardware and software parameters.

Configuration name	Parameters
Operating System	Windows 10 Professional 64-bit
Code running environment	Python3.7
Deep learning framework	Pytorch1.80
GPU model	NVIDIA GeForce RTX 2080
Processor	Intel Core i7-8700 CPU@3.20GHz

2.7.3 Density estimation evaluation index

In density estimation, root mean squared error (RMSE) is often used to measure the difference in each pixel value between the predicted density map and the true density map. The mean absolute error (MAE) calculates the average of the absolute value of the difference between the predicted value and the true value. Compared with MSE, MAE is more robust to outliers. To evaluate the wheat ear counting performance of FIDMT-GhostNet, MAE, RMSE, and R^2 are used as evaluation indicators of the model (Qiao et al., 2023; Zhang et al., 2022b).

$$MAE = \frac{1}{N}\sum_{i=1}^{n}\left|x_i^{GT} - x_i^{P}\right| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left|x_i^{GT} - x_i^{P}\right|^2} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(x_i^{GT} - x_i^{P})}{\sum_{i=1}^{N}(x_i^{GT} - \bar{x}_i^{P})} \tag{10}$$

In the formula, N is the number of tested wheat ear samples, x_i^{GT} and x_i^{P} are respectively the Ground Truth and estimated wheat ear count corresponding to the ith wheat ear sample, MAE can directly reflect the accuracy of the wheat ear counting model, and $RMSE$ can better reflect Out of the robustness of the wheat ear counting model, R^2 reflects the degree of fit between the predicted value of the wheat ear and the Ground Truth. The closer the value is to 1, the higher the degree of fitting and the higher the reliability of the trend line.

In addition, parameters (Parameter), Floating Point Operations (FLOPs), Model Size, and FPS (Frames Per Second) indicators are also introduced to evaluate the wheat ear counting efficiency of the model.

TABLE 3 Training parameters of FIDMT-GhostNet model.

Parameter name	Value
Epoch	1000
Batch size	2
Learning rate	0.0001
Weight decay	0.0005
Optimizer	Adam

3 Results

3.1 Experimental parameter settings

In this study, the PyTorch deep learning framework is used to build the network model. The system information and other software and hardware related information are listed in Table 2.

The parameter settings during model training are shown in Table 3. When the Epoch value is greater than or equal to 100, verification is performed every 5 Epoch intervals.

3.2 Wheat ear counting results

To verify the effectiveness and feasibility of counting method proposed for the wheat ear in this study, and to evaluate the universality and transferability of the FIDMT-GhostNet model, we used WEC, GWHD, and WEDD data sets to carry out wheat ear counting. The counting results are shown in Table 4. From the detailed data in Table 4, we can observe the counting performance of the model under different experimental data. An in-depth analysis of these performance metrics reveals several noteworthy trends or characteristics. Among the three data sets, the FIDMT-GhostNet model performs best with the WEC data set. In particular, the WEC data set test has the lowest MAE and RMSE, and the highest R^2, reaching 3.25, 3.56, and 0.9279 respectively. Compared with the GWHD and WEDD data sets, the MAE of WEC decreased by 2.23 and 1.62 respectively, the RMSE decreased by 2.67 and 2.02 respectively, and the R^2 increased by 0.1153 and 0.0067 respectively. The generalization performance of the FIDMT-GhostNet model on the GWHD and WEDD test sets is lower than that on the WEC test set. It may be that the GWHD data set comes from seven different countries, and there are certain differences in the size, shape, and color of wheat ears. The wheat ears in the WEDD data set are more densely distributed and have a higher degree of occlusion. However, the wheat ears in the WEC data set have a low degree of occlusion, and the differences between wheat ears are small. Therefore, the optimized FIDMT-GhostNet model has demonstrated excellent counting capabilities in multiple data sets. Especially when processing wheat ear images with severe occlusion and complex backgrounds, the model can still maintain a high counting accuracy. This further verifies the effectiveness and robustness of the model design.

To further prove the wheat counting performance of the FIDMT-GhostNet model, Figure 8 shows the density plot of the output of some samples of the data set. These sample data come from different data sets and wheat ears from different countries. It can be seen from the wheat ear counting results in Figure 8 that this model has certain universality and generalization in the wheat ear counting task.

In the wheat image analysis shown in Figure 8, we can observe that from top to bottom are the original wheat image, the FIDT label density map, the density map predicted by the model, the wheat ear positioning map predicted by the model, and the map

TABLE 4 Wheat ear count results from different datasets.

Dataset	Number of images	MAE	RMSE	R^2
WEC	50	3.25	3.56	0.9279
GWHD	50	5.48	6.23	0.8126
WEDD	20	4.87	5.58	0.9212

generated based on point annotation. Although there are significant differences in the distribution, occlusion degree, shape and growth cycle of the wheat ears in the test image, the density map predicted based on the FIDMT-GhostNet model and the FIDT label density map show a high degree of consistency in distribution, which fully demonstrates that the FIDMT-GhostNet model has the ability to handle differences between wheat ears. Furthermore, for the wheat ears in the original image, the wheat ear counting results of the FIDMT-GhostNet model is very close to the ground truth, which shows not only the accuracy of the model in the counting task, but also its reliability and effectiveness in practical applications. In particular, the wheat ear detection frame predicted based on the FIDMT-GhostNet model can more accurately cover the complete wheat ear, which further verifies the performance of the model in positioning and scale estimation. Therefore, the test results of different data sets show that the proposed model can not only achieve the task of counting wheat ears, but also obtain accurate position and scale information of wheat ears.

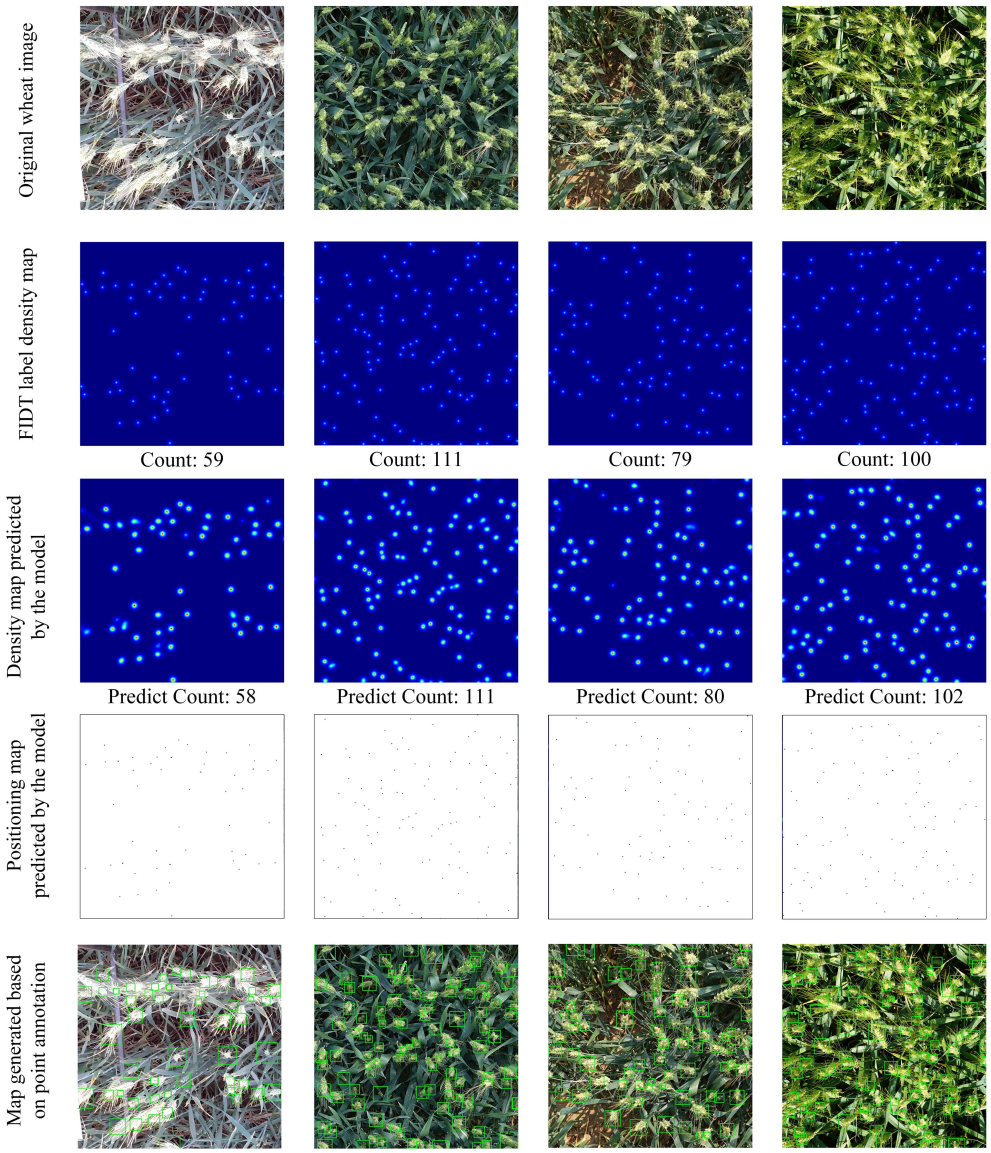


FIGURE 8
Results of wheat ear counts from FIDMT-GhostNet model.

10.3389/fpls.2024.1435042

4 Discussion and analysis

4.1 Counting results of wheat ears with different degrees of occlusion

Wheat ears in the field often show varying degrees of occlusion due to overlapping in the middle and late stages of growth, which increases the difficulty of automatic counting. Occlusion will not only cause some wheat ears to be difficult to identify in the image, but may also affect the accuracy and stability of the counting algorithm. To evaluate the wheat ear counting effect of the proposed FIDMT-GhostNet model under different degrees of occlusion, wheat images with mild occlusion, moderate occlusion and severe occlusion were specially selected for testing in the test set, aiming to intuitively demonstrate the model's ability to count different Adaptability to occlusion situations and counting performance. Figure 9 shows the test results of the model. From left to right, they are the wheat ear image, density map of FIDT label generated using point annotation, the predicted density map, and positioning map of wheat ear.

As can be seen from Figure 9, the prediction results of wheat ear counting are very close to ground truth, and the prediction of wheat ear detection frame can accurately locate most wheat ears. The FIDMT-GhostNet model shows excellent counting capabilities in the wheat ear counting task, mainly due to the SE-block module embedded in its structure. By introducing an adaptive feature recalibration mechanism, SE-block significantly enhances the model's ability to express features of important channels while suppressing the influence of useless channels. In wheat ear counting tasks, occlusion between wheat ears is a common and thorny problem. However, SE-block can enhance the model's focus on important channels, allowing the model to more accurately identify occluded wheat ear features. At the same time, by suppressing the influence of useless channels, the FIDMT-GhostNet model can reduce interference caused by background noise or other irrelevant features, further improving the accuracy of wheat ear counting. As shown from the first row in Figure 9, the wheat ears have higher definition in the image, and there is almost no occlusion. The output of the density map is basically the same as the ground truth density map. As shown in the second and third rows of Figure 9, the FIDMT-GhostNet model can handle the scene where wheat leaves block wheat ears to a certain extent. Even in the case of overlapping wheat ears, the wheat ear density map estimated using the FIDMT-GhostNet model has a smaller error than the ground truth density map. The test results show that the wheat attention mechanism in the model plays an indispensable role. It enables the model to

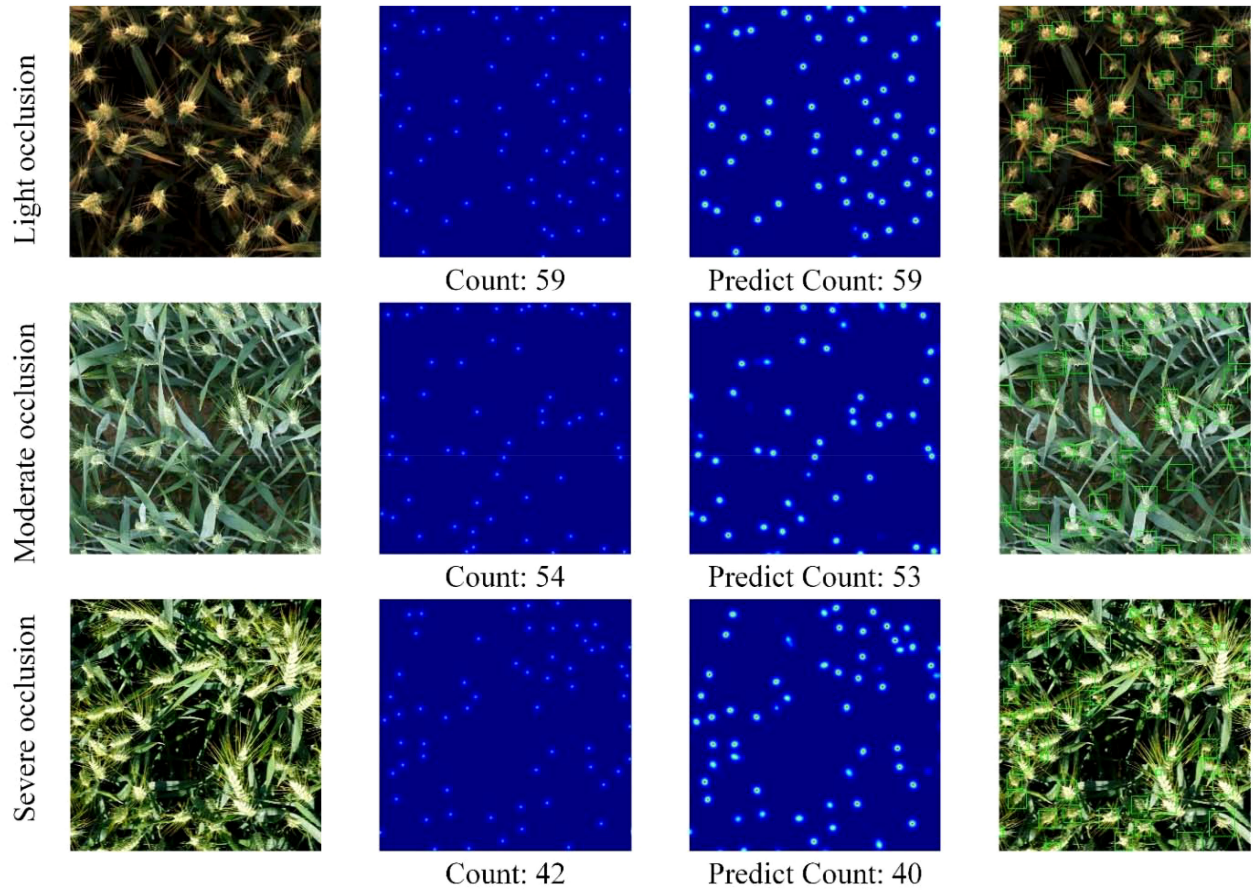


FIGURE 9
Counting results of wheat ears with different occlusion levels.

automatically identify and focus on the wheat ear area in the image, effectively suppressing the interference of background noise and other irrelevant elements. This mechanism not only improves the robustness of the model, but also enables the model to maintain stable performance in complex and changeable field environments.

4.2 Wheat ear counting results with different upsampling algorithms

To verify the effectiveness of the dense upsampling algorithm used in this article, three different upsampling algorithms were designed, including GhostNet+bilinear interpolation, GhostNet +dense upsampling, and GhostNet+deconvolution. When the same parameters are set during model training and testing, the test results are shown in Table 5.

It can be seen from Table 5 that when the backbone feature network is GhostNet, the wheat ear counting effect is the best when the model uses dense upsampling convolution, and its MAE, RMSE, and R^2 are 4.46, 5.87, and 0.9145. Compared with using the bilinear interpolation algorithm and deconvolution operation, the MAE of the model with dense upsampling convolution was reduced by 0.26 and 0.12 respectively, the RMSE was reduced by 0.29 and 0.15, and the R^2 was increased by 0.0722 and 0.033, respectively.

In fact, the bilinear interpolation algorithm may cause the loss of some wheat ear features. Because bilinear interpolation estimates the value of the middle pixel based on the values of the surrounding four pixels, this method performs better when dealing with smoothly changing areas. However, when dealing with complex structures and edge information like wheat ears, the bilinear interpolation algorithm may lose some subtle features, thereby affecting the accuracy of wheat ear counting. The deconvolution operation may indeed produce some unnecessary overlap and aliasing, resulting in image blur and distortion. This is because the deconvolution process is essentially a convolution operation with special parameter settings. If the parameter settings are improper or the convolution kernel design is unreasonable, the above problems may occur. In addition, deconvolution itself will introduce a lot of matrix multiplications during backpropagation, which may indeed increase the risk of gradient disappearance or explosion, thereby affecting the training and testing effects of the model. The dense upsampling algorithms allows the model to recover more spatial detail information, which plays a crucial role in accurately locating and counting wheat ears. This algorithm effectively improves the image resolution, enhances the model's ability to identify wheat ears, and further optimizes the counting results.

TABLE 5 Counting results of wheat ears by different upsampling algorithms.

Upsampling algorithm	MAE	RMSE	R^2
GhostNet + bilinear interpolation	4.72	6.16	0.8423
GhostNet + deconvolution	4.58	6.02	0.8815
GhostNet + dense upsampling	4.46	5.87	0.9145

4.3 Comparison of wheat ear counting results based on different models

To evaluate the wheat ear counting performance of the FIDMT-GhostNet model proposed in this article, we selected three other density estimation models based on convolutional neural networks to compare with the model proposed in this article, including MCNN (Zhang et al., 2016), CSRNet (Li et al., 2018), and FIDTM (Liang et al., 2021). The data used for training and testing of these three models are distributed according to Table 1, while ensuring that the remaining parameter settings remain consistent to obtain the test results of wheat ear counting. The results of wheat ear counting are shown in Table 6. Among them, the leftmost column shows several classic counting network models and our model. The second, third and fourth columns are MAE, RMSE and R^2.

As can be seen from Table 6, the MAE of FIDMT-GhostNet reaches 4.46, which is 4.05, 2.81, lower than that of MCNN, CSRNet, respectively, and is close to that of FIDTM. The RMSE of FIDMT-GhostNet is 5.87, which is 6.61, 2.82, and 0.37 lower than that of MCNN, CSRNet, and FIDTM. The R^2 of FIDMT-GhostNet is 0.9145, which is 0.4301, 0.2046, and 0.0513 higher than that of MCNN, CSRNet, and FIDTM. Therefore, experimental results show that our model outperforms other models on multiple evaluation metrics.

We believe that this result may be due to the following two reasons. On the one hand, the irregular distribution of wheat ears and the large differences in size, shape, density, aspect ratio, etc. of wheat ears make it difficult for MCNN and CSRNet networks to generate high-quality label density maps through Gaussian kernel functions. Of course, in future research, the settings of the multi-column convolution kernel of the MCNN model and the hole rate of CSRNet need to be further optimized to better extract the effective features of wheat ears, thereby improving the accurate counting of wheat ears.

On the other hand, we believe that the good performance of FIDMT-GhostNet comes from the introduction of the Ghost module and the dense upsampling convolution module, which enables the network to adapt to the diversity of wheat ears and focus on wheat ears more effectively. In addition, the FIDTM network can solve the current situation of wheat ear scale changes

TABLE 6 Wheat ear count results for different models.

Model	MAE	RMSE	R^2
MCNN	8.51	12.48	0.4844
CSRNet	7.27	8.69	0.7099
FIDTM	4.39	6.24	0.8632
Ours	4.46	5.87	0.9145

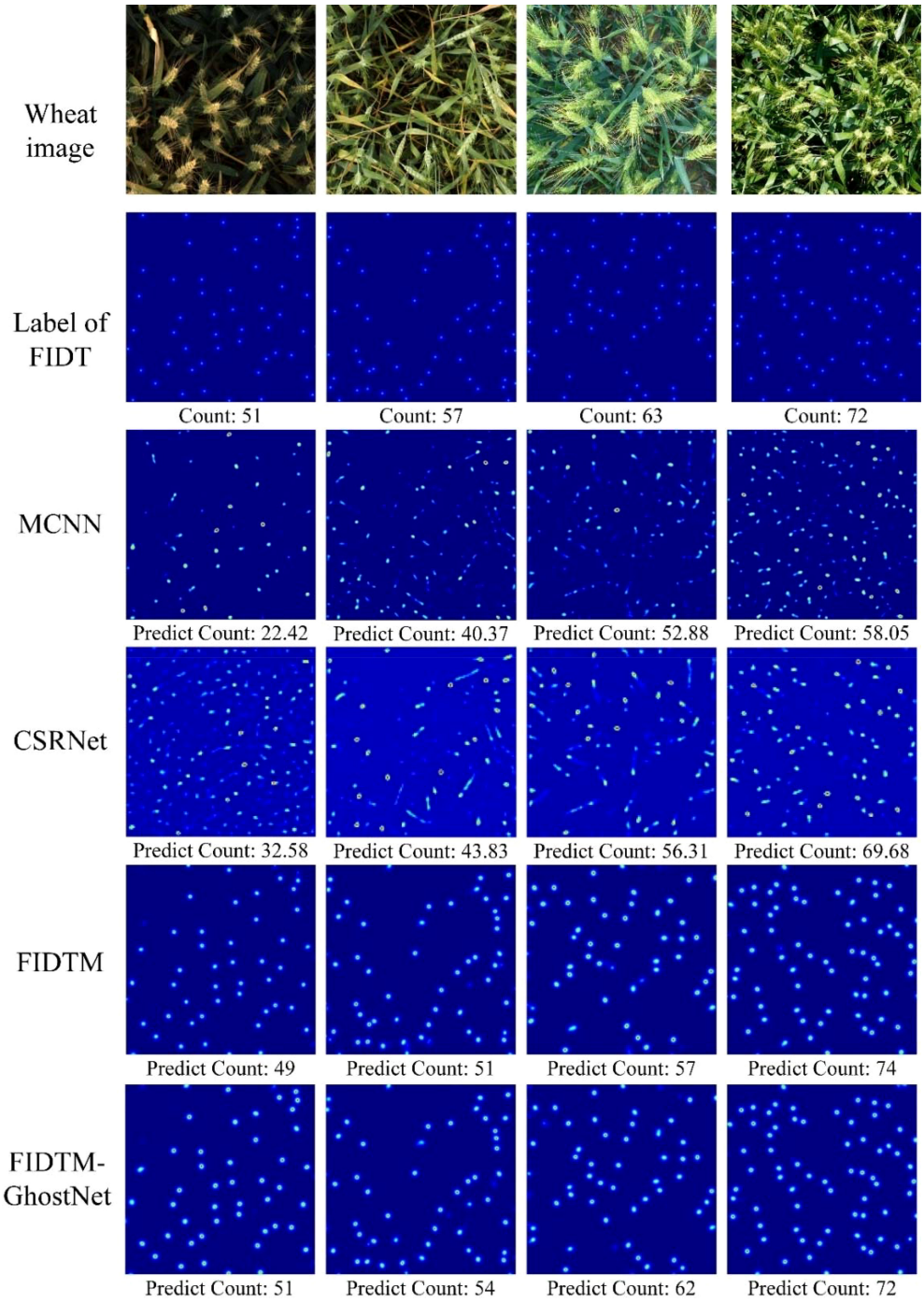


FIGURE 10
Results of wheat ear counts for different models.

and unbalanced density distribution. We also found that the MAE and RMSE of FIDMT-GhostNet model test results were reduced by 63% and 48.04% respectively compared to CSRNet, and R^2 increased by 22.37%. The possible reason is that the backend of the CSRNet model uses atrous convolution to expand the receptive field, which affects the extraction of small target features of wheat ears. It may be that the atrous convolution in the back end of the CSRNet model expands the receptive field, which affects the extraction of small target features of wheat ears.

Figure 10 shows random samples of specific density estimation results for the above four models. The wheat ear counting results of the four density estimation models in Figure 10. We found that the wheat ear density map predicted based on these four models compared with the ground truth label density map, the errors from large to small are MCNN, CSRNet, FIDTM and FIDMT-GhostNet, respectively. Among them, the number of wheat ears predicted based on the FIDMT-GhostNet model is closest to the ground truth. In the third row of Figure 10, the wheat ear density

map estimated based on MCNN was showed. In this part, we see that the estimated density map in the third row is quite different from the real density map in the second row, which indicates that the wheat ear counting effect of the multi-column convolutional neural network MCNN is poor. The possible reason is that MCNN has network structure redundancy, which results in low feature extraction efficiency. Although it has successfully overcome problems such as target occlusion and scale differences to a certain extent, it is difficult to effectively extract the global features of complex wheat field scenes. To solve the above problems, Li et al. (2018) introduced a single-column convolutional neural network CSRNet, which uses dilated convolution to expand the receptive field. In particular, as the depth of the convolutional neural network deepens, the performance of the network structure will gradually increase. Compared with the multi-column structure, the single column convolutional neural network has certain advantages.

As can be seen from Figure 10, the detection accuracy of CSRNet model with the single column is higher than that of MCNN model with multi-column structure. At the same time, we also provide the density map estimated using FIDMT-GhostNet in the sixth row of Figure 10, which shows that the FIDMT-GhostNet model using the dense upsampling algorithm has better results in counting wheat ears in the field. In summary, compared with other common density map models, the FIDMT-GhostNet model proposed in this article has the highest accuracy and the lowest RMSE, indicating that it has good robustness.

4.4 Performance comparison of different models

Table 7 shows the performance parameters of the four models. Among them, the Parameter, FLOPs, Model Size, and FPS of the proposed FIDMT-GhostNet reached 8.42M, 134.09G, 96.8568MB, and 9.26 respectively. Compared with CSRNet, the Parameter of our model is reduced by 48.2% and the FLOP is reduced by 69.1%. Previous studies have shown that Bao et al. (2020) used CSRNet to count wheat. The poor performance of wheat ear counting based on the CSRNet model is mainly due to its deep network structure, which results in a large number of parameters. Although this design can capture more complex wheat ear features in images, it may also cause problems such as high computational complexity, difficulty in model training, and overfitting. Therefore, when counting wheat ears in the field, it is necessary to weigh the balance between the depth and performance of the network to achieve a more efficient wheat ear counting task.

TABLE 7 Wheat ear count results for different models.

Model	Parameter (M)	FLOPs (G)	Model Size (MB)	FPS
MCNN	0.13	28.23	0.5186	52.63
CSRNet	16.26	433.36	62.0512	8.92
FIDTM	66.58	569.59	763.5506	3.13
Ours	8.42	134.09	96.8568	9.16

M/G in the table represent 10^6/10^9 respectively.

According to the wheat ear counting efficiency evaluation indicators in Table 7, MCNN performs well for the Parameter, FLOPs, Model Size, and FPS performance evaluation results of the model. On the one hand, MCNN removes the fully connected layer, resulting in reduced network parameters and a simple structure. In the experimental results, MCNN has the smallest Parameter, reaching 0.13M. Model Size reaches 0.5186MB. On the other hand, the MCNN model contains multiple columns of networks, so network training takes longer than an end-to-end network. This results in the performance indicators of wheat ear counting results based on MCNN, including MAE, RMSE, and R^2, being worse than the other three models.

In addition, it can be seen from Table 7 that the number of parameters in the FIDTM model is 66.58M, and its FLOPs are 569.59G. The parameter size of lightweight FIDMT-GhostNet is 8.42M, and the FLOPs are 134.09G. Compared with the original model, the Parameter, FLOPs and FPS of our proposed lightweight model were reduced by 87.4%, 76.5% and 87.3% respectively. In fact, the FIDTM model has good robustness in scenarios with unbalanced density distribution, but it counts wheat ears evenly distributed in the field. After extensive compression, some parameters in the model were removed, which improved the generalization ability of the FIDMT-GhostNet model, increasing R^2 by 5.6% and reducing RMSE by 5.9%.

5 Conclusions and future work

In this study, an automatic positioning and counting method based on FIDMT-GhostNet was proposed to address the challenges faced by counting wheat ears, including complex backgrounds, dense ears, and different sizes. This method achieves precise positioning and counting of wheat ears through multi-scale feature extraction and point labeling networks, combined with dense upsampling and local maximum detection strategies. Three wheat ear databases including WEC, WEDD and GWHD were used for model training and testing. Experimental results show that the FIDMT-GhostNet model achieved high accuracy on the wheat image data set, and the number of parameters was small. RMSE and R^2 reached 5.87 and 0.9145 respectively, and the number of parameters and FPS reached 8.42M and 9.16 respectively. Therefore, the experimental results show that the FIDMT-GhostNet has good robustness for wheat ear counting. With the continuous advancement of technology, the wheat ear counting model based on FIDMT-GhostNet is expected to be directly deployed to edge devices and play a greater role in precision agriculture with fewer parameters, faster inference speed, and good counting performance.

To further enhance the robustness and generalization ability of the model, we plan to continue to collect more wheat datasets in the future. These new datasets will not only focus on the increase in quantity, but also emphasize the diversity of data, covering wheat images of different varieties, different growth stages, and different environmental conditions, to ensure that the model can perform well in various practical scenarios.

GhostNet has achieved remarkable results in reducing the computational requirements of the model with its efficient network structure. To ensure that this combined model can be widely used in real-world scenarios, especially in resource-constrained agricultural environments, we will conduct in-depth analysis and optimize its computational efficiency. At the same time, in view of the diversity of agricultural image data scale and complexity, exploring how to adjust the model structure to better adapt to these data and enhance the scalability of the model will become a key direction for future research. To this end, we can use model compression and acceleration techniques, such as parameter pruning, quantization, and hardware acceleration, to further optimize the performance of the combined model. These advanced technologies can significantly reduce the storage occupancy and computational requirements of the model without significantly sacrificing model performance, thereby providing strong support for the widespread application of the combined model in resource-constrained agricultural environments.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

BY: Writing – review & editing, Writing – original draft. RC: Writing – original draft, Software, Methodology. ZG: Writing – original draft, Software, Methodology. HZ: Writing – original draft, Data curation.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Major Science and Technology Projects in Anhui Province (202203a06020007), the Opening Project of Key Laboratory of Power Electronics and Motion Control of Anhui Higher Education Institutions (PEMC2001), the Open Fund of State Key Laboratory of Tea Plant Biology and Utilization (SKLTOF20200116), Anhui Province Digital Agriculture Industry Technology System Post Expert (23548009), Anhui Province Graduate Innovation and Entrepreneurship Practice Project (2023cxcysj045).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Atamanyuk, I., Havrysh, V., Nitsenko, V., Diachenko, O., Tepliuk, M., Chebakova, T., et al. (2023). Forecasting of winter wheat yield: A mathematical model and field experiments. *Agriculture* 13 (1), 41. doi: 10.3390/agriculture13010041

Bao, W., Zhang, X., Hu, G., Huang, L., Liang, D., and Lin, Z. (2020). Estimation and counting of wheat ears density in field based on deep convolutional neural network. *Trans. CSAE* 36, 186–193. doi: 10.11975/j.issn.1002-6819.2020.21.022

Dandrifosse, S., Ennadifi, E., Carlier, A., Gosselin, B., Dumont, B., and Mercatoris, B. (2022). Deep learning for wheat ear segmentation and ear density measurement: From heading to maturity. *Comput. Electron. Agric.* 199, 107161. doi: 10.1016/j.compag.2022.107161

Faisal, S., Pan, Z., and Junaid, F. A. (2023). Transpose convolution based model for super-resolution image reconstruction. *Appl. Intell.* 53, 10574–10584. doi: 10.1007/s10489-022-03745-4

Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., and Xu, C. (2019). "GhostNet: More features from cheap operations," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*). (Seattle, WA, USA). 1577–1586.

Hasan, M. M., Chopin, J. P., Laga, H., and Miklavcic, S. J. (2018). Detection and analysis of wheat spikes using convolutional neural networks. *Plant Methods* 14, 1–13. doi: 10.1186/s13007-018-0366-8

Li, R., and Wu, Y. (2022). Improved YOLO v5 wheat ear detection algorithm based on attention mechanism. *Electronics* 11 (11), 1673. doi: 10.3390/electronics11111673

Li, Y., Zhang, X., and Chen, D. (2018). "Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Salt Lake City, UT, USA), 1091–1100.

Liang, D., Xu, W., Zhu, Y., and Zhou, Y. (2022). Focal inverse distance transform maps for crowd localization and counting in dense crowd. *arXiv 2021 arXiv preprint arXiv:2102.07925*, 25, 6040–6052.

Lu, H., Liu, L., Li, Y. N., Zhao, X. M., Wang, X. Q., and Cao, Z. G. (2022). TasselNetV3: Explainable plant counting with guided upsampling and background suppression. *IEEE Trans. Geosci. Remote Sens.* 60, 3058962. doi: 10.1109/TGRS.2021.3058962

Ma, J., Li, Y., Liu, H., Du, K., Zheng, F., Wu, Y., et al. (2020). Improving segmentation accuracy for ears of winter wheat at flowering stage by semantic segmentation. *Comput. Electron. Agric.* 176, 105662. doi: 10.1016/j.compag.2020.105662

Ma, J., Li, Y., Liu, H., Wu, Y., and Zhang, L. (2022). Towards improved accuracy of UAV-based wheat ears counting: A transfer learning method of the ground-based fully convolutional network. *Expert Syst. Appl.* 191, 116226. doi: 10.1016/j.eswa.2021.116226

Madec, S., Jin, X., Lu, H., De, S., Liu, S., Duyme, F., et al. (2019). Ear density estimation from high resolution RGB imagery using deep learning technique. *Agric. For. Meteorol.* 264, 225–234. doi: 10.1016/j.agrformet.2018.10.013

Misra, T., Arora, A., Marwaha, S., Chinnusamy, V., Rao, A. R., Jain, R., et al. (2020). SpikeSegNet-a deep learning approach utilizing encoder-decoder network with hourglass for spike segmentation and counting in wheat plant from visual imaging. *Plant Methods* 16, 1–20. doi: 10.1186/s13007-020-00582-9

Qiao, Y., Guo, Y., and He, D. (2023). Cattle body detection based on YOLOv5-ASFF for precision livestock farming. *Comput. Electron. Agric.* 204, 107579. doi: 10.1016/j.compag.2022.107579

Simão, L. M., Peterson, D., Roozeboom, K. L., Rice, C. W., Du, J., Lin, X., et al. (2023). Crop rotation and tillage impact yield performance of soybean, sorghum, and wheat. *Agron. J.* 115, 658–673. doi: 10.1002/agj2.v115.2

Soh, K., Zhao, L., Peng, M., Lu, J., Sun, W., and Tongngern, S. (2024). SAR marine oil spill detection based on an encoder-decoder network. *Int. J. Remote Sens.* 45, 587–608. doi: 10.1080/01431161.2023.2299274

Sun, J., Yang, K., Luo, Y., Shen, J., Wu, X., and Qian, L (2021). Method for the multiscale perceptual counting of wheat ears based on UAV images. *Trans. Chin. Soc. Agric. Eng.* 37, 136–144. doi: 10.11975/j.issn.1002-6819.2021.23.016

Sun, K., Xiao, B., Liu, D., and Wang, J. (2019). "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Long Beach, CA, USA). 5693–5703.

Wang, D., Fu, Y., Yang, G., Yang, X., Liang, D., Zhou, C., et al. (2019). Combined use of FCN and Harris corner detection for counting wheat ears in field conditions. *IEEE Access* 7, 178930–178941. doi: 10.1109/Access.6287639

Wang, D., Zhang, D., Yang, G., Xu, B., Luo, Y., and Yang, X. (2021a). SSRNet: In-field counting wheat ears using multi-stage convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* 60, 1–11. doi: 10.1109/TGRS.2021.3093041

Wang, Y., Qin, Y., and Cui, J. (2021b). Occlusion robust wheat ear counting algorithm based on deep learning. *Front. Plant Sci.* 12, 645899. doi: 10.3389/fpls.2021.645899

Wu, W., Zhong, X., Lei, C., Zhao, Y., Liu, T., Sun, C., et al. (2023). Sampling survey method of wheat ear number based on UAV images and density map regression algorithm. *Remote Sens.* 15, 1280. doi: 10.3390/rs15051280

Xu, X., Geng, Q., Gao, F., Xiong, D., Qiao, H., and Ma, X. (2023). Segmentation and counting of wheat spike grains based on deep learning and textural feature. *Plant Methods* 19, 77. doi: 10.1186/s13007-023-01062-6

Yang, B., Gao, Z., Gao, Y., and Zhu, Y. (2021). Rapid detection and counting of wheat ears in the field using YOLOv4 with attention module. *Agronomy* 11, 1202. doi: 10.3390/agronomy11061202

Zhang, J., Min, A., Steffenson, B. J., Su, W., Hirsch, C. D., Anderson, J., et al. (2022a). Wheat-net: An automatic dense wheat spike segmentation method based on an optimized hybrid task cascade model. *Front. Plant Sci.* 13, 834938. doi: 10.3389/fpls.2022.834938

Zhang, Z., Qiao, Y., Guo, Y., and He, D. (2022b). Deep learning based automatic grape downy mildew detection. *Front. Plant Sci.* 13, 872107. doi: 10.3389/fpls.2022.872107

Zhang, Y., Zhou, D., Chen, S., Gao, S., and Ma, Y. (2016). "Single-image crowd counting via multi-column convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Las Vegas, NV, USA.) 589–597.

Zhou, Y., Chang, H., Lu, X., and Lu, Y. (2022). Denseunet: improved image classification method using standard convolution and dense transposed convolution. *Knowl. Based Syst.* 254, 109658. doi: 10.1016/j.knosys.2022.109658

Zou, R., Zhang, Y., Chen, J., Li, J., Dai, W., and Mu, S. (2023). Density estimation method of mature wheat based on point cloud segmentation and clustering. *Comput. Electron. Agric.* 205, 107626. doi: 10.1016/j.compag.2023.107626

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 31 October 2024
DOI 10.3389/fpls.2024.1460060

Check for updates

OPEN ACCESS

EDITED BY
Liujun Xiao,
Nanjing Agricultural University, China

REVIEWED BY
Yunchao Tang,
Dongguan University of Technology, China
Xiujuan Chai,
Chinese Academy of Agricultural Sciences (CAAS), China

*CORRESPONDENCE
Qian Wu
wuqian@jaas.ac.cn
Ni Ren
rn@jaas.ac.cn

RECEIVED 05 July 2024
ACCEPTED 30 September 2024
PUBLISHED 31 October 2024

CITATION
Zhou L, Hu A, Cheng Y, Zhang W, Zhang B, Lu X, Wu Q and Ren N (2024) Barrier-free tomato fruit selection and location based on optimized semantic segmentation and obstacle perception algorithm.
Front. Plant Sci. 15:1460060.
doi: 10.3389/fpls.2024.1460060

COPYRIGHT
© 2024 Zhou, Hu, Cheng, Zhang, Zhang, Lu, Wu and Ren. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Barrier-free tomato fruit selection and location based on optimized semantic segmentation and obstacle perception algorithm

Lingli Zhou[1,2], Anqi Hu[1,2], Yawen Cheng[1,2], Wenxiang Zhang[1,2], Bingyuan Zhang[1,2], Xinyu Lu[1,2], Qian Wu[1,2]* and Ni Ren[1,2]*

[1]Institute of Agricultural Information, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [2]Key Laboratory of Intelligent Agricultural Technology (Changjiang Delta), Ministry of Agriculture and Rural Affairs, Nanjing, China

With the advancement of computer vision technology, vision-based target perception has emerged as a predominant approach for harvesting robots to identify and locate fruits. However, little attention has been paid to the fact that fruits may be obscured by stems or other objects. In order to improve the vision detection ability of fruit harvesting robot, a fruit target selection and location approach considering obstacle perception was proposed. To enrich the dataset for tomato harvesting, synthetic data were generated by rendering a 3D simulated model of the tomato greenhouse environment, and automatically producing corresponding pixel-level semantic segmentation labels. An attention-based spatial-relationship feature extraction module (SFM) with lower computational complexity was designed to enhance the ability of semantic segmentation network DeepLab v3+ in accurately segmenting linear-structured obstructions such as stems and wires. An adaptive K-means clustering method was developed to distinguish individual instances of fruits. Furthermore, a barrier-free fruit selection algorithm that integrates information of obstacles and fruit instances was proposed to identify the closest and largest non-occluded fruit as the optimal picking target. The improved semantic segmentation network exhibited enhanced performance, achieving an accuracy of 96.75%. Notably, the Intersection-over-Union (*IoU*) of wire and stem classes was improved by 5.0% and 2.3%, respectively. Our target selection method demonstrated accurate identification of obstacle types (96.15%) and effectively excluding fruits obstructed by strongly resistant objects (86.67%). Compared to the fruit detection method without visual obstacle avoidance (Yolo v5), our approach exhibited an 18.9% increase in selection precision and a 1.3% reduction in location error. The improved semantic segmentation algorithm significantly increased the segmentation accuracy of linear-structured obstacles, and the obstacle perception algorithm effectively avoided occluded fruits. The proposed method demonstrated an appreciable ability in precisely selecting and locating barrier-free fruits within non-structural environments, especially avoiding

fruits obscured by stems or wires. This approach provides a more reliable and practical solution for fruit selection and localization for harvesting robots, while also being applicable to other fruits and vegetables such as sweet peppers and kiwis.

KEYWORDS

harvesting robot, image semantic segmentation, obstacle perception, deep learning, fruit selection, positioning, tomato

1 Introduction

Fruit picking is a time-consuming and laborious work, accounting for 33%-50% of the total production workload. With the advancement of agricultural modernization, the research on harvesting robots has attracted much attention. Harvesting robots will be the key technology to address the issue of labor shortage in the future. This advancement encompasses multidisciplinary techniques spanning kinematics, control science, machine vision, and behavioral science. Following several years of development, substantial progress has been made in addressing key technical challenges such as path planning (Ye et al., 2023; Ghosh et al., 2019; Tang et al., 2024), systematic control (Chen et al., 2024), target recognition and positioning (Guo et al., 2023; Yang et al., 2023), and picking sequence planning (Kurtser and Edan, 2020). Nowadays, various prototypes of single-fruit harvesting robots have been developed, such as tomato harvesting robots (Jun et al., 2021), apple harvesting robots (Zhang et al., 2021; Silwal et al., 2017), kiwifruit harvesting robots (Williams et al., 2019) and so on. The visual servo harvesting robot (Williams et al., 2019; Miao et al., 2023; Jiao et al., 2022) stand out as one of the research hotspots due to their significant advantages in autonomy, precision, and adaptability. The first and most crucial step for these camera-based robots is to detect and locate the target to be harvested in the visual image, and obtain the position of the target in three-dimensional (3D) through coordinate transformation with the help of depth maps or other auxiliary information. Despite significant breakthroughs (Montoya-Cavero et al., 2022; Rong et al., 2022) have been made in fruit recognition and positioning based on visual images, the uncertainty in fruit growth and the complexity of unstructured orchard environments still lead to many problems for the visual system of harvesting robots.

Before deep learning was widely used, traditional machine learning algorithms combined with image processing were the most common methods for fruit recognition. Through a series of pre-processing, such as color space transformation, image denoising, edge detection and region growth (Ouyang et al., 2013; Wachs et al., 2010; Zhao et al., 2016), different levels of features were extracted and served as the input of the machine learning methods for image classification. However, the traditional machine learning algorithms are difficult to deal with images collected in complex natural environments, and their accuracy are affected by light intensity and ray shadow, so their robustness are not enough to meet practical needs.

Compared to traditional machine learning, deep learning characterized by its superior representation, learning, and generalization abilities, has garnered significant attention and widespread application. The deep learning-based computer vison technology has significantly advanced the development of harvesting robots and accelerated their practical application. The most commonly used methods (Divyanth et al., 2022; Kuznetsova et al., 2020; Yan et al., 2021) for harvesting robots are to customize position calculation algorithms for specific species by improving the object detection models, including Faster Regional-Convolutional Neural Network (Faster R-CNN) (Ren et al., 2016) and YOU LOOK ONLY ONCE (YOLO) (Redmon et al., 2016). In order to achieve the real-time detection of apples or oranges, Kuznetsova et al (Kuznetsova et al., 2020). designed the pre- and post-processing techniques based on YOLO v3 algorithm, which shortened the average detection time to 19 ms. Although the algorithms based on object detection have achieved promising results in terms of speed, they can only estimate the location and size of the target. For some soft-rind fruits harvested by cutting the stems to prevent damage to the epidermis, more appearance information, such as the contour of targets, is required to determine the positions of the cutting points. The image semantic segmentation and instance segmentation provide good methods to meet this kind of harvesting demand. The work presented by Yu et al (Yu et al., 2019). was a typical case of applying the instance model Mask Region Convolutional Neural Network (Mask R-CNN) (He et al., 2020) to strawberry harvesting, which calculated picking points on mask images generated by Mask R-CNN. In order to recognize and segment overlapped apples, Jia et al (Jia et al., 2020). optimized Mask R-CNN by combining Residual Network (ResNet) (He et al., 2016) with Densely Connected Convolutional Networks (DenseNet) as an alternative to its original backbone network for reducing input parameters, which was ideally effective in terms of both speed and accuracy of target positioning tested on a random test set. It is a trend to use semantic segmentation (Li et al., 2021; Kang and Chen, 2020) and instance segmentation (Luo et al., 2022; Zheng et al., 2021) to solve target recognition and location in the vision system of harvesting robots, as it can provide more appearance information and realize the separation of fruits and background, and individual targets.

Although these methods have achieved notable breakthroughs, they still face difficulties in addressing the challenges presented by

complex growth environments. We have noticed that the tomato fruits grown under natural conditions are often obscured by various objects such as leaves, stems, wires and other fruits. The presence obstacles, particularly wires and stems, poses a significant challenge to mechanical harvesting by impeding the movement of robotic arms. The entanglement between these slender objects with strong resistance and the robotic arm may lead to fatal errors, so the issue of obstacle occlusion has gradually attracted the attention of researchers. Divyanth et al (Divyanth et al., 2022). has improved the Faster R-CNN model by adding an attention mechanism to detect non-occluded coconuts and leaf-occluded coconuts. However, this method cannot meet the perceptual need for more intricate scenarios. The information of the obstacles, including stems, wires, branches and petioles, need to be captured in semantic level because they directly affect the selection of picking targets and the path planning of picking execution agencies. In recent years, semantic segmentation algorithms have been employed to segment objects with linear structure (such as stems, branches, etc.) in unstructured environments. Song et al. (Song et al., 2021) proposed a branches and wires segmentation and reconstruction method for kiwifruit. Wang et al. (Wang et al., 2023) proposed a parallel network structure (DualSeg) to segment branches and fruits for grapes. Although these studies have successfully implemented pixel-level perception of obstacles through image semantic segmentation, their approach lacks the integration of both obstacle perception and fruit instance discrimination, as well as an analysis of fruit occlusion in the orchard environment.

Thus, we proposed our method to locate fruit instances while recognizing obstacles based on the following two considerations. Firstly, in order to distinguish fruits from different obstacles, especially linear-structured obstacles, an improved image semantic segmentation algorithm was proposed. Due to the slender structure of major obstacles such as stems and wires, we added spatial-relationship features to the semantic network, which was inspired by the research (Pan et al., 2018; Gioi et al., 2012) used to identify slender objects such as lane lines and transmission lines. Different from the slice-by-slice convolution used in Spatial Convolutional Neural Network (SCNN) (Pan et al., 2018), we redesigned an attention-based spatial-relationship feature extraction module with lower computational complexity to enhance the ability of semantic segmentation network to recognize obstacles with strong shape priors. The module skillfully applied spatial attention masks to transmit information in the rows and columns of features, thereby changing the attention allocation of network and achieving more attention to slender structural objects to improve their segmentation accuracy. Secondly, an adaptive K-means (MacQueen, 1967) pixel clustering algorithm was designed to segment fruit instances based on the characteristic of different fruit depths and positions, effectively addressing the challenges associated with unsatisfactory clustering performance due to a fixed K value and the uncertain number of fruits in each image. Subsequently, a straightforward yet efficient barrier-free fruit selection algorithm combining the information of obstacles and fruit instances was proposed to select the closest and largest non-occluded fruit as the ultimately picking target. The feasibility of our method was validated on our Tomato dataset, which performed well in both semantic segmentation, and selection and location of barrier-free fruits.

2 Materials and methods

2.1 Tomato dataset

2.1.1 Image acquisition

The images were acquired at the Tomato Intelligent Production Greenhouse of Jiangsu Academy of Agricultural Sciences, Nanjing, China (32.03° N, 118.87° E) in 2022-2023. All tomatoes were grown using soilless culture, and two tomato cultivars “Sufen No.6” and “Fatalong” were selected for data acquisition.

A tomato harvesting robot platform composed of a picking robotic arm, a visual system, a control system and a walking

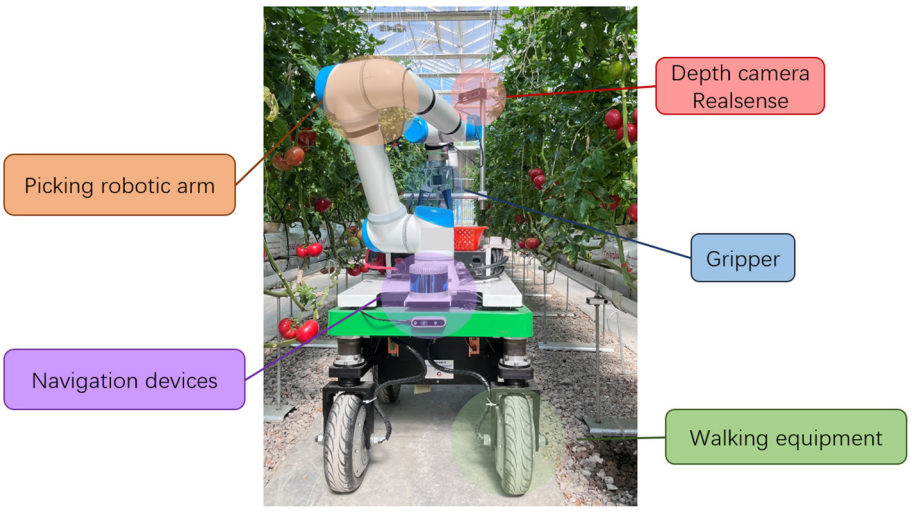


FIGURE 1
The developed tomato harvesting robot.

equipment was developed (Figure 1), and the intelligent picking operations, such as picking path planning, picking target identification and positioning, and non-destructive picking, were realized. A depth camera Realsense (D435i, Intel, America) mounted on the tomato harvesting robot was used to automatically capture RGB images and depth images with a resolution of 600×800 pixels. The tomato harvesting robot stopped every 1 meter for tomato data collection. The images were taken irregularly between 9 a.m. and 5 p.m. under different weather conditions (sunny and cloudy) from February to April. Images that were too bright, too dark or in other colors that do not meet the standards were removed, and finally 170 RGB images and corresponding depth maps were selected.

2.1.2 Image annotation and dataset production

Due to the limited number of empirical images, we used the synthetic data generated by the method provided by Barth et al. (Barth et al., 2018) to pre-train the weights of the semantic segmentation model. 3250 synthetic tomato images and corresponding pixel-level synthetic semantic segmentation label images were generated by rendering on the 3D simulated model of tomato greenhouse environment produced by Blender. The resolution of all synthetic images was 1600×1200 pixels, and each pixel in the synthetic images corresponds to one of nine classes (background, mature fruits, immature fruits, peduncles, stems, branches and petioles, wires, cuts, leaves) represented by a color in the label images.

170 empirical RGB images were labelled with pixel-level semantic segmentation using the open source software Labelme (Russell et al., 2008), with 120 images as training data, 20 images as validation data, and the remaining 30 images as testing data. The annotators annotated the images from back to front, ensuring that the boundary of each object was not marked repeatedly. Considering the labelling speed and quality, only the first row of plants in the images were labelled, and other distant plants can be regarded as background because they have little effect on the perspective of robots. We omitted the annotation of leaves for empirical images, mainly based on the following two considerations: (i) the leaves were regularly removed during the harvesting seasons, and the presence of a limited number of leaves does not significantly impact the operation of the robots; (ii) the morphology of tomato leaves was complex and irregular, resulting in a significant increase in the workload associated with data annotation. Therefore, there were a total of 8 label classes (background, mature fruits, immature fruits, peduncles, stems, branches and petioles, wires, cuts) in empirical label images and the annotation time for each image took about 1 hour.

Figure 2 shows some examples of empirical data and synthetic data. It can be seen that the synthetic images and empirical images have high similarity, and fruits in both are easily obscured by objects such as wires and stems.

2.2 Proposed method for barrier-free fruit selection and location

In order to solve the problem of barrier-free fruit selection and positioning based on monocular images, a segmentation and location approach for vision system of harvesting robots based on RGB images and depth maps was proposed. The whole flowchart was divided into three phases (Figure 3). In phase 1, an improved semantic segmentation network was used to distinguish fruits,

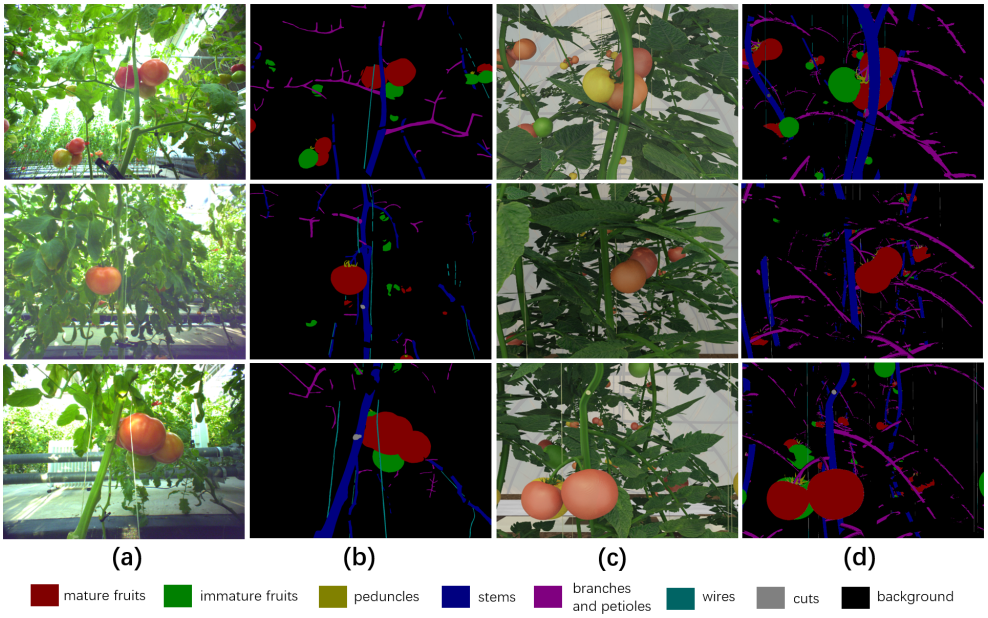


FIGURE 2
Visual comparison between synthetic data and empirical data. **(A)** Three empirical images, **(B)** corresponding semantic segmentation labels of three empirical images, **(C)** three synthetic images, **(D)** corresponding synthetic labels of three synthetic images. For better comparison, the leaves in the synthetic labels are displayed as background.

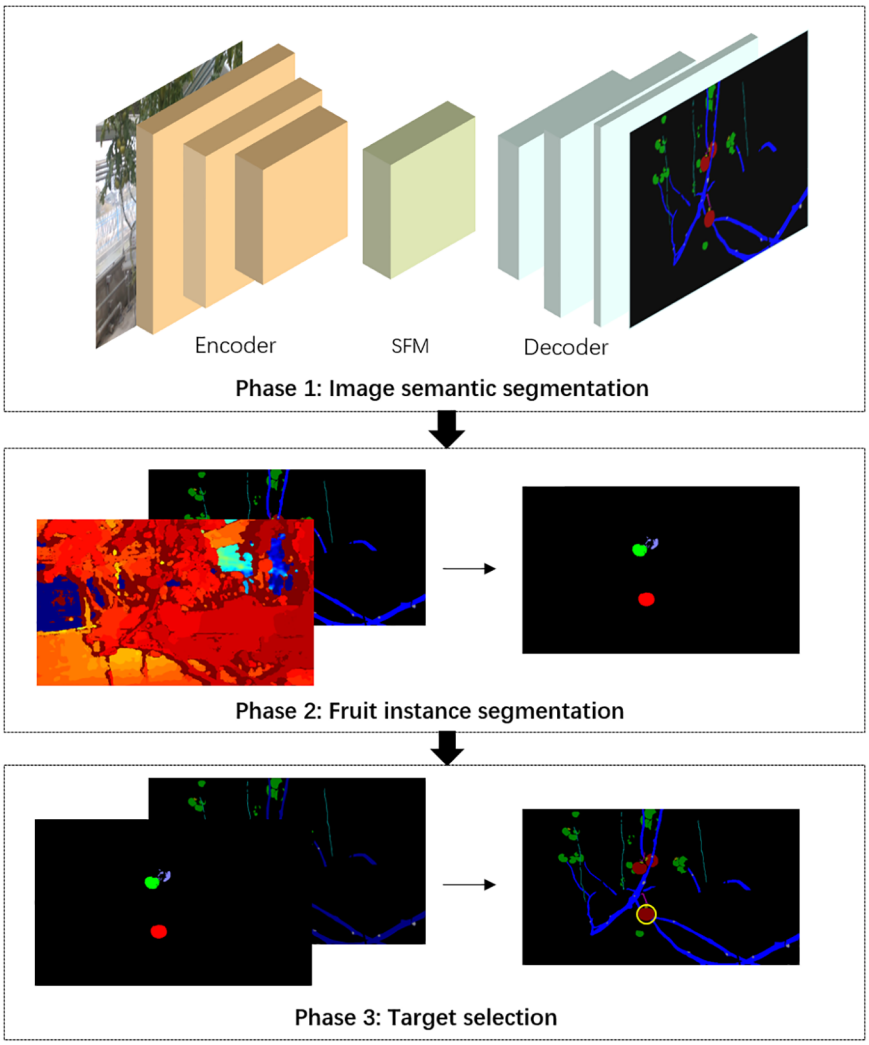


FIGURE 3
The flowchart of the non-occluded fruit selection and location approach for single-fruit harvesting robots.

obstacles and background. The shape and location of the major obstacles, such as stems, branches and petioles, wires, need to be delicately identified. Therefore, we improved the recognition accuracy of obstacles with slender structures by adding spatial-relationship feature module (SFM) into the encoder of DeepLab v3+ (Chen et al., 2017), and achieved high-precision pixel-level segmentation of plant organs and key objects. In phase 2, based on the position information of fruits in the semantic segmentation maps and the depth information provided by depth maps, an adaptive K-means clustering method was applied to cluster pixels of fruits into individual instances. In phase 3, barrier-free harvesting targets were selected by using a proposed obstacle detection method, and the optimal picking target was recommended.

2.2.1 Improved image semantic segmentation network

The widely used DeepLab v3+ (Chen et al., 2017) was adopted as the baseline of the image semantic segmentation network, and its encoder was optimized to enhance the segmentation ability of slender obstacles. Inspired by SCNN for lane detection (Pan et al., 2018), we proposed a spatial-relationship feature module (SFM) and strengthened the spatial relationships of pixels across rows and columns for objects with strong shape priors but weak appearance coherences during feature extraction.

Given the demand for real-time inference in agricultural machinery automation, it is imperative to develop accurate, lightweight, and fast models. Considering the computational complexity of SCNN, we proposed a lighter spatial information perception module drawing on the idea of the spatial attention mechanism (Jaderberg et al., 2015). Generally speaking, the spatial attention mechanism assigns different weights to pixels through a mask, thereby directing more attention towards regions of interest. In our study, it is essential to increase space-attention towards stems and wires to enhance their segmentation accuracy. These objects exhibit obvious directionality in both horizontal and vertical directions in the images. In light of these considerations, we proposed our spatial-relationship feature module (SFM). This module was designed to augment the spatial attention mechanism

by incorporating the inherent directional attributes of the target objects, thereby enhancing the segmentation accuracy.

SFM presented a novel methodology for feature transformation that differs from the traditional slice-by-slice convolution utilized by SCNN. SCNN operated in four computational directions: top to bottom, bottom to top, right to left, and left to right. In each direction, the features were sliced, with individual convolutions applied to each slice and addition operations conducted between slices. In contrast, SFM utilized two masks derived from stripe pooling of the original features along their spatial directions to facilitate information transmission across the rows and columns of features. This innovative approach significantly reduced the parameter count, thereby enhancing model efficiency. As the masks were derived from the features, the spatial information of the original features was retained. Subsequently, the two masks were dot-multiplied with the original features in their respective responsible directions, thereby reallocating attention to the original features. This process effectively reallocated the focus within the original features, facilitating a more nuanced and context-aware representation of features. As shown in Figure 4, letting $F_o^{(C,H,W)}$ denote the 3-D features generated from the feature extraction network ResNet101 of DeepLab v3+ for an image. The extraction of spatial features was conducted through a dual-axis approach, focusing on both the horizontal and vertical orientations within the image. Horizontal stripe pooling with the kernel size of 1×N compressed $F_o^{(C,H,W)}$ to the size of C×H×1, followed by an average pooling of the channels to obtain smaller features with a size of 1×H×1. Through the nonlinear activation function ReLU, the attention mask $Mask_h^{(1,H,1)}$ on the horizontal dimension was calculated, as described in Equation 1. Then, the mask was dot multiplied by the original feature $F_o^{(C,H,W)}$ to capture the horizontal dependencies and obtain the horizontal spatial relationship feature $F_h^{(C,H,W)}$, as seen in Equation 2. Similarly, the vertical spatial relationship feature $F_v^{(C,H,W)}$ was calculated. $F_o^{(C,H,W)}$, $F_v^{(C,H,W)}$ and $F_h^{(C,H,W)}$ was merged by cascading, and then the output was fed into the convolution layer to generate the final features $F_t^{(C,H,W)}$ with spatial relationships, as seen in Equation 3.

$$Mask_h^{(1,H,1)} = ReLU(AvgPool(StripPool(F_o^{(C,H,W)}))) \quad 1$$

$$F_h^{(C,H,W)} = F_h^{(C,H,W)} \times Mask_h^{(1,H,1)} \quad 2$$

$$F_t^{(C,H,W)} = Conv(Concat(F_o^{(C,H,W)}, F_h^{(C,H,W)}, F_v^{(C,H,W)})) \quad 3$$

where *C*, *H*, *W* stand for the number of channels, rows, and columns respectively, *ReLU* denotes nonlinear activation function, *AvgPool* denotes average pooling, *StripPool* denotes strip pooling, × denotes dot multiplication operation, *Conv* denotes standard convolution and *Concat* denotes cascade operation.

This specially designed module facilitated the transmission of pixel feature information within the layer space and enabled the network pay more attention to the objects with slender structures in both rows and columns. The attention masks were obtained from spatial feature statistics with specific directionality, which help original features perceive information in the row and column space of the layer. Specially, the SFM module has low computational complexity. Taking a feature of size 1×64×64×128 as an example, the computational cost of performing one SCNN operation on it exceeds 75.5 MFLOPs, whereas the computational cost for SFM is approximately 2.6 MFLOPs. Meanwhile, the design of SFM ensures compatibility with existing feature extraction frameworks, allowing for seamless integration following the feature extraction backbone of the segmentation network.

2.2.2 Segmentation for fruit instance

In general, pixels belonging to the same instance have close positions in semantic segmentation maps and similar depth values in depth maps, even if they were partitioned by obstacles. A clustering method based on K-means (MacQueen, 1967) was chosen to divide the pixels of mature fruits into different instances with the help of depth information.

The clustering algorithms could represent a more efficient method in contexts with constrained computational resources. K-means clustering is a straightforward and efficient unsupervised learning method that offers the advantage of applicability to various types of data, thereby making it widely utilized in image segmentation tasks (Zhang and Peng, 2022). The basic idea of the algorithm is to find the optimal partition scheme of K-means clustering by minimizing the loss function. The loss function is the sum of the distances from all the elements to the center of the cluster, we defined our loss function *L* in Equation 4,

$$L = \sum_{j=1}^{k}\sum_{i=1}^{N_j} \| p(X_i) - p(Z_i) \|^2 + \alpha \| d(X_i) - d(Z_j) \|^2, X_i \in S_j \quad 4$$

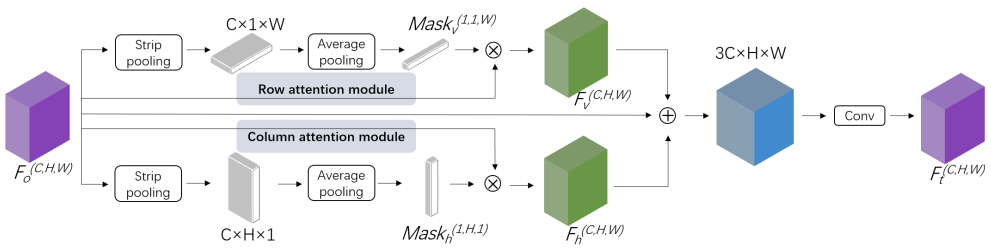


FIGURE 4
Sketch of spatial-relationship feature module (SFM) of our approach.

where, k is the number of clusters, S_j is the cluster set, and N_j and Z_j are the number of pixels and the cluster center belonging to S_j respectively, $p(x)$ denotes the position of x and $d(x)$ denotes the depth value of x, $||\cdot||^2$ is the $L2$-norm, α is used to balance the proportion of two parts.

The clustering centers were continuously optimized through many iterations until they no longer changes. The resolution of the original images needs to be reduced before pixel clustering in order to reduce the noise and improve the operational efficiency. The initial clustering centers were randomly selected from the established mature fruit categories to improve clustering efficiency. Considering the different number of fruits in different images, we adopted an adaptive K-means clustering algorithm for each image. Specifically, we initially selected a larger value as the original K-value based on the count of mature tomatoes in the foreground of the images from previous trials. After clustering on this K-value, the results were checked and the closest clustering centers were merged. Then, the K-value was updated, and the next round of clustering continued until no new K-value was generated.

2.2.3 Target selection and location

A target selection method was proposed to screen the fruit instances obtained in the second phase based on the obstruction information obtained in the first phase. The barrier-free instances were retained and prioritized based on their area and depth values. Ideally, a picking instance would be a connected region (Figure 5A), while the instances occluded by objects (such as stems) would be divided into multiple regions (Figure 5B). Therefore, we first used a basic connected region analysis algorithm (Di Stefano and Bulgarelli, 1999) to label connected regions of the binary image mask generated from each instance. Then, we randomly selected any two points from two non-connected regions, and connected the four points into a quadrilateral, which was defined as a blind spot, as shown in the green area in Figure 5C. If an object divided an instance into two regions, the object was likely to exist in the corresponding blind spot region on the semantic segmentation map. Therefore, all blind spots were sent to semantic segmentation maps to check for the presence of obstacles.

By correlating the blind spots with the semantic segmentation map, the corresponding categories of obstacles, including stems, wires, branches and petioles, were identified. In the absence of a corresponding category within the blind spots, which signified an empty background, the obstacles were initially classified as leaves. This assumption was primarily based on the observation that only leaves or other obstacles could partition a single fruit instance into multiple regions, thereby creating blind spots; conversely, the background itself could not divide fruit instances and generate these blind spots. To enhance the rigor of obstacle detection, we employed the Otsu threshold segmentation algorithm (Yousefi, 2011) based on color features to determine whether the obstacles in the blind spots belong to leaves or unknown categories. Specifically, the Otsu algorithm performed rapid binary segmentation on the remaining portion of the original image identified as background, effectively distinguishing leaves from other elements within that background based on their distinct color features. Subsequently, the obstacles obstructing the fruit were further clarified by referencing the binary segmentation result of the non-fruit area in the blind spots.

When an instance presented obstacles (stems, branches and petioles, wires) in its blind spots, it would be discarded. If an instance was obstructed by leaves, it would still be retained as the obstruction caused by leaves does not affect the mechanical picking. The center coordinates and radii of the remaining instances were obtained by delineating the minimum circumscribed circles of the connected regions. Then, the closest and largest non-occluded fruit was selected as the ultimate harvesting target according to the comprehensive evaluation scores calculated from their areas and depth values. The algorithm flow was shown in Supplementary Material 1.

2.3 Experiments and performance evaluation

2.3.1 The optimized algorithm enhances the performance of image semantic segmentation

The method presented above was tested on the on our Tomato dataset. The whole approach was implemented in the Python programming language. For image semantic segmentation, deep CNNs was implemented on TensorFlow (Cordts et al., 2016),

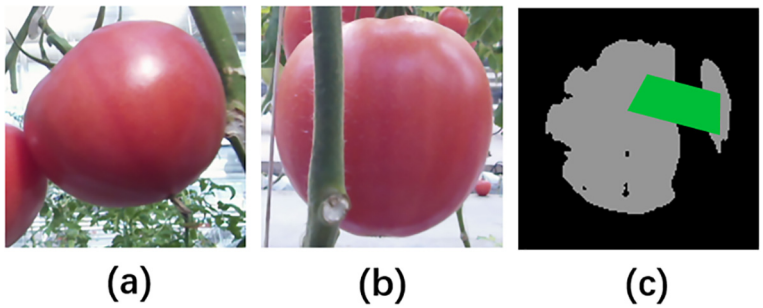


FIGURE 5
Tomato fruits with different regional connectivity and fruit blind spot identification. **(A)** A fruit without any obstruction, **(B)** a fruit occluded by a stem, **(C)** a blind spot generated by two connected regions represented by a green region.

a public deep learning architecture. Adam optimizer (Kingma and Ba, 2014), which is computationally efficient and widely used in many models, was used to improve the performance of the networks. Meanwhile, the "exponential decay" policy was used to control the updating speed of parameters and accelerate the convergence speed of the networks. During the whole training process, original input was randomly cropped. The crop size was 640×640 pixels and batch size was 4. Some layers were added a "dropout" strategy to prevent over-fitting during the training process. The networks were trained with the pre-trained parameters from ImageNet (Russakovsky et al., 2015), and each model converged after approximately 60k iterations. For fruit instances segmentation and the post-processing including the selection of picking targets, their implementation relied on OpenCV, an open-source library for computer vision.

2.3.2 The evaluation of image semantic segmentation

The image semantic segmentation performance of DeepLab v3+ and DeepLab v3+ with SFM were compared. According to the conventional evaluation criteria, Pixel Accuracy (*acc*) and Intersection-over-Union (*IoU*) (Everingham et al., 2015) were adopted as evaluation criteria for image semantic segmentation. The *acc* was defined as the ratio of all correctly classified pixels to all valid pixels in the image (Equation 5). The *IoU* evaluated the similarity between the portion parsed by the network and ground truth related to a specific class (Equation 6).

$$acc = \frac{1}{N}\sum_{i=1}^{N}\frac{R_i}{V_i} \times 100\,\% \quad 5$$

$$IoU = \frac{1}{N}\sum_{i=1}^{N}\frac{P_i \cap G_i}{P_i \cup G_i} \times 100\,\% \quad 6$$

where N is the number of images in the testing set; R_i and V_i are the total number of correctly classified pixels and the total number of valid pixels in image i, respectively; P_i denotes the area predicted as the target class, G_i represents the area of the target class in the ground truth. Therefore, the value of *IoU* was calculated by dividing the intersection area of the two regions by the union area of two regions. *IoU* was a supplement to *acc*, it took into account the class imbalance that existed in most datasets. In addition, the mean intersection-over-union (*mIoU*) of all classes was added as another indicator.

2.3.3 The evaluation of fruit instance segmentation

The effectiveness of the proposed adaptive K-means clustering algorithm was validated by evaluating the fruit instance segmentation performance using different K values. Generally, the number of mature fruits in each image's foreground should not exceed 8, hence we varied the K value within a range from 2 to 8. Precision, Recall and F1 were employed as evaluation metrics for fruit instance segmentation performance assessment. Usually, Precision and Recall are contradictory, with high Precision resulting in low Recall, and vice versa. Therefore, F1 Score was often used to comprehensively measure the performance of a model while balancing Precision and Recall. The three indicators were defined in Equations 7–9,

$$Precision = \frac{N_{TP}}{N_{TP} + N_{FP}} \times 100\,\% \quad 7$$

$$Recall = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100\,\% \quad 8$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\,\% \quad 9$$

where N_{TP} is the number of positive samples correctly identified as positive, N_{FP} is the number of negative samples incorrectly identified as positive, and N_{FN} is the number of positive samples incorrectly identified as negative. If the ratio of the correctly detected area to the ground truth exceeds a threshold, the target was considered to be correctly detected.

For the evaluation of fruit instance segmentation performance, the threshold of *IoU* was set to 0.5, indicating that if the overlapping area between the segmented fruit instance and manually labelled instance exceeded 50% of the latter, it would be marked as a true positive. Normally, *IoU* values exceeding 0.5 could support the harvesting robot for picking operations.

2.3.4 The evaluation of target selection and location approach

The performance of the approach in picking-targets selection was evaluated from two aspects of the target selection and the location precision. *Precision*, *Recall* and *F1* were commonly used metrics for target detection. The central position and radius of the fruit were the key parameters for the harvesting robots to perceive the target position. The positioning error (*PE*) for a target was defined in Equation 10,

$$PE = \frac{\| C_p - C_g \|^2}{R} \times 100\,\% = \frac{\sqrt{(x_p - x_g)^2 + (y_p - y_g)}}{R} \times 100\,\% \quad 10$$

where C_p and C_g and are the center coordinates predicted by the algorithm and the manually labeled center, respectively, and R is the target radius. *PE* was essentially a relative error between positions.

3 Results

3.1 The diversity and robustness of the empirical data

The empirical images were collected at various temporal intervals and under diverse lighting conditions to cope with different harvesting situations. Different weather conditions have a significant impact on the quality of image acquisition and subsequent image processing tasks. Under overcast conditions, inadequate illumination may produce excessively dark images, resulting in a loss of detail, whereas bright light can lead to

overexposure. Collecting images under different weather conditions can effectively enhance the richness and robustness of the image dataset, thereby providing substantial value in improving the model's applicability across different weather conditions.

To assess the effectiveness, diversity and robustness of our empirical data, we created two training datasets by dividing the original dataset into two categories based on the weather conditions. Then, we evaluated the semantic segmentation performance of DeepLab v3+ trained on different datasets, including images captured under diverse weather conditions, as well as those taken in sunny and cloudy weather conditions (Table 1). All three models were pre-trained on the synthetic data and the training images were augmented through geometric transformation and color transformation. The results indicated that the model trained on the dataset comprising images captured under diverse weather conditions significantly outperformed the other two models in terms of accuracy (*acc*) and mean Intersection over Union (*mIoU*), as both of their training datasets exhibited notable biases. This finding demonstrated the diversity and robustness of our empirical training data.

3.2 The optimized algorithm enhances the performance of image semantic segmentation

The performance of image semantic segmentation has a significant impact on the subsequent single-fruit segmentation, obstacles perception and target selection. The performance of the widely used semantic segmentation network DeepLab v3+ and the proposed DeepLab v3+ with SFM were compared on the Tomato dataset. The visual comparison results showed that the segmentation of wires and stems were significantly enhanced by using DeepLab v3+ with SFM (Figure 6), demonstrating that SFM can effectively assist semantic segmentation networks to parse objects with slender structures more accurately.

In the quantitative comparison, we included U-net network (Ronneberger et al., 2015) as an additional baseline network and compared it with the model with SFM. The quantitative analysis results showed that both models with SFM achieved better performance, while DeepLab v3+ with SFM exhibited the best semantic segmentation performance (Table 2). The *acc* of DeepLab v3+ with SFM reached 96.75% and the *mIoU* exceeded that of DeepLab v3+ by a gap of 2.19%. The *IoU* values of most classes have been improved to varying degrees due to the introduction of spatial relationship through SFM (Table 3), especially for obstructions with slender structures. Notably, significant enhancements were observed in the classes of stems and wires (Table 3), where U-net with SFM showed a respective increase of 2.48% and 5.07% and DeepLab v3+ with SFM exhibited an increase of 2.30% and 5.00%. Overall, the optimized approach resulted in more accurate and smooth segmentation due to the spatial-relationship perception ability of SFM, which was beneficial to the recognition of obstructed fruits.

Moreover, the improvement of semantic segmentation performance by introducing SFM was also validated on a Pepper dataset (Barth et al., 2018) (Supplementary Material 2). The *acc* of DeepLab v3+ with SFM reached 90.56% and the *mIoU* exceeded that of DeepLab v3+ by a gap of 4.93% (Supplementary Table S1). Notably, significant enhancements were observed in the classes of stems, branches and leaf stems, and wires (Supplementary Table S2), where DeepLab v3+ with SFM exhibited an increase of 9.74%, 4.49% and 11.47% in *IoU* values. The result indicated that our algorithm exhibited strong generalization ability.

TABLE 1 The accuracy (*acc*) and mean Intersection over Union (*mIoU*) of DeepLab v3+ trained on different datasets.

Training data type	*acc* (%)	*mIoU* (%)
Images taken under diverse weather conditions	96.53	62.12
Images taken under sunny weather conditions	92.00	45.40
Images taken under cloudy weather conditions	92.41	45.27

3.3 Positive effect of depth maps and adaptive K-means clustering on fruit instance segmentation

The depth maps were used to assist in the segmentation of fruit instances in our method. The positive effect of depth maps on pixel clustering of fruit instances was shown in Figure 7, where the combination of depth maps and 2D images enabled the differentiation of adjacent fruits in 2D images based on different depth values (the first and second rows in Figure 7), while similar depth values also prevented oversized fruits from being cut apart (the third row in Figure 7). The reason for the preferable performance of instance segmentation was that depth maps provided more discriminating information.

The quantitative comparison results further proved the positive role of depth maps in segmentation of mature fruit instances (Table 4). When the depth maps were used as input of pixel clustering, the statistical results of fruit instance segmentation were significantly higher where the *Precision*, *Recall* and *F1* were improved by 4.8%, 12.8% and 9.44%, respectively.

Different K values were used to demonstrate the effectiveness of the proposed adaptive K-means clustering algorithm (Table 5). The statistical results showed that the adaptive K-means clustering had obvious advantages over different K-values in fruit instance segmentation. Although the Precision reached 94.11% when K was set to 2, both *Recall* and *F1* score were unsatisfactory (Table 5). When k was set to 4, *Precision*, *Recall* and *F1* reached 87.80%, 72.00% and 79.11% respectively, however, they still fell short of the adaptive K-means clustering method by 4.2%, 11.63% and 8.5%.

3.4 The obstacles perception effectively assists target selection and location

The target fruits were selected or discarded according to the judgment of the categories of the obstacles (Figure 8). The fruit was discarded if the obstacle was a stem or wire that would hinder the

10.3389/fpls.2024.1460060

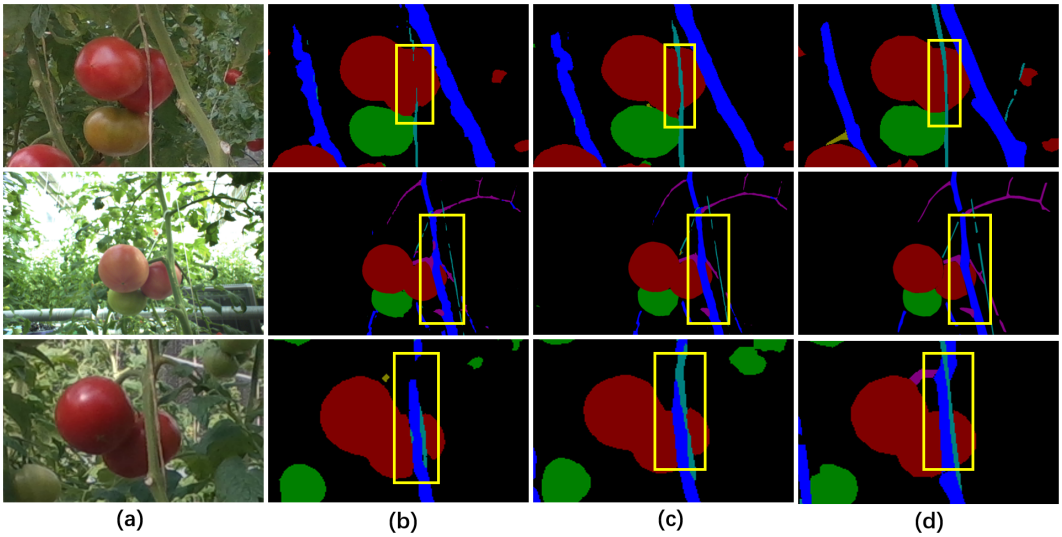


FIGURE 6
Visual comparison of the parsing results of different semantic segmentation methods. **(A)** Original images, **(B)** the segmentation results of DeepLab v3+, **(C)** the segmentation results of DeepLab v3+ with SFM, **(D)** the ground truth labels. Class labels: black indicates background, red indicates mature fruits, green indicates immature fruits, yellow indicates peduncles, dark blue indicates stems, purple indicates branches and petioles, light blue indicates wires and gray indicates cuts.

operation of the robotic arm (the first to third columns of Figure 8), whereas the fruit was retained if the obstacle to the fruit was a leaf (the last column of Figure 8).

To further validate the performance of our method in obstacle perception, we conducted an analysis on all occluded fruits in the validation and testing set (Table 6, Supplementary Material 3). The results showed that 86.67% of the occluded fruits were successfully perceived, with 96.15% of the obstacle categories were correctly recognized. Specifically, for different types of obstacles including stems, wires, branches and petioles, and leaves, the obstacle detection accuracies were 85.71%, 66.67%, 100%, and 100%, respectively (Table 6, Supplementary Material 3). The corresponding false detection rates for these obstacle types were 8.33%, 0%, 0%, and 0%, respectively. For different densities of obstacles, such as single obstacle and multiple obstacles, the obstacle detection accuracies were 84.62% and 100%, while the false detection rates were 4.55% and 75%. This indicated that in low-density obstacles environments, there was an increased likelihood of missing fine obstacles such as wires; however, the accuracy of detecting the obstacles categories was high at 95.45%. Conversely, in high-density obstacles scenarios, the obstacles detection accuracy was high at 100%, while accurately identifying all categories of obstacles proved to be more challenging. Additionally, our method achieved an accuracy rate of 86.67% in distinguishing between pickable (obscured by leaves) and non-pickable (obscured by stems, wires or branches) fruits. Specifically, for different types of obstacles, including stems, wires, branches and petioles, and leaves, the judgment accuracies of fruit pickability were 85.71%, 66.67%, 100%, and 100%, respectively (Table 6, Supplementary Material 3). For different densities of obstacles, such as single obstacle and multiple obstacles, the judgment accuracies of fruit pickability were 84.64% and 100%. This indicated that the accuracy of determining whether the fruit can be picked or not was higher for high-density obstacles compared to low-density obstacles. The above results demonstrated the effectiveness of our approach in recognizing obstructed fruits and preventing potential issues such as mechanical arm entanglement.

TABLE 2 The accuracy (*acc*) and mean Intersection over Union (*mIoU*) of different semantic segmentation methods.

Method	*acc* (%)	*mIoU* (%)
U-net	91.74	50.28
U-net with SFM	92.94	52.81
DeepLab v3+	96.53	62.12
DeepLab v3+ with SFM	96.75	64.31

The proposed method was compared with the fruit selection and localization methods based on Yolo v5 model (Redmon et al., 2016; Tong et al., 2017) to verify its ability of visual obstacle avoidance. Yolo v5 was employed to detect mature and harvestable fruits. By incorporating depth information, the closest and largest fruit was selected as the target for picking, with its center designated as the picking point. Yolo v5 demonstrated high detection performance for mature fruits in our dataset, achieving a *Precision* of 96.4% and a *Recall* of 92.2%. Although the fruit detection efficiency of Yolo v5 was commendable, numerous occluded fruits were observed in its detection results (Figure 9), which would also be selected as picking targets. Figure 10 showed some selection results of Yolov5-based method and our method, where the targets detected by Yolov5-based method (the fruit selection and localization methods based on Yolo v5) were obstructed by stems or wires (Figure 10B), whereas our method can successfully avoid these occluded fruits (Figure 10C).

10.3389/fpls.2024.1460060

TABLE 3 The *IoU* of different semantic segmentation methods for different classes.

Method	Background (%)	Mature fruits (%)	Immature fruits (%)	Stems (%)	Wires (%)	Branches and petioles (%)
U-net	92.28	81.36	60.98	51.59	31.18	36.22
U-net with SFM	92.87	83.10	68.43	54.07	36.25	35.69
DeepLab v3+	96.60	90.01	84.62	72.71	44.02	42.15
DeepLab v3+ with SFM	97.01	90.88	84.82	75.01	49.02	42.56

The *Precision*, *Recall* and *F1* of our method in picking targets selection and location were 88.9%, 84.2% and 86.5% (Table 7), which were 18.9%, 10.5% and 14.7% higher than that of Yolov5-based method. The better performance of our method was largely attributed to the fact that our method can remove the occluded fruits. The mean *PE* between the detected results and the annotations of picking targets were calculated, and the mean *PE* of our method and that of Yolov5-based method were compared (Table 7). The mean *PE* of our method was 9.5% with an average radius of 48.85 pixels, which was obviously superior than that of Yolov5-based method, with a proportion exceeding 10%, partly because our method used the minimum circumcircle of the visible and blind plots of the fruit as the picking target, which reduced the center point offset when part of the fruit was missing due to the occlusion of obstacles.

Additionally, we categorized the images into two groups (one group comprised 26 images featuring dense obstacles, whereas the other included 24 images characterized by sparse obstacles) to evaluate the performance of our method under varying obstacle complexities and densities. Figure 11 illustrated the selection results of both Yolov5-based method and our proposed method across varying complexities and densities of obstacles. The visualization results indicated that both methods effectively selected the correct target in scenarios with low obstacle complexity and density, however, as occlusion intensified, Yolov5-based method faced increased challenges in avoiding occluded fruits. The quantitative comparison results indicated that our proposed method (precision=87.5%) significantly outperformed Yolov5-based method (precision=58.3%) in scenarios with dense obstacles (Table 8), largely due to its effective obstacle detection capabilities. In scenarios with fewer obstacles (Table 9), both our method and Yolov5-based method exhibited commendable fruit selection performance, achieving precision of 90.9% and 90.4%, respectively. Our method demonstrated effective fruit selection and localization performance across scenes with varying obstacle complexity and density, thereby validating its robustness and applicability.

3.5 Computational efficiency analysis

On the images with a resolution of 600×800 pixels, the average time consuming for each phase of our method was 0.018s, 0.8s, and 0.013s, respectively, and the total time consuming was less than 1s. We also investigated the time consumption of the same input on Yolov5-based method. Both methods were implemented based on Python programming language and tested on single NVIDIA GeForce GTX1660. Although our method took about 0.6 seconds longer than Yolov5-based method, the total time was still very short and acceptable for current harvesting robots. This can be understood as sacrificing a reasonable time to perceive obstacles within the fruits.

4 Discussion

Robotic harvesting is a complex task that integrates multidisciplinary expertise, including kinematics, control systems,

10.3389/fpls.2024.1460060

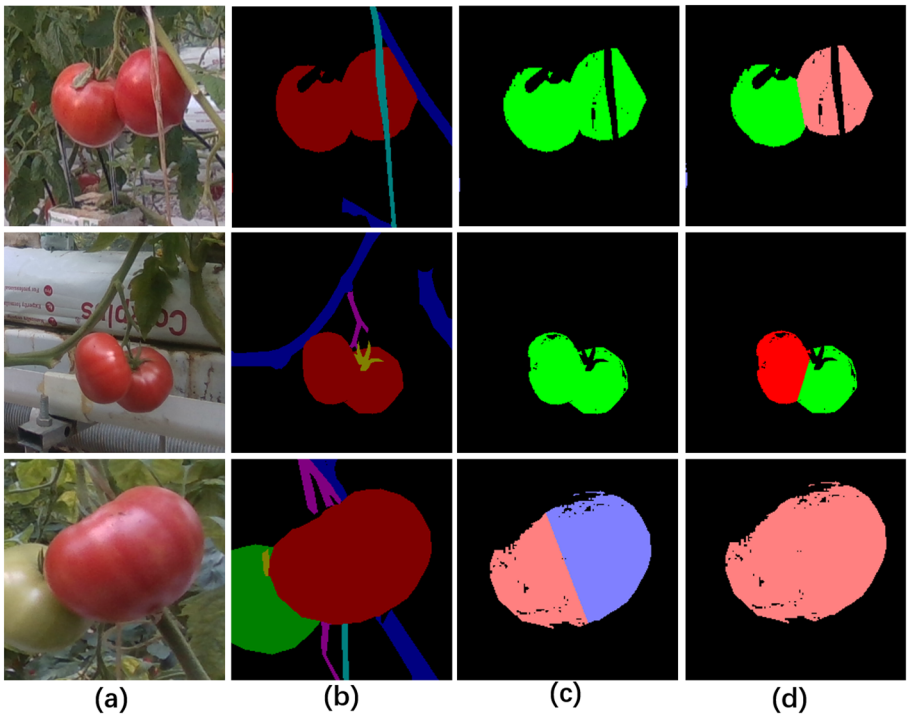


FIGURE 7
Pixel clustering with or without the information of depth maps. **(A)** Original RGB images, **(B)** image semantic segmentation for fruit differentiation, **(C)** visualization of pixel clustering of mature fruits without the information of depth maps, **(D)** visualization of pixel clustering of mature fruits with the information of depth maps.

machine vision, and behavioral science. A pivotal initial step for harvesting robots is to accurately perceive the target fruit and its surrounding environment through advanced machine vision techniques. To enhance the efficiency of the picking process, it is imperative that the vision system of harvesting robots delivers comprehensive information, encompassing fruit recognition, ripeness assessment, 3D coordinate localization, obstacle detection, and estimation of the fruit's 3D posture, and so on. Objective detection algorithms can effectively distinguish ripe and unripe fruits, and by integrating depth image information, the 3D coordinates of the fruits can be accurately located. While, obstacle recognition has always been a challenge and is often overlooked in the vision systems of harvesting robots. Accurate obstacle recognition is crucial for determining harvestable targets and planning efficient harvesting paths.

To address the issue of fruit occlusion during picking, a prioritized selection method for unobstructed fruits was proposed in this study. The proposed method achieved target recognition and localization while also perceiving obstacles, demonstrating its effectiveness in picking target selection and location in intricate scenarios. However, revisions were still necessary to further enhance the capacity of the method. The accuracy of target selection and location largely depends on the performance of the image semantic segmentation and the pixel clustering for instance differentiation. The segmentation of obstacles such as stems basically meet the requirement of the target selection algorithm, but the segmentation accuracy of extremely slender obstacles such

TABLE 4 The performance of instance segmentation of mature fruits with and without the depth maps as input.

Method	*Precision* (%)	*Recall* (%)	*F1* (%)
without depth maps	87.20	70.83	78.17
with depth maps	92.00	83.63	87.61

TABLE 5 Comparison of instance segmentation performance of mature tomato fruits under different K values.

K	*Precision* (%)	*Recall* (%)	*F1* (%)
2	94.11	24.62	30.03
3	87.88	41.43	56.31
4	87.80	72.00	79.11
5	70.45	51.67	59.61
6	72.41	56.21	63.28
7	69.73	79.10	74.12
8	58.06	70.58	63.71
Adaptive	92.00	83.63	87.61

10.3389/fpls.2024.1460060

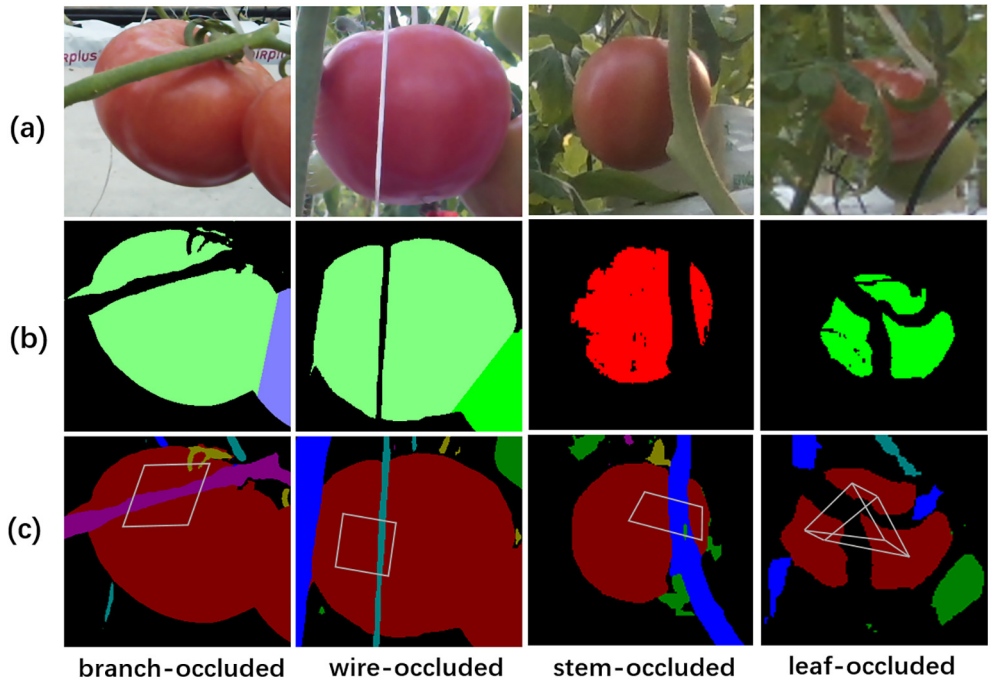


FIGURE 8
The judgment of the categories of obstructions on occluded tomatoes. **(A)** original RGB images, **(B)** the instances of tomato fruits, **(C)** the semantic segmentation maps and blind spots enclosed by gray quadrilaterals.

as wires still need to be further improved. Also, two typical issues may arise during the processes of instance differentiation and obstacle perception. The first issue pertains to the unsuccessful separation of tightly overlapping fruits, resulting in their classification as a single harvesting target. To alleviate this problem, it is imperative to enhance pixel clustering by incorporating more comprehensive judgments of the instances for improved segmentation performance, or consider employing deep learning for fruit instance segmentation. The second issue pertains to the neglect of obstacles located on the edge of the fruits by the perception algorithm. Therefore, further optimization of the obstacle perception algorithm was essential, such as augmenting its ability to detect obstacles surrounding fruit edges, or avoiding this potential risk by providing more fruit pose information to coordinate with fruit grasping actions.

Our proposed method demonstrated an appreciable ability in selecting and locating barrier-free mature fruits within non-structural environments, especially avoiding fruits obscured by stems or wires. This advancement offers a more reliable and practical solution for fruit recognition and localization of harvesting robots. However, there are still certain limitations, such as the absence of detailed information regarding the posture of fruit. Accurate fruit pose estimation is crucial for identifying optimal grasping points and planning the most efficient harvesting path and actions. Achieving 3D pose estimation of tomatoes within complex growth environments, while synchronizing it with the posture of the end-effector, is essential for minimizing damage to surrounding fruits and foliage, significantly improving the success rate of harvesting. In the future, we intend to estimate the posture of fruits through the integration of 2D and 3D visual features in

TABLE 6 The accuracy of obstacle perception for occluded fruits.

	Categories of obstructions						
	Stems	Wires	Branches and petioles	Leaves	Single obstacle	Multiple obstacles	All fruits
The perception accuracy of the presence of obstacles (%)	85.71	66.67	100.00	100.00	84.62	100.00	86.67
The recognition accuracy of obstruction categories (%)	91.67	100.0	100.00	100.00	95.45	25.00	96.15
The judgment accuracy of fruit pickability (%)	85.71	66.67	100.00	100.00	84.64	100.00	86.67

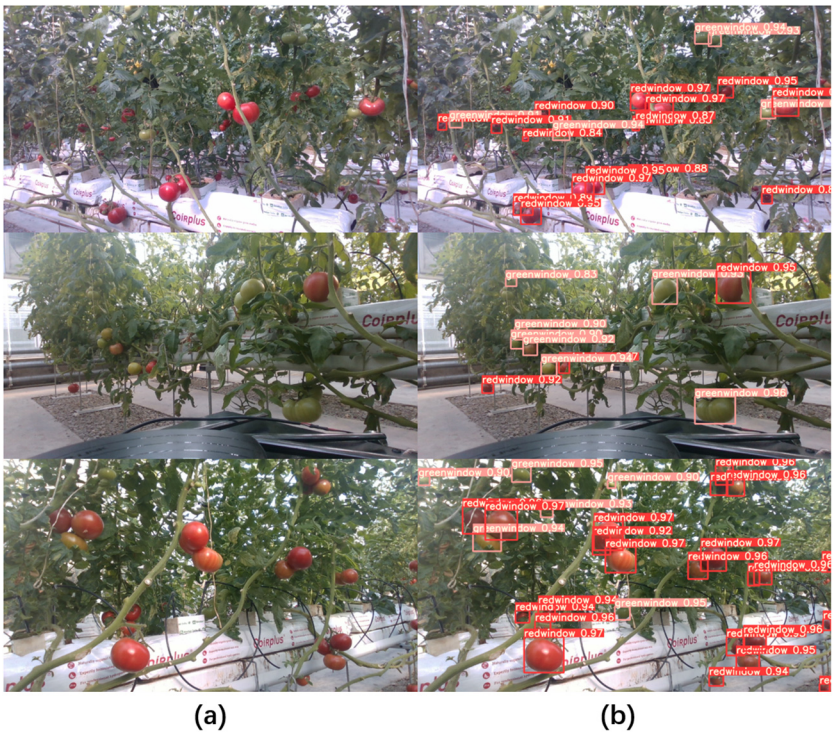


FIGURE 9
Visualization of the detection results of Yolo v5. **(A)** Original RGB images, **(B)** the detection results, where red window represented mature fruit, green window represented immature fruit and the number represented confidence score.

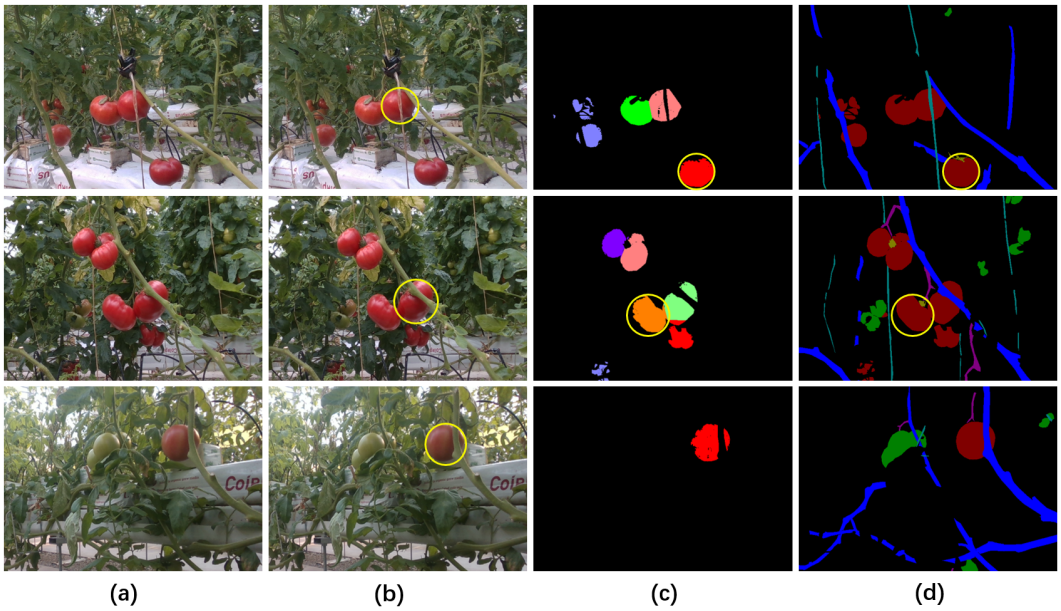


FIGURE 10
Visualization of the selection results of Yolov5-based method and our method. **(A)** Original RGB images, **(B, C)** the outcomes of Yolov5-based method and our method, respectively, **(D)** manual annotations of the targets in the ground truth images. All detected fruits and annotations were circled in yellow.

10.3389/fpls.2024.1460060

TABLE 7 Quantitative evaluation of the performance of two methods in picking-target selection and location.

Method	*Precision* (%)	*Recall* (%)	*F1* (%)	mean *PE* (%)
Yolov5-based method	70.0	73.7	71.8	10.8
Our method	88.9	84.2	86.5	9.5

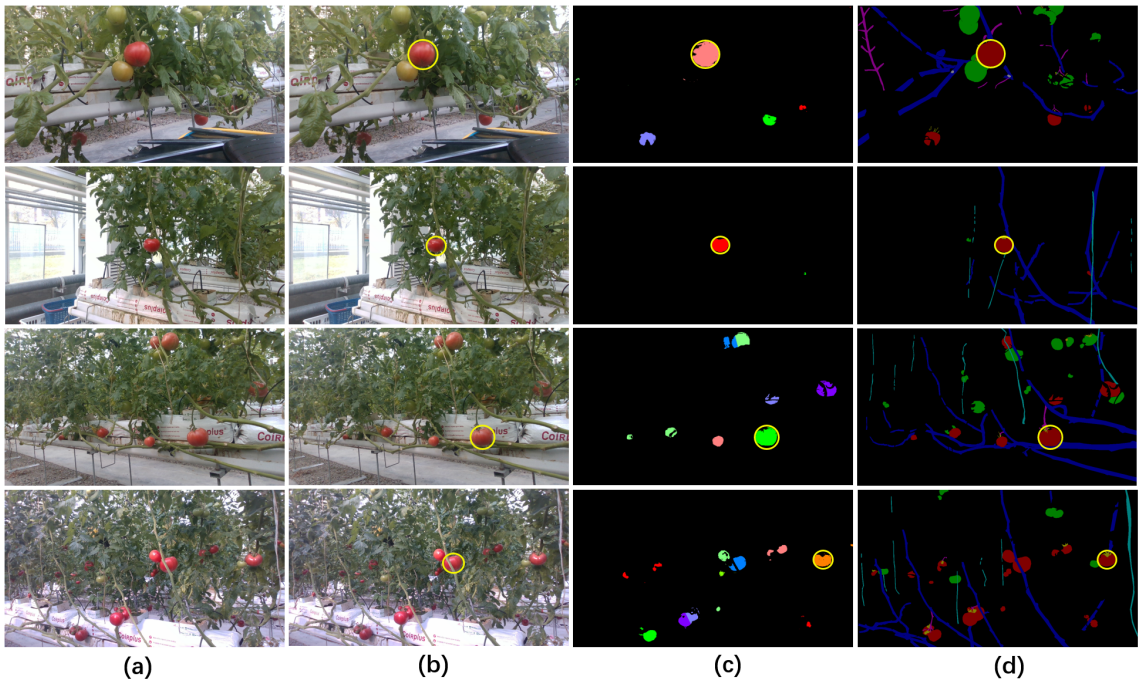


FIGURE 11
Visualization of the selection results of Yolov5-based method and our method under different obstacle complexities and densities. The two upper rows illustrated results under sparse obstacles, while the two lower rows illustrated results under dense obstacles. **(A)** Original RGB images, **(B)** and **(C)** the outcomes of Yolov5-based method and our method, respectively, **(D)** manual annotations of the targets in the ground truth images. All selected fruits and annotations were circled in yellow.

conjunction with keypoint detection, serving the motion path planning of the robotic arm and guiding the grasping posture of the gripper, thereby enabling high-efficiency and higher-quality fruit picking. Additionally, the vision system must be capable of addressing various weather conditions and changes in lighting, as well as adapting in real-time to dynamic changes such as the movement of fruits or foliage, ensuring sustained accuracy in fruit recognition and positioning. Consequently, significant efforts are still required to enhance the robustness and real-time performance of the algorithms.

TABLE 8 The quantitative comparison of our method and Yolov5-based method on images with dense obstacles.

Method	*Precision* (%)	*Recall* (%)	*F1* (%)
Yolov5-based method	58.3	63.6	60.8
Our method	87.5	80.8	84.0

TABLE 9 Comparison results of our method and Yolov5-based method on images with sparse obstacles.

Method	*Precision* (%)	*Recall* (%)	*F1* (%)
Yolov5-based method	90.4	86.3	88.3
Our method	90.9	86.9	88.9

5 Conclusions

Encountering obstacles during the mechanical picking process of tomatoes is inevitable, despite the growing environments are gradually transitioning from unstructured to semi-structured. These obstacles, such as slender stems obstructing the fruits, often entangled with robotic arms and result in picking failures. Our research introduced a novel method for selecting and locating barrier-free fruits based on semantic segmentation and obstacle perception algorithm that offers a solution for selecting safe targets for harvesting robots in practical applications. Each phase of the proposed method offers distinct advantages. Firstly, the easy-to-use spatial-relationship feature module designed for the image semantic

segmentation enabled finer segmentation of slender structural objects by discovering the spatial relevance with features. Secondly, incorporating depth maps as input of adaptive K-means clustering significantly enhanced the performance of fruit instance segmentation. Lastly, our proposed obstacle perception and target selection algorithm can effectively select and discard both non-occluded and occluded fruits. Experiments conducted on our Tomato Dataset and our Tomato Harvesting Robot Platform demonstrated that the proposed spatial-relationship features greatly improve semantic segmentation performance while showcasing our method's ability to exclude obstructed fruits in target selection and location. It is worth noting that our proposed method possesses the potential for extension to a wide range of similar target recognition and localization tasks related to commonly cultivated fruits and vegetables, such as peppers, apples, and kiwis, demonstrating universality and versatility applicability. The model's generalization ability suggests that it can be calibrated and deployed across various agricultural scenarios and scales, thereby providing a robust solution for automated detection systems within the smart farming sector.

The current methodology, while promising, still faces several challenges in the visual task for robotic fruit harvesting. Firstly, there exists a deficiency in the provision of comprehensive posture information regarding fruits and their stems, which is essential for achieving precise robotic manipulation. Additionally, the segmentation accuracy for extremely slender obstacles, such as wires, necessitates further enhancement. The algorithm also faces difficulties in effectively distinguishing and segregating fruits that are heavily overlapping or entangled. Furthermore, to effectively perceive the dynamic changes of fruits, branches, and other obstacles during the harvesting process, there is a pressing need to enhance both the computational efficiency and real-time performance of the algorithm.

In our future research, we intend to construct an integrated end-to-end network informed by the concepts presented in this paper. Multi-task learning networks present a promising approach to achieving this objective, as they effectively balance efficiency while addressing multiple parsing tasks within the visual perception system of fruit harvesting robots. Furthermore, we will refine the backbone of the network to enhance its ability for segmenting extremely slender obstacles, such as wires. Concurrently, we will utilize the powerful feature extraction capabilities of the neural network to enhance the performance of fruit instance segmentation. Additionally, in the new design, we also plan to add the fruit pose estimation through the integration of 2D and 3D visual features in conjunction with keypoint detection. This capability is crucial for guiding robotic grippers to approach and grasp the fruits from the most advantageous angles, thereby minimizing the risk of damage. This integrated network will not only facilitate accurate fruit selection and positioning, but also support motion path planning for robotic arms and guide gripping postures during fruit harvesting. These enhancements will yield more comprehensive, accurate and efficient visual information, while enhancing the harvesting robot's ability to adapt to the unique posture of each fruit and to avoid complex obstacles surrounding it. This is invaluable for further improve the success rate and efficiency of fruit harvesting in unstructured complex environments.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

LZ: Writing – original draft, Validation, Software, Methodology, Conceptualization. AH: Writing – original draft, Investigation. YC: Writing – original draft, Data curation. WZ: Writing – original draft, Resources, Data curation. BZ: Writing – original draft, Validation. XL: Writing – original draft, Resources. QW: Writing – review & editing, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. NR: Writing – review & editing, Funding acquisition.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the National Natural Science Foundation of China (grant number 32201644), Jiangsu Agricultural Science and Technology Innovation Fund (grant number CX(22)3110, CX (22)5009), Natural Science Foundation of Jiangsu Province (grant number BK20200277), Suzhou Science and Technology Development Plan Project (grant number SNG2023005).

Acknowledgments

We sincere gratitude to Dr. Duan and Dr. Chen for their input and guidance on this research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1460060/full#supplementary-material

References

Barth, R., IJsselmuiden, J., Hemming, J., and Van Henten, E. J. (2018). Data synthesis methods for semantic segmentation in agriculture: A Capsicum annuum dataset. *Comput. Electron. Agric.* 144, 284–296. doi: 10.1016/j.compag.2017.12.001

Chen, M., Chen, Z., Luo, L., Tang, Y., Cheng, J., Wei, H., et al. (2024). Dynamic visual servo control methods for continuous operation of a fruit harvesting robot working throughout an orchard. *Comput. Electron. Agric.* 219, 108774. doi: 10.1016/j.compag.2024.108774

Chen, L. C., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arxiv preprint arxiv:1706.05587*. doi: 10.48550/arXiv.1706.05587

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., et al. (2016). The cityscapes dataset for semantic urban scene understanding. *In Proc. IEEE Conf. Comput. Vision Pattern recognition* pp, 3213–3223). doi: 10.1109/CVPR.2016.350

Di Stefano, L., and Bulgarelli, A. (1999). "A simple and efficient connected components labeling algorithm," in *In Proceedings 10th international conference on image analysis and processing*, (Venice, Italy: IEEE), 322–327. doi: 10.1109/ICIAP.1999.797615

Divyanth, L. G., Soni, P., Pareek, C. M., Machavaram, R., Nadimi, M., and Paliwal, J. (2022). Detection of coconut clusters based on occlusion condition using attention-guided faster R-CNN for robotic harvesting. *Foods* 11, p.3903. doi: 10.3390/foods11233903

Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *Int. J. Comput. Vision* 111, 98–136. doi: 10.1007/s11263-014-0733-5

Ghosh, D., Nandakumar, G., Narayanan, K., Honkote, V., and Sharma, S. (2019). "Kinematic constraints based bi-directional RRT (KB-RRT) with parameterized trajectories for robot path planning in cluttered environment," in *In 2019 International Conference on Robotics and Automation (ICRA)*, (Brisbane, Australia: IEEE), 8627–8633.

Gioi, R. G., Jakubowicz, J., Morel, J. M., and Randall, G. (2012). LSD: A line segment detector. *Image Process. On Line* 2, 35–55. doi: 10.5201/ipol

Guo, J., Xiao, X., Miao, J., Tian, B., Zhao, J., and Lan, Y. (2023). Design and experiment of a visual detection system for zanthoxylum-harvesting robot based on improved YOLOv5 model. *Agriculture* 13, 821. doi: 10.3390/agriculture13040821

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 42, 386–397. doi: 10.1109/TPAMI.2018.2844175

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, Las Vegas, America. (IEEE), 770–778.

Jaderberg, M., Simonyan, K., and Zisserman, A. (2015). Spatial Transformer Networks. *arxiv* [preprint]. doi: 10.48550/arXiv.1506.02025

Jia, W., Tian, Y., Luo, R., Zhang, Z., Lian, J., and Zheng, Y. (2020). Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot. *Comput. Electron. Agric.* 172, 105380. doi: 10.1016/j.compag.2020.105380

Jiao, Z., Huang, K., Jia, G., Lei, H., Cai, Y., and Zhong, Z. (2022). An effective litchi detection method based on edge devices in a complex scene. *Biosyst. Eng.* 222, 15–28. doi: 10.1016/j.biosystemseng.2022.07.009

Jun, J., Kim, J., Seol, J., Kim, J., and Son, H. I. (2021). Towards an efficient tomato harvesting robot: 3d perception, manipulation, and end-effector. *IEEE Access* 9, 17631–17640. doi: 10.1109/ACCESS.2021.3052240

Kang, H., and Chen, C. (2020). Fruit detection, segmentation and 3D visualisation of environments in apple orchards. *Comput. Electron. Agric.* 171, 105302. doi: 10.1016/j.compag.2020.105302

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arxiv preprint arxiv:1412.6980*. doi: 10.48550/arXiv.1412.6980

Kurtser, P., and Edan, Y. (2020). Planning the sequence of tasks for harvesting robots. *Robotics Autonomous Syst.* 131, p.103591. doi: 10.1016/j.robot.2020.103591

Kuznetsova, A., Maleva, T., and Soloviev, V. (2020). Using YOLOv3 algorithm with pre-and post-processing for apple detection in fruit-harvesting robot. *Agronomy* 10, , 1016. doi: 10.3390/agronomy10071016

Li, Q., Jia, W., Sun, M., Hou, S., and Zheng, Y. (2021). A novel green apple segmentation algorithm based on ensemble U-Net under complex orchard environment. *Comput. Electron. Agric.* 180, 105900. doi: 10.1016/j.compag.2020.105900

Luo, L., Yin, W., Ning, Z., Wang, J., Wei, H., Chen, W., et al. (2022). In-field pose estimation of grape clusters with combined point cloud segmentation and geometric analysis. *Comput. Electron. Agric.* 200, p.107197. doi: 10.1016/j.compag.2022.107197

MacQueen, J. (1967). "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (University of California Press), Vol. 1. 281–297.

Miao, Z., Yu, X., Li, N., Zhang, Z., He, C., Li, Z., et al. (2023). Efficient tomato harvesting robot based on image processing and deep learning. *Precis. Agric.* 24, 254–287. doi: 10.1007/s11119-022-09944-w

Montoya-Cavero, L. E., de León Torres, R. D., Gómez-Espinosa, A., and Cabello, J. A. E. (2022). Vision systems for harvesting robots: Produce detection and localization. *Comput. Electron. Agric.* 192, 106562. doi: 10.1016/j.compag.2021.106562

Ouyang, C., Li, D., Wang, J., Wang, S., and Han, Y. (2013). "The research of the strawberry disease identification based on image processing and pattern recognition," In *Computer and Computing Technologies in Agriculture VI: 6th IFIP WG 5.14 International Conference, CCTA 2012*, Zhangjiajie, China. 69–77 (Berlin Heidelberg: Springer), October 19-21, 2012, Revised Selected Papers, Part I 6.

Pan, X., Shi, J., Luo, P., Wang, X., and Tang, X. (2018). "Spatial as deep: Spatial cnn for traffic scene understanding," in *Proceedings of the AAAI conference on artificial intelligence*, New Orleans, America. Vol. 32. doi: 10.48550/arXiv.1712.06080

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Brisbane, Australia: IEEE), 779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*. 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Rong, J., Wang, P., Wang, T., Hu, L., and Yuan, T. (2022). Fruit pose recognition and directional orderly gras** strategies for tomato harvesting robots. *Comput. Electron. Agric.* 202, 107430. doi: 10.1016/j.compag.2022.107430

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference*, Munich, Germany, October 5-9, 2015, proceedings, part III 18. (Springer International Publishing), 234–241.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet large scale visual recognition challenge. *Int. J. Comput. Vision* 115, 211–252. doi: 10.1007/s11263-015-0816-y

Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2008). LabelMe: a database and web-based tool for image annotation. *Int. J. Comput. Vision* 77, 157–173. doi: 10.1007/s11263-007-0090-8

Silwal, A., Davidson, J. R., Karkee, M., Mo, C., Zhang, Q., and Lewis, K. (2017). Design, integration, and field evaluation of a robotic apple harvester. *J. Field Robotics* 34, 1140–1159. doi: 10.1002/rob.21715

Song, Z., Zhou, Z., Wang, W., Gao, F., Fu, L., Li, R., et al. (2021). Canopy segmentation and wire reconstruction for kiwifruit robotic harvesting. *Comput. Electron. Agric.* 181, 105933. doi: 10.1016/j.compag.2020.105933

Tang, Y., Qi, S., Zhu, L., Zhuo, X., Zhang, Y., and Meng, F. (2024). Obstacle avoidance motion in mobile robotics. *J. System Simulation* 36, 1–26. doi: 10.16182/j.issn1004731x.joss.23-1297E

Tong, T., Li, G., Liu, X., and Gao, Q. (2017). "Image super-resolution using dense skip connections," in *Proceedings of the IEEE international conference on computer vision*, (Venice, Italy: IEEE), 4799–4807.

Wachs, J. P., Stern, H. I., Burks, T., and Alchanatis, V. (2010). Low and high-level visual feature-based apple detection from multi-modal images. *Precis. Agric.* 11, 717–735. doi: 10.1007/s11119-010-9198-x

Wang, J., Zhang, Z., Luo, L., Wei, H., Wang, W., Chen, M., et al. (2023). DualSeg: Fusing transformer and CNN structure for image segmentation in complex vineyard environment. *Comput. Electron. Agric.* 206, p.107682. doi: 10.1016/j.compag.2023.107682

Williams, H. A., Jones, M. H., Nejati, M., Seabright, M. J., Bell, J., Penhall, N. D., et al. (2019). Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms. *Biosyst. Eng.* 181, 140–156. doi: 10.1016/j.biosystemseng.2019.03.007

Yan, B., Fan, P., Lei, X., Liu, Z., and Yang, F. (2021). A real-time apple targets detection method for picking robot based on improved YOLOv5. *Remote Sens.* 13, 1619. doi: 10.3390/rs13091619

Yang, Y., Han, Y., Li, S., Yang, Y., Zhang, M., and Li, H. (2023). Vision based fruit recognition and positioning technology for harvesting robots. *Comput. Electron. Agric.* 213, 108258. doi: 10.1016/j.compag.2023.108258

Ye, L., Wu, F., Zou, X., and Li, J. (2023). Path planning for mobile robots in unstructured orchard environments: An improved kinematically constrained bi-directional RRT approach. *Comput. Electron. Agric.* 215, p.108453. doi: 10.1016/j.compag.2023.108453

Yousefi, J. (2011). *Image binarization using Otsu thresholding algorithm* Vol. 10 (Ontario, Canada: University of Guelph). doi: 10.13140/RG.2.1.4758.9284

Yu, Y., Zhang, K., Yang, L., and Zhang, D. (2019). Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Comput. Electron. Agric.* 163, 104846. doi: 10.1016/j.compag.2019.06.001

Zhang, X., Karkee, M., Zhang, Q., and Whiting, M. D. (2021). Computer vision-based tree trunk and branch identification and shaking points detection in Dense-Foliage canopy for automated harvesting of apples. *J. Field Robotics* 38, 476–493. doi: 10.1002/rob.21998

Zhang, H., and Peng, Q. (2022). PSO and K-means-based semantic segmentation toward agricultural products. *Future Generation Comput. Syst.* 126, 82–87. doi: 10.1016/j.future.2021.06.059

Zhao, Y., Gong, L., Huang, Y., and Liu, C. (2016). A review of key techniques of vision-based control for harvesting robot. *Comput. Electron. Agric.* 127, 311–323. doi: 10.1016/j.compag.2016.06.022

Zheng, C., Chen, P., Pang, J., Yang, X., Chen, C., Tu, S., et al. (2021). A mango picking vision algorithm on instance segmentation and key point detection from RGB images in an open orchard. *Biosyst. Eng.* 206, 32–54. doi: 10.1016/j.biosystemseng.2021.03.012

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 November 2024
DOI 10.3389/fpls.2024.1456713

Check for updates

OPEN ACCESS

EDITED BY
Syed Tahir Ata-Ul-Karim,
Aarhus University, Denmark

REVIEWED BY
Jiajun Xu,
Michigan State University, United States
Parvathaneni Naga Srinivasu,
Amrita Vishwa Vidyapeetham University, India
Ghulam Mustafa,
Hohai University, China

*CORRESPONDENCE
Xuejun Yue
yuexuejun@scau.edu.cn
Lifei Huang
huanglifei@gdaas.cn

RECEIVED 29 June 2024
ACCEPTED 17 October 2024
PUBLISHED 08 November 2024

CITATION
Ding Z, Zeng F, Li H, Zheng J, Chen J, Chen B, Zhong W, Li X, Wang Z, Huang L and Yue X (2024) Identification of sweetpotato virus disease-infected leaves from field images using deep learning.
Front. Plant Sci. 15:1456713.
doi: 10.3389/fpls.2024.1456713

COPYRIGHT
© 2024 Ding, Zeng, Li, Zheng, Chen, Chen, Zhong, Li, Wang, Huang and Yue. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification of sweetpotato virus disease-infected leaves from field images using deep learning

Ziyu Ding[1], Fanguo Zeng[1], Haifeng Li[1], Jianyu Zheng[1], Junzhi Chen[1], Biao Chen[1], Wenshan Zhong[1], Xuantian Li[1], Zhangying Wang[2], Lifei Huang[2]* and Xuejun Yue[1,3]*

[1]College of Electronic Engineering (College of Artificial Intelligence), South China Agricultural University, Guangzhou, China, [2]Guangdong Provincial Key Laboratory of Crops Genetics and Improvement, Crop Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, Guangdong, China, [3]National Key Laboratory of Green Pesticides, South China Agricultural University, Guangzhou, China

Introduction: Sweetpotato virus disease (SPVD) is widespread and causes significant economic losses. Current diagnostic methods are either costly or labor-intensive, limiting both efficiency and scalability.

Methods: The segmentation algorithm proposed in this study can rapidly and accurately identify SPVD lesions from field-captured photos of sweetpotato leaves. Two custom datasets, DS-1 and DS-2, are utilized, containing meticulously annotated images of sweetpotato leaves affected by SPVD. DS-1 is used for training, validation, and testing the model, while DS-2 is exclusively employed to validate the model's reliability. This study employs a deep learning-based semantic segmentation network, DeepLabV3+, integrated with an Attention Pyramid Fusion (APF) module. The APF module combines a channel attention mechanism with multi-scale feature fusion to enhance the model's performance in disease pixel segmentation. Additionally, a novel data augmentation technique is utilized to improve recognition accuracy in the edge background areas of real large images, addressing issues of poor segmentation precision in these regions. Transfer learning is applied to enhance the model's generalization capabilities.

Results: The experimental results indicate that the model, with 62.57M parameters and 253.92 Giga Floating Point Operations Per Second (GFLOPs), achieves a mean Intersection over Union (mIoU) of 94.63% and a mean accuracy (mAcc) of 96.99% on the DS-1 test set, and an mIoU of 78.59% and an mAcc of 79.47% on the DS-2 dataset.

Discussion: Ablation studies confirm the effectiveness of the proposed data augmentation and APF methods, while comparative experiments demonstrate the model's superiority across various metrics. The proposed method also

exhibits excellent detection results in simulated scenarios. In summary, this study successfully deploys a deep learning framework to segment SPVD lesions from field images of sweetpotato foliage, which will contribute to the rapid and intelligent detection of sweetpotato diseases.

KEYWORDS

sweetpotato, virus disease, deep learning, RGB image, semantic segmentation

1 Introduction

Sweetpotato (*Ipomoea batatas* L.) is one of the top ten food crops in the world and an important source of nutrients for the human body (Lee et al., 2019). Currently, China has the largest sweetpotato planting area of about 6.6 million hm^2 and an output of about 100 million tons, accounting for roughly 80% of the world's total output annually (Gai et al., 2016; Zhang et al., 2019). Sweetpotatoes are propagated through vines and tubers, making them especially vulnerable to viral diseases that can be transmitted from one generation to the next. This mode of transmission often leads to the rapid degradation of new sweetpotato varieties and significant yield losses (McGregor et al., 2009; Clark et al., 2012). The most serious of these viral diseases is sweetpotato virus disease (SPVD), which results from a symbiotic infection of the sweetpotato feathery mottle virus (SPFMV) and the sweetpotato chlorotic stunt virus (SPCSV) (Rännäli et al., 2009; Wanjala et al., 2020).

Since 2012, SPVD has rapidly spread across China, severely affecting key sweetpotato-growing regions. The disease has a devastating impact on sweetpotato production, causing yield losses of 90-100% (Karyeija et al., 1998). For instance, a major SPVD outbreak occurred in Zhanjiang City, Guangdong Province, in early 2015, where 1,257 hectares of sweetpotato fields were severely impacted, with an incidence rate exceeding 50% (Zhang et al., 2019). The prevalence of SPVD in China led to a restructuring of the sweetpotato market, making the production of virus-free seedlings the new standard. However, the lack of efficient virus detection technologies has hindered the supply and breeding of detoxified seedlings. In particular, there is an urgent need for simple and rapid detection methods for use at the grassroots level and for timely testing of seedling quality. Therefore, developing a reliable SPVD detection system is essential for managing the disease and supporting the breeding of virus-free sweetpotato seedlings.

Current research on SPVD has primarily focused on defense mechanisms, particularly virus detection and pathogenesis. The main detection techniques for SPVD include biological, serological, and molecular methods. Among these, enzyme-linked immunosorbent assay (ELISA) and PCR are the most widely used serological methods for SPVD detection (Wanjala et al., 2020; David et al., 2022). However, both ELISA and PCR are labor-intensive, expensive, and time-consuming. As China's sweetpotato industry has expanded, the limitations of these methods have made it increasingly difficult to meet the growing demand for efficient SPVD detection. Thus, there is an urgent need for new, faster, and more efficient detection solutions, such as artificial intelligence (AI) and deep learning technologies, to overcome the shortcomings of manual and conventional SPVD identification methods.

The field of computer vision (CV) has achieved significant success across various applications, particularly in the area of plant disease detection (Srinivasu et al., 2024a; Patil and Manohar, 2022). Recent advancements in convolutional neural networks (CNNs) have revolutionized disease recognition, exemplified by their application in potato disease detection using RGB images, where high precision has been attained (Oppenheim et al., 2019). Further illustrating the efficacy of deep learning, researchers have successfully employed these techniques to identify northern leaf blight in corn, achieving notable accuracy in field conditions (DeChant et al., 2017).

Progress has been marked by a shift from basic classification tasks to more complex challenges, such as pinpointing infected areas and assessing disease severity (Kalaivani et al., 2020). For instance, Zhou et al. utilized an enhanced DeepLabV3+ model for efficient recognition of tea leaf diseases, where the integration of an attention mechanism notably improved the model's sensitivity to subtle lesions (Zhou et al., 2024). Similarly, Yang et al. advanced the detection of rice blast by adopting a U-Net architecture combined with a multi-scale feature fusion strategy, thus improving detection accuracy (Yang et al., 2023). Additionally, Li et al. proposed a Transformer-based framework for the real-time monitoring of wheat stripe rust, showcasing the capacity of deep learning for timely disease detection in agricultural settings (Zhu et al., 2022). Moreover, the evaluation of disease incidence on cucumber leaves under natural conditions has been facilitated by the improved DeepLabV3+ network, further demonstrating the robustness of CV technologies in various crop contexts (Wang et al., 2021; Li et al., 2022).

Despite these advancements, the application of CV technology in sweetpotato disease detection remains underexplored. Current research primarily focuses on other crops such as potatoes, corn, tomatoes, rice, and wheat, leaving a gap in knowledge regarding sweetpotato diseases. Although deep learning techniques have been utilized for counting sweetpotato leaves (Wang et al., 2023), their

impact on disease management and detection in practical scenarios is limited. This is particularly critical given the lack of simple and rapid detection methods at the grassroots level. Therefore, timely and accurate detection of sweetpotato virus diseases is essential for effective disease management and agricultural sustainability.

Motivated by these technical gaps and the need for efficient SPVD detection, this study employs deep learning techniques to identify SPVD lesions on sweetpotato leaves. RGB photos of sweetpotato leaves from the field that showed typical SPVD symptoms at the "branching tuber stage" were gathered, and the DeepLabV3+ deep learning segmentation model was utilized to precisely identify the affected regions. Data enhancement, transfer learning, and a proposed APF module were implemented to enhance the recognition performance. The model's SPVD detection efficacy was evaluated on two datasets (DS-1 and DS-2). The proposed system aims to provide growers with an efficient, easy-to-apply, and affordable solution for SPVD diagnosis, facilitating SPVD management and reducing economic losses caused by the disease.

The main contributions of this study are as follows:

- A data augmentation method is proposed to process edge background areas through a specific strategy, which significantly improves the model's recognition accuracy of edge background areas in real large images.
- The proposed attention pyramid fusion (APF) significantly enhances the model's features at different scales by introducing a channel attention module and fusing multi-scale features, thereby enhancing its performance in disease pixel segmentation.
- Through ablation studies and comparative experiments on a self-made SPVD dataset, the effectiveness of data augmentation and APF methods are verified, and excellent detection results are shown in simulated scenarios.

The rest of this paper is organized as follows: Section 2 introduces the relevant materials and methods. The materials include the obtained datasets and how to process them, while the methodology is a description of the details of the proposed improved deeplabv3+. Section 3 analyzes the experimental results and discusses the impact of the network module. Section 4 discusses and makes suggestions for future research.

2 Materials and methods

2.1 SPVD dataset collection

Due to the lack of publicly available datasets for studying SPVD, this study collected and constructed two distinct RGB datasets from field environments, referred to as DS-1 and DS-2. To ensure the representativeness of DS-1, data were gathered from a 100-acre farmland near Maoming City, Guangdong Province, China, where the local variety "Sweet Fragrant Potato" had been planted for 40-50 days, corresponding to the "Branching Tuber Stage." The top leaves of the plants were green and pointed heart-shaped, while the mature leaves were green and heart-shaped (Figure 1A). Suspected infected plants exhibited typical symptoms of SPVD, such as stunted growth, twisted leaf veins, chlorosis of leaves and petiole leaflets, as well as yellowing of leaf veins. Experts in plant diseases initially classified these symptoms as late-stage SPVD infection, which is the maximum degree of severity. This made it easier to gather photographic evidence and build the study's model.

In January 2022, 216 suspected infected plants were photographed using a handheld SONY DSC-RX100M6 digital camera (Sony Group Corporation, Tokyo, Japan) with close-up and horizontal scanning perspectives under the guidance of a photographer. To obtain detailed information about SPVD detection, images with less obvious symptoms were removed, and 300 high-resolution RGB images in JPEG format (5472×3648 pixels) were selected. Additionally, in September 2023, virus disease symptoms were observed in the sweetpotato resource nursery (45 acres)in Guangzhou, China, and plant leaves with typical SPVD symptoms were selected. Considering the reliance of computer vision technology on visual features, two varieties with appearances similar to "Sweet Fragrant Potato" in DS-1 were chosen, and 12 high-resolution images were captured to

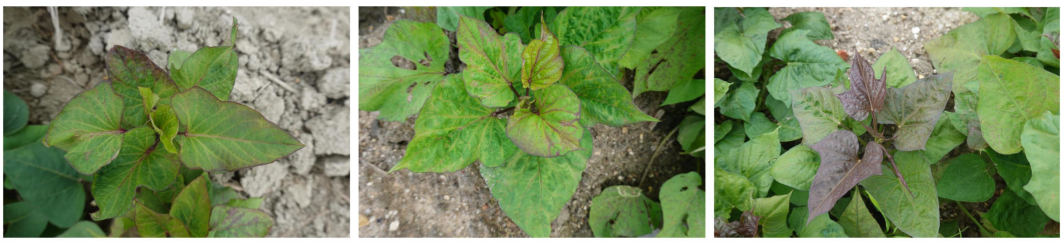


FIGURE 1
Example images in the SPVD dataset. Typical symptoms of SPVD in the late stage of infection can be identified by yellowed, veined, deformed leaves, and dwarfed plants. **(A)**, "Fragrant Pink Potato"; the top leaves were green pointed heart-shaped, and the mature leaves were green heart-shaped. **(B)**, "Guangshu 22-18"; the top leaves were green pointed heart-shaped, and the mature leaves were green heart-shaped. **(C)**, "Guangshu 22-15"; both top and mature leaves were shallow split single notch-shaped, with the top leaves being purple and mature leaves being green.

TABLE 1 SPVD dataset information.

Dataset	Subset	Resolution	Number of images
DS-1	Training	5472×3648	300
	Validation		
	Test-1		
DS-2	Test-2	5472×3648	12

constitute DS-2. Among them, the variety 'Guangshu 22-18' comprised 8 plants, characterized by green pointed heart-shaped top leaves and green heart-shaped mature leaves (Figure 1B); the variety 'Guangshu 22-15' comprised 4 plants. The top leaf color of the plant was purple, and the leaf shape was shallow split single notch, while the mature leaf color was green, and the leaf shape was shallow split single notch (Figure 1C). The variety "Guangshu 22-18" was highly similar to "Sweet Fragrant Potato" in terms of leaf color and shape, while the variety "Guangshu 22-15" had slight differences in leaf shape and completely different top leaf color. These sweetpotatoes had been planted for 50-60 days ("Branching tuber stage "), and experts preliminarily identified them as being in the late stage of SPVD infection. In summary, the DS-1 dataset is used to train, validate, and test the performance of the model. The DS-2 dataset is specifically used to test the generalization ability of the model. Statistical details are provided in Table 1. In Section 2.3, the DS-1 and DS-2 datasets were further preprocessed to ensure suitability for specific task requirements. Samples collected from both locations were promptly placed in ice chests, frozen rapidly in liquid nitrogen, and stored in a -80°C freezer for molecular biology identification purposes.

2.2 Materials and methods for virus identification

Materials: Plant RNA Kit, Omega Bio-Tek, USA; EasyScript OneStep gDNA Removal and cDNA Synthesis SuperMix Kit, TransGen Biotech Co., Ltd., China; 2×TaqMaster Mix (With Dye), APExBIO Technology LLC, USA; Primers synthesized by Sangon Biotech (Shanghai) Co., Ltd., China; Biometra TAdvanced PCR Instrument, Biometra GmbH, Germany; DYY-6C Electrophoresis Apparatus, Beijing Liuyi Biotechnology Co., Ltd., China; Tanon 4100 Gel Imaging System, Shanghai Tianeng Technology Co., Ltd., China.

The collected sweetpotato plant leaves were ground into powder in liquid nitrogen. The total RNA was extracted using a Total RNA Extraction Kit, followed by cDNA synthesis using a reverse transcription kit. The reverse transcription reaction system was 20 μL. For the reverse transcription, 1 ug of the RNA template was denatured at 65°C for 5 minutes and then placed in an ice bath for 2 minutes. Next, 1μL of Anchored Oligo $(dT)^{18}$ Primer (0.5 μg/μL), 1μL of EasyScript® RT/RI Enzyme Mix, and 10 μL of 2×ES Reaction Mix were added and topped up to 20 μL with RNase-Free H_2O. The process was maintained at 42°C for 15 minutes and then 85°C for 5 seconds to complete the reverse transcription reaction. RT-PCR amplification was performed using RNA virus-specific primers (primer sequences listed in Table 2) (Li et al., 2012; Pan et al., 2013). The RT-PCR reaction system (25 μL) included 2 μL of synthesized cDNA, 12.5 μL of 2×Taq Master Mix (With Dye), 1μL each of upstream and downstream primers (10 μmol/L), and 8.5 μL of ddH_2O. The amplification conditions were as follows: 94°C for 2 minutes; 30 cycles of 94°C for 30 seconds, annealing at 45°C (temperature specified in Table 2) for 45 seconds, and extension at 68°C for 1 minute; followed by a final extension at 68°C for 10 minutes. All assays included a negative control using water as a template. Next, the 5 μL amplified products were detected by 1% agarose gel electrophoresis and photographed by gel imager (Xinliang et al., 2022).

2.3 SPVD dataset process

Due to the complexity of processing high-quality initial images and the limited GPU memory allocation, the original photos were clipped to 512 × 512 pixels (Figures 2A, B). The image background, such as the soil (Figure 2C) and the residual parts (Figure 2E) were discarded. Finally, 5,723 small-sized images containing only the diseased leaves portion (Figure 2D) were selected. Besides, images in DS-2 underwent the same

TABLE 2 Primers used in RT-PCR for the detection of sweetpotato viruses[a].

Primer	Sequence[b](5′-3′)	Product size/bp	Annealing Temperature/°C	Target virus
SPF-F1	GGATTAYGGTGTTGACGACACA	589	60	SPFMV
SPC-F1	GTGAGAAAYCTATGCGCTCTGTT	836		SPVC
SPG-F5	GTATGAAGACTCTCTGACAAATTTTG	1191		SPVG
SP2-F10	CGTACATTGAAAAGAGAAACAGGATA	369		SPV2
SPFCG-R2	TCGGGACTGAARGAYACGAATTTAA	–		SPVG,SPVC, SPFMV, SPV2
CSV1	AGTGGTGAYGTAATAGTCGGTGG	365	61.06	SPCSV
CSV2	GCTAACGATTCACADACAGACTTCA	365	59.24	

[a]F represents forward primer; R represents reverse primer.
[b]Y= C or T; R=A or G.

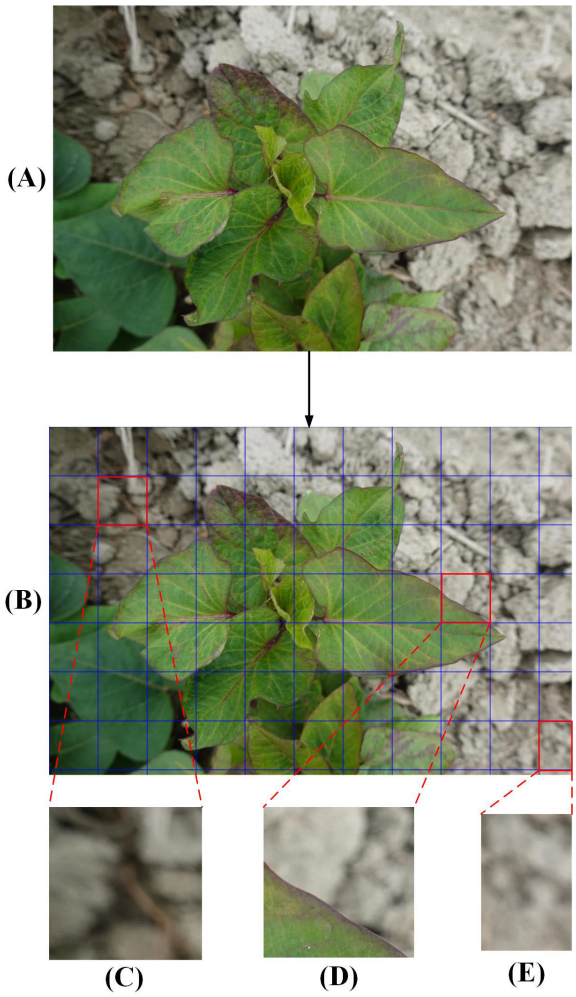


FIGURE 2
The process of dividing the original images into smaller images. **(A)**, A full-sized raw image in DS-1 before segmentation. **(B)**, Small-size image areas to be segmented. **(C)**, Original image background. **(D)**, Diseased leaf image. **(E)**, Cut residual image.

processing, resulting in a total of 710 images sized at 512×512 pixels, which were not used in the model training process rather were only used to verify the model generalization ability and reliability.

After the original image was segmented, the DS-1 and DS-2 images were labeled as infected or background. The leaf-diseased area was marked with polygons using the open-source computer vision annotation tool (CVAT). The generated images were saved in PNG format with ground truth annotations (Figure 3A). The ground truth labels were then divided into training, verification, and test sets in the ratio 8.1.1 with 4578, 572, and 573 images, respectively. To avoid over-fitting, the image augmentation method (Zou et al., 2021) was adopted to enhance the image diversity. Briefly, the infected area was cut out using the label as the foreground and overlaid on the background image to create a composite image (Figure 3B). To obtain a richer background, 165 background, and 27 foreground images were selected, generating 4455 composite images. After supplementing the training set with the obtained composite images, 9033 images were obtained. In addition, the process is shown in Figure 4.

2.4 Deep learning algorithm

The Efficient Channel Attention (ECA) module is a local cross-channel interaction strategy without dimensionality reduction (Wang et al., 2020), which can be effectively implemented by one-dimensional (1D) convolution, and is an effective channel attention learning method (Figure 5A). It is a lightweight general-purpose module that can be easily embedded into any CNN framework to achieve end-to-end training and significantly improve network performance, and its implementation process is shown in (Figure 5B). This method first uses global average pooling (GAP) to obtain X_0 for each feature channel of the input feature map X_{input}, then uses 1D convolution to capture local cross-channel interaction information to obtain X_1, and then uses the Sigmoid function to generate channel weights X_2, to get normalized weights between 0 and 1. Finally, the original feature image X_{input} with a matrix size of H×W×C is multiplied by the weight generated by the Sigmoid function to obtain a new feature image X_{output}.

ResNet101 is a deep convolutional neural network architecture that introduces skip connections, allowing gradients to flow directly from later layers back to earlier layers during backpropagation (Zhang, 2022). The architecture typically serves as a backbone network for feature extraction and is composed of four stages,

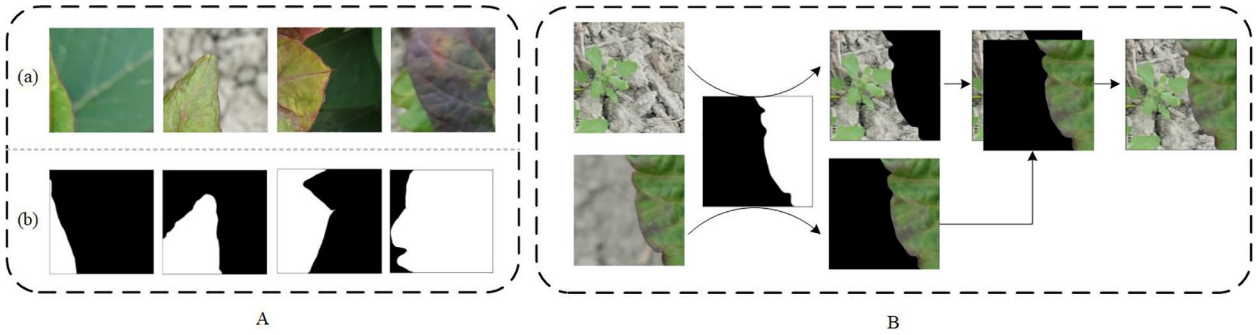


FIGURE 3
Data annotation and augmentation. **(A)**, Small-sized images and annotated images. **(B)**, The flowchart for making a composite image.

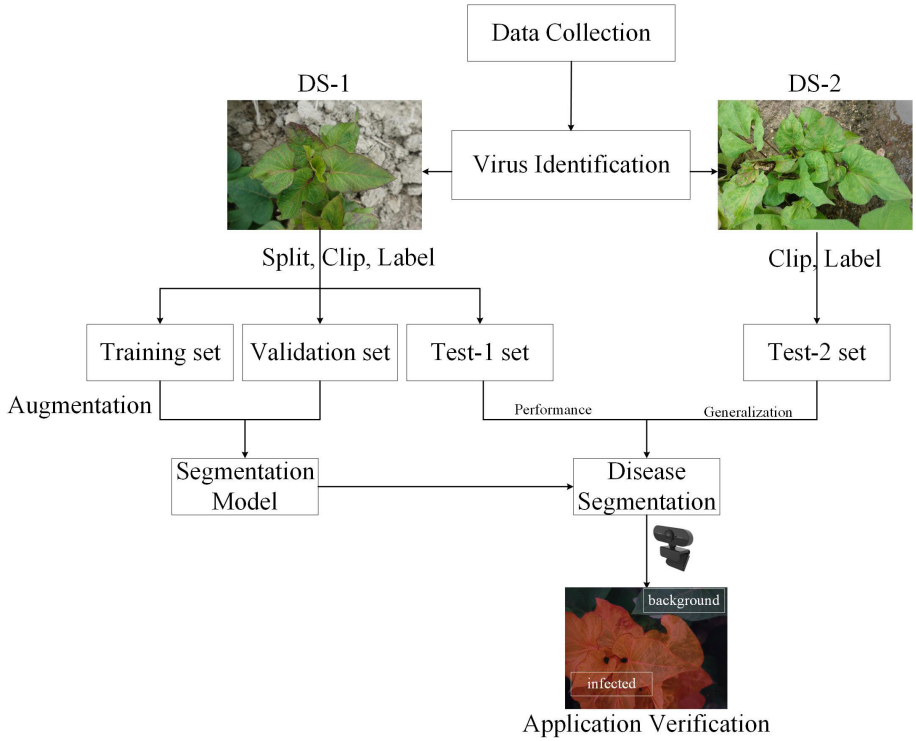


FIGURE 4
Overall flowchart of the methodology.

each containing multiple residual blocks. These stages serve as input features for downstream tasks, denoted as Block1, Block2, Block3, and Block4, with corresponding strides of (1, 2, 1, 1). A stride of 1 indicates that the output feature map size remains the same as the input feature map, without downsampling; a stride of 2 indicates that the output feature map size is half of the input feature map, indicating downsampling. After the second stage, the feature map size remains unchanged. Such a design helps maintain the resolution of the feature maps, which is particularly important for accurate segmentation in semantic segmentation tasks. Additionally, the activation function used is ReLU.

In the process of extracting image features using deep learning, the resolution of the image gradually decreases due to the continuous application of deep convolution operations, resulting in lower resolution deep features. This phenomenon is particularly detrimental to small objects within the image, leading to recognition errors. To address this issue, combining features from different levels during network training can significantly enhance the accuracy of multi-scale detection. Feature Pyramid Network (Lin et al., 2017) is an effective feature fusion method that combines feature maps from different layers to obtain feature representations that reflect semantic information at various scales. Different from FPN, the proposed attention pyramid fusion (APF) introduces a channel attention module on features of different scales, and then fuses the enhanced feature representations of different scales, further enhancing the representation ability of feature maps at different scales (Figure 5B, where (C, H, W) represents the feature map (channels, height, width). Specifically, Feature maps generated by Block1, Block2, and Block3 of the DeepLab v3+ backbone network ResNet101 are fused. These feature maps have sizes of 1/4, 1/8, and 1/8 of the input image, with initial channel numbers of 256, 512, and 1024, respectively. In APF, a 1×1 convolution is first applied to the feature maps of Block1, Block2, and Block3 for dimensionality reduction, optimizing computational efficiency and feature representation. Specifically, the number of channels in Block1's feature map is reduced from 256 to 48, Block2's feature map from 512 to 256, and Block3's feature map from 1024 to 256. To unify the sizes of the feature maps, the feature maps of Block2 and Block3 are upsampled, increasing their sizes from 1/8 to 1/4 to match the size of Block1's feature map. Finally, these three processed feature maps are combined to obtain the fused feature map. This fused feature map not only contains feature information from three different levels but also possesses richer semantic and spatial information due to the introduction of the channel attention module, significantly enhancing the performance of the DeepLab v3+ network in image segmentation tasks.

The DeepLabV3+ used in this article is one of the most popular semantic segmentation networks (Chen et al., 2018), which consists of two parts: encoder and decoder, and uses depthwise separable convolution (Chollet, 2017), which is more efficient. The encoder module reduces the spatial resolution and captures semantic information, and the decoder module recovers the spatial information and leads to a cleaner segmentation. The flow chart of the algorithm is shown in Figure 5, where (C, H, W) represents the feature map (channels, height, width). First, input the 512×512 pixels picture containing part of the diseased leaves into the

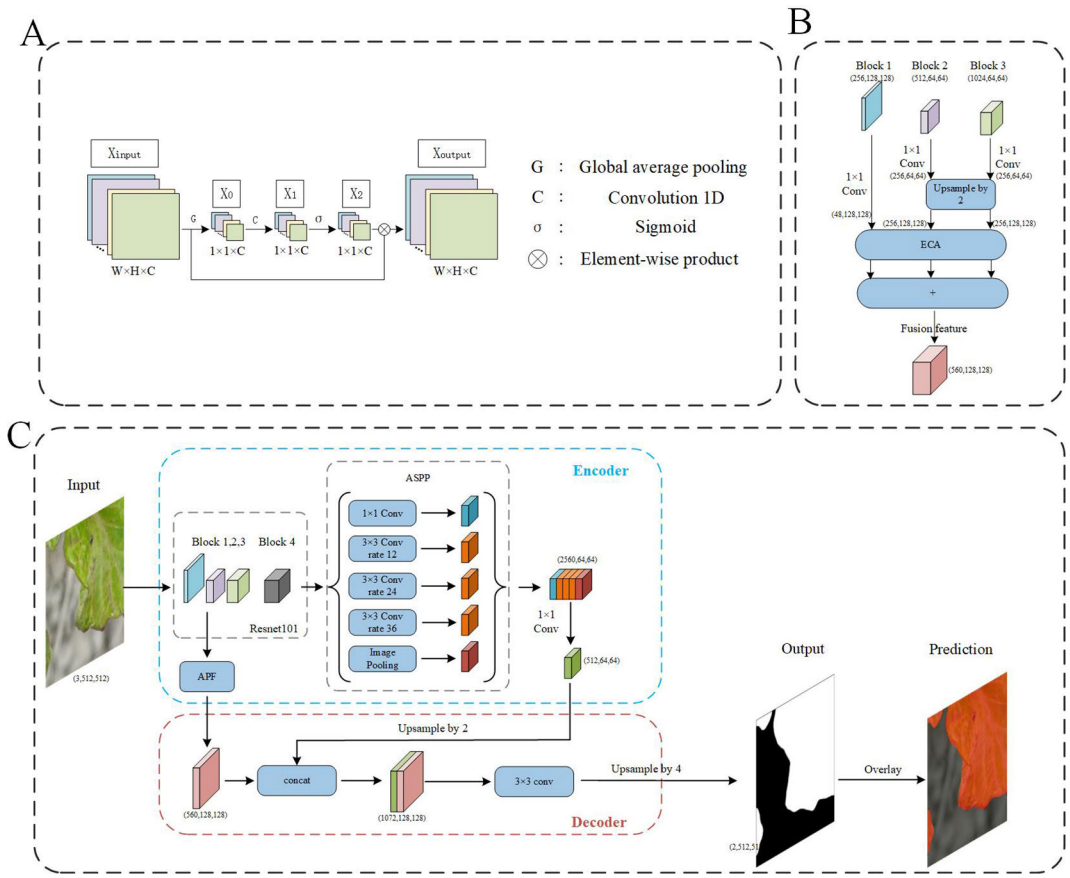


FIGURE 5
The flowchart of model. **(A)** The flowchart of ECA module. **(B)** The flowchart of APF module, **(C)** The flowchart of the segmentation network from small-sized images to prediction outputs.

backbone network ResNet101 in the encoder, which can be used to extract the diseased features, in which the representation information is extracted in the Block 1,2,3, and the semantic information is extracted in the Block4 (1024 channels and 1/8 input size). The Atrous Spatial Pyramid Pooling (ASPP) (Chen et al., 2017) mechanism with different atrous rates of atrous convolution is used in DeepLabV3+. Since the diseased area in those small images is generally larger, a higher hole rate (rate = 12, 24, 36) is used compared to the original network (rate = 6, 12, 18). At the same time, ASPP also contains a 1×1 dilated convolution and an image pooling layer. After splicing the multi-scale feature maps, use 1×1 atrous convolution to fuse information and reduce dimensions. The processed feature map is then input to the decoder module. The output of the encoder module is upsampled by a factor of 2 and then fused with the multi-scale enhanced features that have passed through the APF module. Afterwards, a 3×3 convolutional layer is used to extract features, followed by another simple bilinear upsampling by a factor of 4. Finally, classify each pixel to get the binarized prediction result. According to the binarized prediction result, the diseased areas in the small-scale image can be marked in red, which can clearly show the diseased leaves. Moreover, to speed up the convergence process and improve the generalization ability of the model, a transfer learning strategy was adopted for the backbone ResNet101 parameters. Specifically, the backbone parameters trained on public datasets were used as the initial parameters of the network.

2.5 Model verification process

Model detection was verified under simulated conditions using a 480p Guke camera with a resolution of 640×480 pixels per frame. These frames were subsequently resized to 512×512 pixels and input into the model for processing to obtain 512×512 pixels binarized prediction result. According to the binarized prediction result, the diseased areas in the small-scale image were marked in red. Finally, the 512×512 pixels binarized prediction images were resized to 640×480 pixels and displayed in a small window.

2.6 Training procedure and performance evaluation

The system hardware consists of an Intel Core i9-10900X CPU, 32 GB of RAM, and an NVIDIA GeForce RTX 3090 GPU with 24 GB of VRAM. The software environment comprises Python 3.9, PyTorch 1.12.0, and a WSL system running on Windows 11. The optimizer employs the SGD algorithm with a weight decay factor of

0.0005 and a momentum factor of 0.9. The initial learning rate is set to 0.01 and varies according to the poly strategy. The batch size is set at 4, with a maximum of 20,000 iterations. The loss function is based on cross-entropy. The paper (Srinivasu et al., 2024b) is referenced, and some important parameters are presented in Table 3.

To demonstrate the enhanced capabilities of the proposed SPVD recognition model, a comprehensive comparative analysis of pioneering and state-of-the-art deep learning architectures is conducted. This includes: 1. Fully Convolutional Networks (FCN): Introduced by Long et al (Long et al., 2015), FCN revolutionizes semantic segmentation by enabling end-to-end learning without fully connected layers, making segmentation more efficient and accurate. 2. Pyramid Scene Parsing Network (PSPNet): Developed by Zhao et al (Zhao et al., 2017), PSPNet introduces the pyramid pooling module, effectively aggregating context information at different scales. This is crucial for distinguishing fine-grained categories in SPVD recognition. 3. Segmenter Series: Specifically, Segmenter ViT-B (Vision Transformer Base version) and Segmenter ViT-S (Vision Transformer Small version). These models showcase the potential of Transformer-based approaches in semantic segmentation by modeling long-range dependencies through self-attention mechanisms. 4. SETR ViT-L: The Segmentation Transformer with a Large Vision Transformer variant demonstrates the power of large-scale Vision Transformers in the field of semantic segmentation, particularly adept at handling wide spatial variations. 5. SegFormer MiT-B4: A combination of lightweight Transformer blocks and multi-scale feature representations, SegFormer efficiently integrates these components to maintain high precision while keeping computational complexity low.

These benchmark algorithms are sourced from the mmsegmentation library, an open-source semantic segmentation toolbox based on PyTorch. This library provides implementations of various semantic segmentation models and supports multiple mainstream segmentation frameworks. The library is highly extensible and includes many advanced technologies built-in, which facilitate the acceleration of the development and training processes.

TABLE 3 Training parameters used in proposed model.

Training parameters	Values
Input Size	(512, 512)
Learning policy	Poly
Optimizer Type	SGD
Learning Rate	0.01
Momentum	0.9
Weight Decay	0.0005
Loss function	CrossEntropy
Batch size	4
Atrous rates	[12, 24, 36]

The evaluation indicators cover the number of parameters (Params, M), number of floating-point operations (FLOPs, GFLOPs), mean intersection over union (mIoU, %), mean accuracy (mAcc, %), and the segmentation time (ST, ms). Among them, Params and FLOPs are statistically analyzed. ST is calculated by dividing the sum of processing time for all images in the test set by their total number. mIoU is the commonly used evaluation index in semantic segmentation methods, revealing the overlap degree between the predicted and actual areas. The mAcc was the ratio of the number of correctly predicted pixels to the total number of pixels, indicating the accuracy of the prediction results.

mIoU and mAcc are calculated using formulas (1) and (2) below, where k+1 represents the number of categories, which is set to 2 in this case, implying that pixels in each image are classified into two categories: SPVD-infected areas and the background. p_{ij} represents the number of pixels of category i predicted to be category j, p_{ii} indicates the number of pixels of category i predicted to be category i, and p_{ji} indicates the number of pixels of category j predicted to be class i.

$$mIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \quad (1)$$

$$mAcc = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}pij} \quad (2)$$

3 Results

3.1 RT-PCR detection results

Only a distinct band of approximately 600 bp within the 500-750 bp range was obtained based on the multiple detections of the four viruses, including SPVG, SPVC, SPFMV, and SPV2 using RT-PCR (Figure 6A). According to (Li et al., 2012), only 589 bp of SPFMV was amplified in four samples. Also, the presence of SPVG, SPVC, SPV2 was not detected. With the additional testing for the SPCSV virus, only a 365bp band was amplified (Figure 6B), confirming that the virus disease infecting the sweetpotato in the field was SPVD. Therefore, the diseased sweetpotato samples contained only two viruses, SPFMV and SPCSV. This, combined with the dwarfing and yellowing of sweetpotato plants in the field, indicates that the RGB data used for modeling analysis in this study is the sweetpotato data set infected with SPVD.

3.2 Ablation studies

The results from the ablation study revealed that the baseline model achieved a mIoU of 85.14% (Table 4). Compared to the baseline, the mIoU of baseline+DA was increased by 3.04% and mAcc by 0.45%. The mIoU of baseline+APF was 89.64%, and mAcc was 93.49%, which was better than the former. After incorporating

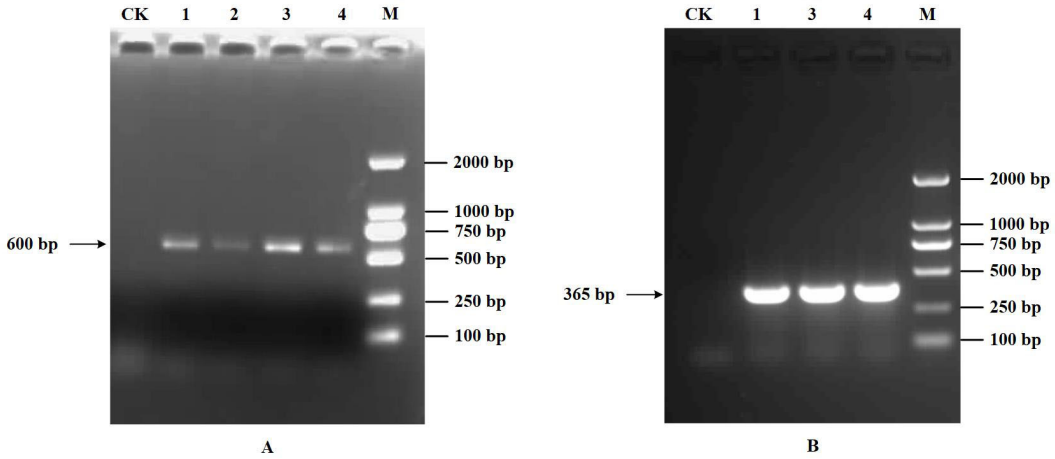


FIGURE 6
Detection results of RT-PCR. **(A)**, Detection results of RT-PCR. 1-4: SPFMV. **(B)**, Identification result of SPCSV. 1,3,4: SPCSV. M: DL2000 DNA maker, CK: healthy plant control.

APF into the baseline model with data augmentation (DA), the mIoU decreased by 0.12%. The performance improvement of baseline+TL was the most obvious, with mIoU and mAcc of 92.05 and 94.52%, respectively. Upon adding DA to the baseline with TL, mIoU and mAcc increased to 93.92% and 96.38%, respectively. After adding APF to the baseline model with TL, mIoU and mAcc increased by 2.58% and 2.47%, respectively.

By comprehensively utilizing the APF and TL methods, the highest mIoU and mAcc scores achieved were 94.63 and 96.99%, respectively (a portion in Figure 7A). However, when the model was used to process pure background area, unsatisfactory performance was observed (Figure 7B). Specifically, the segmentation performance on the infected leaves was notably impressive, showcasing high accuracy. However, the recognition performance on the surrounding background was deficient, appearing as square red areas (Figure 7C). Although the mIoU and mAcc were slightly lower (0.2 and 0.27%, respectively) than the former after adding the DA, the overall effect was good (Figure 7D).

3.3 Comparison with other segmentation methods

The mIoU and mAcc scores of the FCN, PSPNet and proposed network under the same conditions were higher than 85 and 90%, respectively, indicating that they can successfully complete the recognition challenge (Table 5). Segmenter-S is lighter in terms of parameter count and computational complexity (25.98 MB, 37.36 GFLOPs), making it suitable for deployment in resource-constrained environments. Although its performance (87.28% mIoU, 91.12% mAcc) is slightly lower than that of Segmenter-B (88.11% mIoU, 92.94% mAcc), it still demonstrates effective segmentation results, proving that lightweight models can achieve high segmentation accuracy while maintaining efficiency. SegFormer stands out with its relatively low parameter count (64.52 MB) and exceptional segmentation performance (91.27% mIoU, 94.45% mAcc). This is largely due to its innovative Transformer encoder design, which effectively leverages the self-attention mechanism to capture long-

TABLE 4 Performance comparison of different improvements on the network using the DS-1 test set[a].

Methods	Baseline	DA	APF	TL	mIoU(%)	mAcc(%)
Baseline	√				85.14	93.39
Baseline + DA	√	√			88.18	93.84
Baseline + APF	√		√		89.64	93.49
Baseline + TL	√			√	92.05	94.52
Baseline + DA + APF	√	√	√		88.06	94.04
Baseline + DA + TL	√	√		√	93.92	96.38
Baseline + APF + TL	√		√	√	94.63	96.99
Baseline + DA + APF + TL **(Proposed)**	√	√	√	√	94.43	96.72

[a]Where √ indicates the strategy was used, blank means the strategy was not used. Baseline = DeepLabV3+ network with ResNet101 backbone, DA, data augmentation; APF, APF module; and TL, transfer learning; MIoU, mean intersection over union; MAcc, mean accuracy.

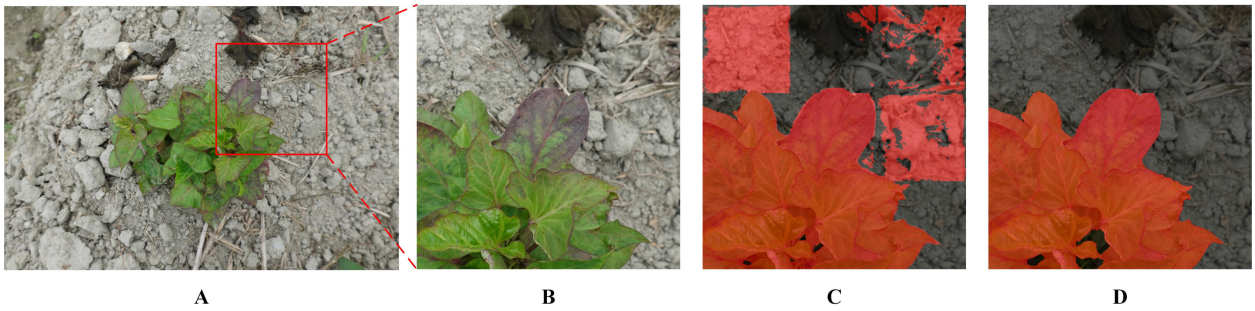


FIGURE 7
SPVD identification performance with and without DA. **(A)** A raw image. **(B)** The selected area on the raw image. **(C)** Recognition result based on baseline + APF + TL method. **(D)** Recognition result based on baseline + DA + APF + TL method (Proposed).

range dependencies in images while reducing computational complexity and memory consumption. SegFormer excels particularly in capturing detailed segmentation, accurately distinguishing between different object classes, as shown in Figure 8, where its segmentation results are often closer to ground truth values. While SETR (Segmentation Transformer) has the largest parameter count (309.33 MB) and relatively high computational complexity (367.21 GFLOPs), its segmentation performance (88.69% mIoU, 92.20% mAcc) did not reach the expected optimal levels. This may be due to the fact that the Transformer structure requires more data to train to fully leverage its advantages, and the training process might be more sensitive and unstable, as evidenced by the noticeable oscillations in the training curves. Although the proposed method shows slightly higher ST compared to other methods, it achieves the highest precision, exhibits stable training performance, and overall delivers the best performance. The training loss and accuracy curves of various semantic segmentation models are shown in Figure 9.

The comparison of segmentation details of different networks is shown in Figure 8, including the high-resolution original image (Figure 8A), its 512×512 pixel partial image (Figure 8B), and the corresponding ground truth labeled region (Figure 8C). The recognition results of FCN, PSPNet and our proposed method are shown in Figures 8D–F, respectively. The segmentation effects of the state-of-art models Segmenter ViT-B, Segmenter ViT-S, SegFormer MiT-B4 and SETR ViT-L are shown in Figures 8G-J, respectively. Among them, Segmenter ViT-B shows a certain accuracy in overall segmentation, but is slightly insufficient in processing details; Segmenter ViT-S shows a certain competitiveness in segmentation results while maintaining a relatively fast inference speed, but its segmentation details are slightly rough in complex backgrounds. SegFormer MiT-B4 has a very good segmentation effect. Its segmentation results are not only highly accurate, but also have smooth edges and well-processed details, which is close to our proposed method. Although SETR ViT-L has a large number of model parameters and high computational complexity, its segmentation effect does not fully meet expectations. The segmentation results of our proposed method are the best and closest to the ground truth.

3.4 Model generalization verification

The model's generalization is tested on DS-2, with the results summarized in Table 6. The mIoU and mAcc values are recorded as 78.59% and 79.47%, respectively, representing a decrease of 16.04% and 17.52%, respectively, compared to the DS-1 test set's metrics of 94.63% and 96.99%. The segmentation results are shown in Figure 10. The model exhibits excellent recognition capability for disease features such as deformity and chlorosis on green top leaves and mature leaves, which are present in the DS-1 training set (Figure 10, first row). However, due to the lack of visual features of SPVD on purple top leaves in the DS-1 training set, the identification results for purple top leaves in DS-2 are poorer. (Figure 10, second row).

TABLE 5 Comparison results of different segmentation methods under enhanced data[a].

Segmentation methods	Params(M)	FLOPs(GFLOPs)	MIoU(%)	MAcc (%)	ST(ms)
FCN	68.48	275.37	86.68	90.16	33.15
PSPNet	67.95	256.13	89.90	93.84	52
Segmenter-B	102.38	126.11	88.11	92.94	33.96
Segmenter-S	**25.98**	**37.36**	87.28	91.12	**26.21**
SegFormer	64.52	350.12	91.27	94.45	75.86
SETR	309.33	367.21	88.69	92.20	81.06
Proposed	62.57	253.92	**94.43**	**96.72**	50.61

[a]Params, parameters; FLOPs, floating-point operations; MIoU, mean intersection over union; MAcc, mean accuracy; ST, segmentation time. Numbers in bold indicate the best performance.

10.3389/fpls.2024.1456713

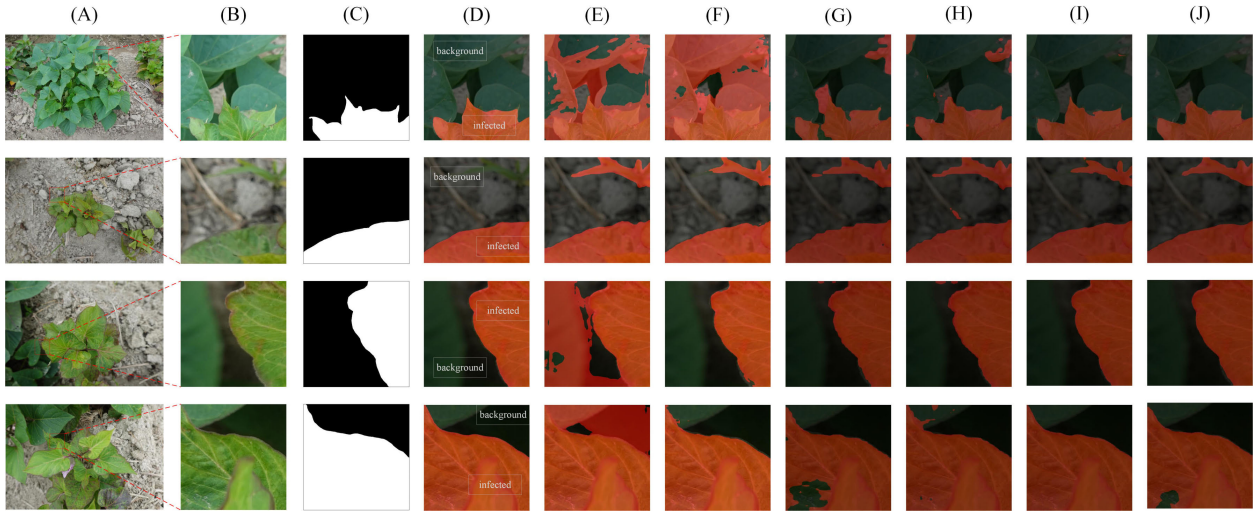


FIGURE 8
Segmentation results using different algorithms on DS-1. **(A)** Original images in DS-1 **(B)** The selected area **(C)** Ground truth **(D)** proposed algorithm **(E)** FCN **(F)** PSPNet **(G)** Segmenter ViT-B **(H)** Segmenter ViT-S **(I)** SegFormer MiT-B4 **(J)** SETR ViT-L.

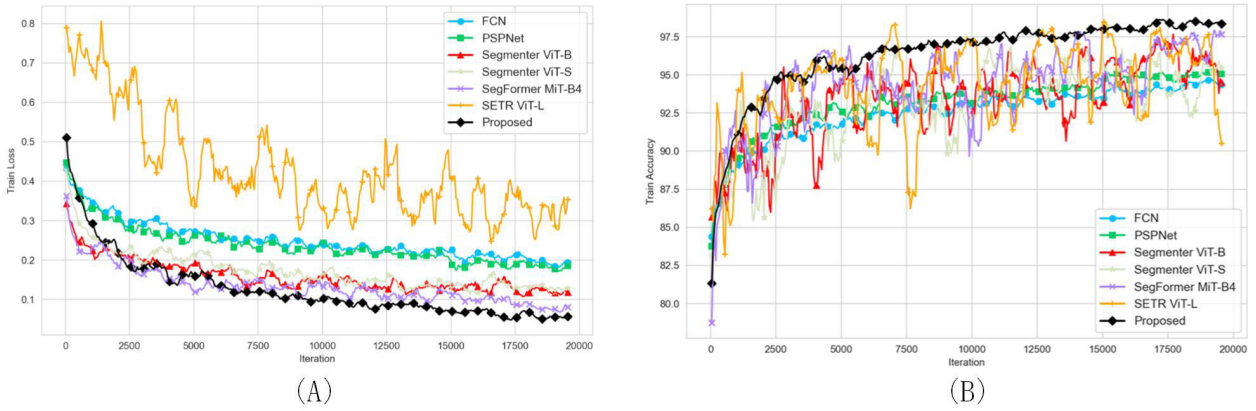


FIGURE 9
Training loss and accuracy curves for various semantic segmentation models. **(A)** Training Loss Curves **(B)** Training Accuracy Curves.

3.5 Model application verification in simulation scenarios

Using images collected from the field to simulate real-world scenarios, and Figure 11 demonstrates the SPVD detection performance of the model. In Figure 11A, labeled as Sample 1, the manually annotated original image shows diseased areas marked in red, with the background consisting of healthy leaves and soil. The white solid line indicates the 480p camera's capture window with a resolution of 480×640 pixels. Figure 11B shows the inference results from the camera frame within the capture window of Figure 11A, where the model accurately identifies the diseased leaves, effectively distinguishing them from the background.

In Figure 11C, labeled as Sample 2, the manually annotated original image includes no diseased leaves, and the background consists of healthy leaves and soil. Figure 11D shows the inference results from the camera frame within the capture window of Figure 11C, where the model accurately avoids misclassifying the healthy leaves in the background as diseased leaves.

TABLE 6 Test results of the proposed method on DS-2[a].

Segmentation method	MIoU(%)	MAcc (%)
Proposed	78.59	79.47

[a]MIoU, mean intersection over union; MAcc, mean accuracy.

4 Discussion

Computer vision and machine learning techniques have gained significant traction in agricultural applications, particularly in plant disease detection. However, despite advancements across various

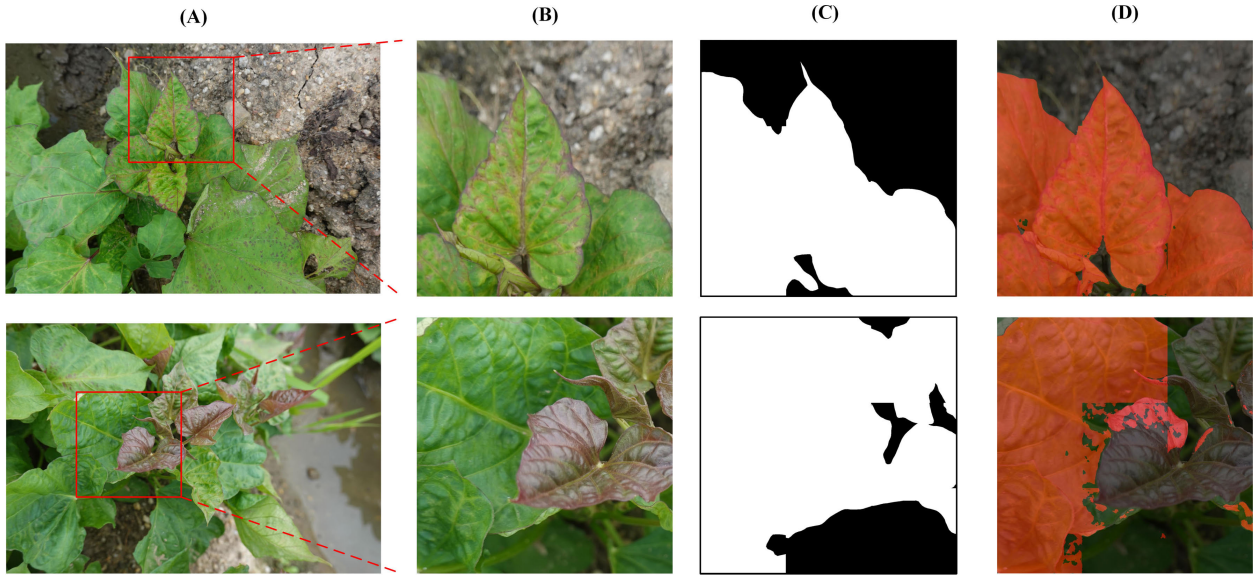


FIGURE 10
Segmentation results using proposed algorithm on DS-2. **(A)** Original images in DS-2. **(B)** The selected area **(C)** Ground truth **(D)** Detection results of proposed algorithm.

crops, there remains a lack of studies specifically addressing sweetpotato virus disease (SPVD). Considering sweetpotato's global significance as a food crop and the severe economic impact of SPVD, this study evaluates the effectiveness of the DeepLabV3+ algorithm in accurately identifying SPVD infections on sweetpotato leaves. The work addresses a critical need for rapid, precise disease detection to support crucial crop management and disease control decisions.

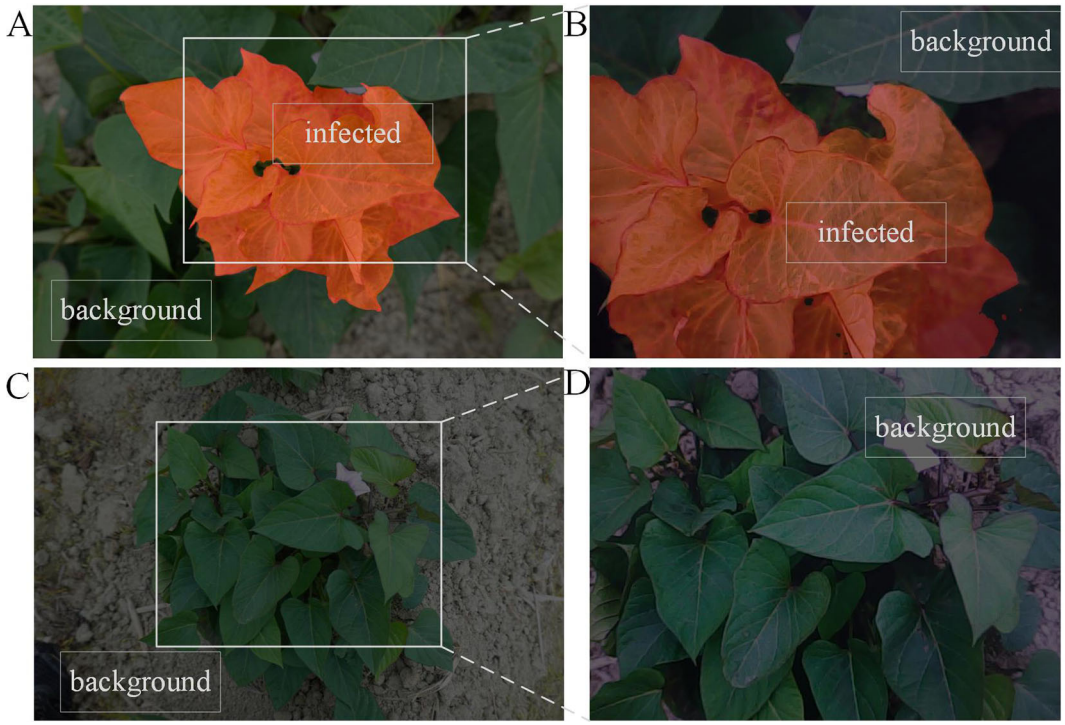


FIGURE 11
Model application verification under simulation scenarios. **(A, C)** Original images with manually annotation **(B, D)** Verification of the SPVD recognition effect of the model.

A common challenge in processing large images for segmentation tasks is the computational limitations that often require either down-sampling or splitting the images into smaller segments. While down-sampling can reduce segmentation accuracy, splitting images can introduce irrelevant background content, leading to data inefficiencies. Researchers such as Chang et al. and Jiang et al. have emphasized the importance of dividing images into smaller segments to address this issue (Chang et al., 2024; Jiang et al., 2024). In this study, the latter approach was adopted, with a focus on optimizing the segmentation process and minimizing the influence of irrelevant background areas.

To simulate real-world agricultural conditions, a cost-effective 480p Guke camera with a resolution of 640×480 pixels was used. This camera reflects the types of equipment that farmers are likely to have access to, offering a practical scenario for evaluating the model. Additionally, the lower resolution provides an opportunity to test the model's robustness in handling low-quality images, which is crucial in field conditions where high-resolution equipment may not always be available.

The DeepLabV3+ model was selected for its ability to capture both global and local features through atrous spatial pyramid pooling (ASPP). While the baseline model performed well, it encountered limitations when processing pure background images, occasionally failing to distinguish between diseased and healthy areas. To mitigate this, attention mechanisms and data augmentation techniques were introduced, significantly enhancing the model's capacity to accurately segment diseased regions while reducing errors in background segmentation. Similar approaches have been noted by researchers, who also highlighted the effectiveness of these methods in improving segmentation performance (Qin and Hu, 2020; Yang et al., 2022).

In the ablation study, the baseline model combined with Attention Pyramid Fusion (APF) and Transfer Learning (TL) demonstrated the highest accuracy on the DS-1 test set. However, since the DS-1 dataset mainly focuses on diseased areas with limited background images, the model's performance on pure background images was less satisfactory, occasionally resulting in misclassifications as square red areas. By adding data augmentation (DA), the model's handling of pure background images improved, providing clearer segmentation and better overall performance, despite a slight reduction in accuracy.

While our improved model achieves good segmentation accuracy, there is still room for improvement, particularly in reducing segmentation time and parameter counts. Although the improved model performs slightly worse on DS-2 compared to DS-1, it still effectively recognizes similar disease features in other environments. For disease appearances not included in the DS-1 training set, particularly the purple diseased apical leaves, detection performance remains an area for improvement. Given the significant differences between various sweetpotato varieties, this result is not surprising and further emphasizes that future efforts could focus on expanding the dataset to include more growth stages and diverse varieties (with varying leaf colors and shapes between top and mature leaves), thereby enhancing the model's generalization capability. Finally, validating the model under real field conditions will provide a comprehensive evaluation of its practical applicability.

5 Conclusion

By leveraging transfer learning, the improved DeepLabV3+ network's accuracy is significantly enhanced. The incorporation of Attention Pyramid Fusion (APF) improves semantic feature representation, while the novel data augmentation technique increases generalization, allowing the model to handle background noise more effectively in large, real-world images. Achieving strong segmentation performance on both DS-1 (mIoU: 94.63%, mAcc: 96.99%) and DS-2 (mIoU: 78.59%, mAcc: 79.47%) datasets, the model also demonstrates robust detection in simulated field conditions. This solution offers a practical approach to SPVD identification, reducing the need for specialized expertise and providing farmers with an accessible, efficient detection tool.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

ZD: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. FZ: Writing – review & editing, Methodology, Conceptualization. HL: Writing – review & editing, Methodology, Data curation. JZ: Writing – review & editing, Software, Methodology. JC: Writing – review & editing, Methodology, Data curation. BC: Writing – review & editing, Data curation. WZ: Writing – review & editing, Resources. XL: Writing – review & editing, Methodology, Visualization. ZW: Writing – review & editing, Resources. LH: Writing – review & editing, Supervision, Funding acquisition. XY: Writing – review & editing, Supervision, Funding acquisition.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by grants from the earmarked fund for CARS-10-Sweetpotato, the Key-Area Research and Development Program of Guangdong Province(no. 2020B020219001), Guangzhou Science and Technology Plan Project (no. 20212100068, no. 202206010088), Guangdong Province Rural Technology Special Envoy Program in Townships and Villages (Yueke Hannongzi (2021) No.1056), and Sweetpotato Potato Innovation Team of Modern Agricultural Industry Technology System in Guangdong Province(2023KJ111).

Open Fund of State Key Laboratory of Green Pesticide, South China Agricultural University(GPLSCAU202408).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Chang, T. V., Seibt, S., and von Rymon Lipinski, B. (2024). "Hierarchical histogram threshold segmentation-auto-terminating high-detail oversegmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE/CVF.

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv*.

Chen, L., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *arXiv*.

Chollet, F. (2017). Xception: deep learning with depthwise separable convolutions. *arXiv*.

Clark, C. A., Davis, J. A., Abad, J. A., Cuellar, W. J., Fuentes, S., Kreuze, J. F., et al. (2012). Sweetpotato viruses: 15 years of progress on understanding and managing complex diseases. *Plant Dis.* 96, 168–185. doi: 10.1094/PDIS-07-11-0550

David, M., Kante, M., Fuentes, S., Eyzaguirre, R., Diaz, F., De Boeck, B., et al. (2023). Early-stage phenotyping of sweet potato virus disease caused by sweet potato chlorotic stunt virus and sweet potato virus c to support breeding. *Plant Dis.* 107, 2061–2069. doi: 10.1094/PDIS-08-21-1650-RE

DeChant, C., Wiesner-Hanks, T., Chen, S., Stewart, E. L., Yosinski, J., Gore, M. A., et al. (2017). Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology* 107, 1426–1432. doi: 10.1094/PHYTO-11-16-0417-R

Gai, Y., Ma, H., Chen, X., Chen, H., and Li, H. (2016). Boeremia tuber rot of sweet potato caused by b. Exigua, a new post-harvest storage disease in China. *Can. J. Plant Pathol.* 38, 243–249. doi: 10.1080/07060661.2016.1158742

Jiang, S., Wu, H., Chen, J., Zhang, Q., and Qin, J. (2024). "Ph-net: semi-supervised breast lesion segmentation via patch-wise hardness," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE/CVF.

Kalaivani, S., Shantharajah, S. P., and Padma, T. (2020). Agricultural leaf blight disease segmentation using indices based histogram intensity segmentation approach. *Multimedia. Tools Appl.* 79, 9145–9159. doi: 10.1007/s11042-018-7126-7

Karyeija, R. F., Gibson, R. W., and Valkonen, J. (1998). The significance of sweet potato feathery mottle virus in subsistence sweet potato production in africa. *Plant Dis.* 82, 4–15. doi: 10.1094/PDIS.1998.82.1.4

Lee, I. H., Shim, D., Jeong, J. C., Sung, Y. W., Nam, K. J., Yang, J.-W., et al. (2019). Transcriptome analysis of root-knot nematode (meloidogyne incognita)-resistant and susceptible sweetpotato cultivars. *Planta* 249, 431–444. doi: 10.1007/s00425-018-3001-z

Li, F., Zuo, R., Abad, J., Xu, D., Bao, G., and Li, R. (2012). Simultaneous detection and differentiation of four closely related sweet potato potyviruses by a multiplex one-step rt-pcr. *J. Virol. Methods* 186, 161–166. doi: 10.1016/j.jviromet.2012.07.021

Li, K., Zhang, L., Li, B., Li, S., and Ma, J. (2022). Attention-optimized deeplab v3 + for automatic estimation of cucumber disease severity. *Plant Methods* 18, 109. doi: 10.1186/s13007-022-00941-8

Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE.

Long, J., Shelhamer, E., and Darrell, T. (2015). *Fully convolutional networks for semantic segmentation* (arXiv).

McGregor, C. E., Miano, D. W., LaBonte, D. R., Hoy, M., Clark, C. A., and Rosa, G. J. M. (2009). Differential gene expression of resistant and susceptible sweetpotato plants after infection with the causal agents of sweet potato virus disease. *J. Am. Soc. Hortic. Sci.* 134, 658–666. doi: 10.21273/JASHS.134.6.658

Oppenheim, D., Shani, G., Erlich, O., and Tsror, L. (2019). Using deep learning for image-based potato tuber disease detection. *Phytopathology* 109, 1083–1087. doi: 10.1094/PHYTO-08-18-0288-R

Pan, Z., Xinzhi, L., Qil, Q., Desheng, Z., Yanhong, Q., Yuting, T., et al. (2013). Development and application of a multiplex rt-pcr detectionmethod for sweet potato virus disease (spvd). *Plant Prot.* 39, 86–90.

Patil, M. A., and Manohar, M. (2022). Enhanced radial basis function neural network for tomato plant disease leaf image segmentation. *Ecol. Inf.* 70, 101752. doi: 10.1016/j.ecoinf.2022.101752

Qin, Q., and Hu, X. (2020). "The application of attention mechanism in semantic image segmentation," in *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. Piscataway, NJ, USA: IEEE.

Rännäli, M., Czekaj, V., Jones, R., Fletcher, J. D., Davis, R. I., Mu, L., et al. (2009). Molecular characterization of sweet potato feathery mottle virus (spfmv) isolates from easter island, French Polynesia, New Zealand, and southern africa. *Plant Dis.* 93, 933–939. doi: 10.1094/PDIS-93-9-0933

Srinivasu, P. N., Ahmed, S., Hassaballah, M., and Almusallam, N. (2024a). An explainable artificial intelligence software system for predicting diabetes. *Heliyon*. 10. doi: 10.1016/j.heliyon.2024.e36112

Srinivasu, P. N., Lakshmi, G. J., Narahari, S. C., Shafi, J., Choi, J., and Ijaz, M. F. (2024b). Enhancing medical image classification via federated learning and pre-trained model. *Egyptian. Inf. J.* 27, 100530. doi: 10.1016/j.eij.2024.100530

Wang, C., Du, P., Wu, H., Li, J., Zhao, C., and Zhu, H. (2021). A cucumber leaf disease severity classification method based on the fusion of deeplabv3+ and u-net. *Comput. Electron. Agric.* 189, 106373. doi: 10.1016/j.compag.2021.106373

Wang, M., Fu, B., Fan, J., Wang, Y., Zhang, L., and Xia, C. (2023). Sweet potato leaf detection in a natural scene based on faster r-cnn with a visual attention mechanism and diou-nms. *Ecol. Inf.* 73, 101931. doi: 10.1016/j.ecoinf.2022.101931

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. (2020). "Eca-net: efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE/CVF.

Wanjala, B. W., Ateka, E. M., Miano, D. W., Low, J. W., and Kreuze, J. F. (2020). Storage root yield of sweetpotato as influenced bysweetpotato leaf curl virusandits interaction withsweetpotato feathery mottle virusandsweetpotato chloroticstunt virusin Kenya. *Plant Dis.* 104, 1477–1486. doi: 10.1094/PDIS-06-19-1196-RE

Xinliang, C., Boping, F., Zhangying, W., Zhufang, Y., Yiling, Y., and Lifei, H. (2022). Identification and analysis of viruses in the national germplasm nurseryfor sweet potato (guangzhou). *Plant Prot.* 48, 258–264. doi: 10.16688/j.zwbh.2020557

Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., and Shen, F. (2022). Image data augmentation for deep learning: A survey. *Arxiv*. [Preprint]. doi: 10.48550/arXiv.2204.08610

Yang, L., Zhang, H., Zuo, Z., Peng, J., Yu, X., Long, H., et al. (2023). Afu-net: a novel u-net network for rice leaf disease segmentation. *Appl. Eng. Agric.* 39, 519–528. doi: 10.13031/aea.15581

Zhang, Q. (2022). A novel resnet101 model based on dense dilated convolution for image classification. *Sn. Appl. Sci.* 4, 1–13. doi: 10.1007/s42452-021-04897-7

Zhang, X. X., Wang, X. F., Ling, J. C., Yu, J. C., Huan, L. F., and Dong, Z. Y. (2019). Sweet potato virus diseases (spvd): research progress. *Chin. Agric. Sci. Bull.* 35, 118–126. doi: 10.11924/j.issn.1000-6850.casb18060093

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). "Pyramid scene parsing network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE.

Zhou, H., Peng, Y., Zhang, R., He, Y., Li, L., and Xiao, W. (2024). Gs-deeplabv3+: a mountain tea disease segmentation network based on improved shuffle attention and gated multidimensional feature extraction. *Crop Prot.* 132, 106762. doi: 10.1016/j.cropro.2024.106762

Zhu, J., Yang, G., Feng, X., Li, X., Fang, H., Zhang, J., et al. (2022). Detecting wheat heads from uav low-altitude remote sensing images using deep learning based on transformer. *Remote Sens.* 14, 5141. doi: 10.3390/rs14205141

Zou, K., Chen, X., Wang, Y., Zhang, C., and Zhang, F. (2021). A modified u-net with a specific data argumentation method for semantic segmentation of weed images in the field. *Comput. Electron. Agric.* 187, 106242. doi: 10.1016/j.compag.2021.106242

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 11 November 2024
DOI 10.3389/fpls.2024.1458337

Check for updates

OPEN ACCESS

EDITED BY
Syed Tahir Ata-Ul-Karim,
Aarhus University, Denmark

REVIEWED BY
Ke Zhang,
Nanjing Agricultural University, China
Haozhou Wang,
The University of Tokyo, Japan

*CORRESPONDENCE
Yongqi Ge
geyongqi@nxu.edu.cn

RECEIVED 02 July 2024
ACCEPTED 17 October 2024
PUBLISHED 11 November 2024

CITATION
Yang S, Ge Y, Wang J, Liu R and Fu L (2024) Predicting alfalfa leaf area index by non-linear models and deep learning models.
Front. Plant Sci. 15:1458337.
doi: 10.3389/fpls.2024.1458337

COPYRIGHT
© 2024 Yang, Ge, Wang, Liu and Fu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Predicting alfalfa leaf area index by non-linear models and deep learning models

Songtao Yang[1], Yongqi Ge[1,2]*, Jing Wang[1], Rui Liu[1,2] and Li Fu[3]

[1]College of Information Engineering, Ningxia University, Yinchuan, China, [2]Ningxia Key Laboratory of Artificial Intelligence and Information Security for Channeling Computing Resources from the East to the West, Yinchuan, China, [3]College of Resources Environment and Life Sciences, Ningxia Normal University, Guyuan, China

Leaf area index (LAI) of alfalfa is a crucial indicator of its growth status and a predictor of yield. The LAI of alfalfa is influenced by environmental factors, and the limitations of non-linear models in integrating these factors affect the accuracy of LAI predictions. This study explores the potential of classical non-linear models and deep learning for predicting alfalfa LAI. Initially, Logistic, Gompertz, and Richards models were developed based on growth days to assess the applicability of nonlinear models for LAI prediction of alfalfa. In contrast, this study combines environmental factors such as temperature and soil moisture, and proposes a time series prediction model based on mutation point detection method and encoder-attention-decoder BiLSTM network (TMEAD-BiLSTM). The model's performance was analyzed and evaluated against LAI data from different years and cuts. The results indicate that the TMEAD-BiLSTM model achieved the highest prediction accuracy (R^2 > 0.99), while the non-linear models exhibited lower accuracy (R^2 > 0.78). The TMEAD-BiLSTM model overcomes the limitations of nonlinear models in integrating environmental factors, enabling rapid and accurate predictions of alfalfa LAI, which can provide valuable references for alfalfa growth monitoring and the establishment of field management practices.

KEYWORDS

alfalfa, leaf area index, non-liner model, deep learning model, MOSUM.

1 Introduction

Alfalfa (*Medicago sativa* L.) is regarded as the "king of forages" upon account of its remarkable forage properties, prolific production, high protein and nutritious content, and good palatability (Baral et al., 2022). Between 2010 and 2020, China's import of alfalfa hay increased from 200,000 tons to 1.4 million tons in response to the growing market demand (Wang and Zhang, 2023). To reduce import dependence, China actively expanded its alfalfa cultivation area and increased alfalfa production (Wang et al., 2023). However, alfalfa production faces significant challenges due to regional climate variations, limited water

resources, and soil conditions (Qian et al., 2020). The leaf area index (LAI) of alfalfa is a crucial indicator for defining the canopy structure of alfalfa, statistically describing the growth and density changes of alfalfa leaf populations (Tooley et al., 2024). It can be used to assess the dynamics of leaf growth and yield in alfalfa (Tripathi et al., 2018). Therefore, predicting the alfalfa leaf area index is significant for monitoring growth dynamics and guiding field management.

Methods for predicting LAI through direct measurement of leaf area are often destructive and can be costly. Numerous studies have developed nonlinear models for LAI prediction by analyzing the optimal growth requirements at each developmental stage of plants. Examples include the Logistic model, Gompertz model, Richards model, and Schnute model (Zhang et al., 2021; Duran and Ünal, 2024). Guo et al. (2022) established a generalized Logistic model using growing degree days (GDD) and relative growing degree days (RGDD) as key parameters to describe the height, leaf area index, and biomass accumulation of maize. Karadavut et al. (2010) was carried out on five maize varieties (Monton, Ranchero, Progen 1550, 35 P12, and TTM 81-19) to explain the fitting performance of the Richards model on leaf data. Chaturvedi et al. (2017) developed allometric models for estimating the leaf area index (LAI) of Tectona grandis (teak) trees across ten diameter classes in India. Non-linear regression models, especially logistic and Gompertz, effectively explained over 60% of LAI variability. The advantages of such models include their simplicity, ease of calculation, and reasonable interpretability. However, the limitations are also obvious. This is mainly reflected in two aspects: first, the parameter settings of such models need to be verified and calibrated through years of field experiments, and their applicability in different regions is poor (Lin et al., 2023; Brogi et al., 2020); second, such models are usually one-dimensional, lacking the consideration of a large number of environmental influences, and the dynamic change of variables over time is not taken into account (Desai et al., 2020).

In contrast, deep learning is more suitable for handling non-linear prediction problems. Deep learning models can improve their predictive ability by self-learning the relationships between data from multiple dimensions based on historical datasets, including crop growth information, crop environmental information, and field management strategies. Ze-hao et al. (2020) used LSTM to model winter wheat temporal LAI and LAI in different growth periods, and the results showed that the LSTM network has good prediction ability. Long Short-Term Memory (LSTM), while capable of obtaining long-term dependent time series estimates from continuous time series data, uses only prior state knowledge and ignores back propagation of information about current vegetation changes (Sherstinsky, 2020). BiLSTM adds a reverse operation based on LSTM, which is better than LSTM at capturing the relationship between sequential features (Siami-Namini et al., 2019). LSTM shows excellent performance in time-series based prediction results, but performs poorly in dealing with data mutations.

LSTM has strong self-learning and self-adaptive ability, which overcomes the shortcomings of classical non-linear models to a certain extent. However, alfalfa is often at the maximum LAI value during the cutting, which leads to LAI data mutations after cutting. The presence of these mutation points may greatly reduce the predictive performance of LSTM. Researchers have proposed detection strategies for the change point problem, including maximum likelihood, least squares, minimum absolute length, and minimum descriptive length (Anastasiou and Fryzlewicz, 2022; Chen et al., 2022). This paper examines the number and location of mutations in alfalfa leaf area index data using moving sum (MOSUM) method. The MOSUM method not only effectively detects data mutations, but also greatly reduces the computational complexity (Burczaniuk and Jastrzębska, 2024).

To overcome the limitations of nonlinear models in integrating environmental factors and the inefficiency of LSTM in handling abrupt changes in LAI data, this study proposes a TMEAD-BiLSTM method for the rapid and accurate prediction of alfalfa leaf area index. The main contributions of this paper are as follows:

1) We established Logistic, Gompertz, and Richards models for predicting alfalfa leaf area index based on growing degree days, achieving a prediction accuracy (R^2) greater than 0.78 for all three models.

2) To address the issue of abrupt data changes after alfalfa harvesting, we proposed a TMEAD-BiLSTM model that combines the MOSUM method with a bidirectional long short-term memory (BiLSTM) encoder-decoder neural network. This model predicts the leaf area index (LAI) of alfalfa by utilizing its annual cycle and different planting strategies. The results demonstrate that this deep learning model achieved the highest prediction accuracy ($R^2 > 0.99$).

2 Materials and methods

This section introduces the experimental dataset, which comes from the publicly available dataset we previously worked on. Secondly, we describe the non-linear model and TMEAD-BiLSTM model used for predicting alfalfa leaf area index, and display the model training parameters and evaluation indicators. Then the training process and techniques used to optimize model performance were discussed. Finally, we propose evaluation metrics for assessing and comparing the accuracy of the models.

2.1 Dataset

The acquisition of alfalfa LAI data requires a significant amount of manpower and resources, and adverse weather conditions may affect the accuracy and feasibility of data collection, further increasing the challenge of obtaining continuous time series data. To the best of our knowledge, only we have publicly released the alfalfa leaf area index dataset (Yang et al., 2024a). The dataset provides growth data of alfalfa under different water and nitrogen treatments, as well as meteorological data for the entire growth period of alfalfa and soil moisture data at different depths (0-10 cm, >10-20 cm, >20-30 cm). The dataset includes field trial data from three years of history (2017-2018, 2022), with a 7-day interval for collecting alfalfa LAI data. During the period from 2019 to 2021, the experimental field was utilized for planting silage maize using a rotation method. The dataset contains a total of 955 leaf area

index data. The interval between meteorological and soil moisture data is one day. Integrating multiple features is beneficial for expanding the applicability of estimation and prediction models and improving their accuracy. This paper used non-linear regression models and deep learning models to model the alfalfa leaf area index.

We used this dataset to model the alfalfa leaf area index, as shown in Figure 1. The dataset contains 955 leaf area index data, the dataset is divided according to the following proportion: (Train: Val = 8: 2): Test = 8: 2. Deep learning models require a large number of samples to participate in training, and we use logistic models with relative leaf area index and growth days for data augmentation. Calculate daily LAI based on the maximum measured LAI value, as shown in equation (14). Thus, a daily LAI dataset with a history of three years is generated, and the partially enhanced dataset is shown in Table 1.

$$LAI = f(GD) \times LAI_{\max} \tag{1}$$

Where, $f(\cdot)$ represents the fitting curve; GD represents the number of growing days; LAI_{max} represents the maximum measured LAI value.

2.2 Non-liner model

The growth of plants over time typically follows an "S" shaped curve, and such growth characteristics are often simulated using mathematical models such as the Logistic model, Gompertz model, and Richards model. Currently, research on modeling the leaf area index of alfalfa using mathematical models is still in its early stages. In this study, the growth days are taken as influencing factors, and single variable Logistic, Gompertz, and Richards models are used to model four different cuttings of alfalfa. The applicability of Logistic, Richards, and Gompertz models in predicting the leaf area index of alfalfa is evaluated.

While the trend of alfalfa LAI growth days is basically consistent, significant differences in the LAI values among different water and nitrogen treatment plots. The relative leaf area index (RLAI) can eliminate the influence of mathematical models on the fitting of alfalfa LAI under different water and nitrogen treatments, thereby allowing for more accurately analyzing the growth status of alfalfa under different treatments (Su et al., 2022; Liu et al., 2020). Using three non-linear regression models, Logistic (Formula 2), Richards (Formula 3), Gompertz (Formula 4), to describe the trend of LAI changes during plant growth (Nhu et al., 2020; Roosa et al., 2020; Vaghi et al., 2020):

$$RLAI = \frac{LAI}{LAI_{\max}} = \frac{1}{1 + e^{a_1 + b_1 x + c_1 x^2}} \tag{2}$$

$$RLAI = \frac{LAI}{LAI_{\max}} = \frac{a_2}{(1 + e^{b_2 - c_2 x})^{\frac{1}{d}}} \tag{3}$$

$$RLAI = \frac{LAI}{LAI_{\max}} = a_3 e^{-c_3 e^{-b_3 x}} \tag{4}$$

Where, $RLAI$ is the relative leaf area index; LAI is leaf area index; LAI_{max} is the theoretical maximum LAI. a_1, b_1, c_1, a_2, b_2, c_2, a_3, b_3, c_3 and d are model fitting parameters; x is the growth days.

2.3 TMEAD-BiLSTM model

2.3.1 Overall framework

The construction of the model can be divided into three parts: (1) The MOSUM method is used to detect mutation points in time series data (Figures 2A, B). (2) A model based on encoder decoder was constructed, where both the encoder and decoder are BiLSTM models (Figures 2C, E). In order to solve the problem of feature

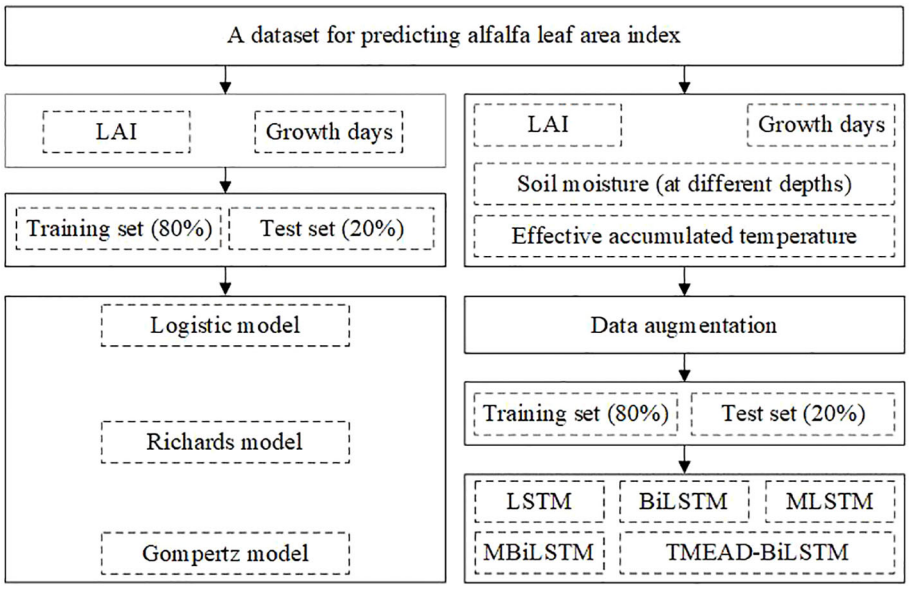


FIGURE 1
Dataset partitioning and prediction strategies.

TABLE 1 Partial data of LAI dataset.

DateTime	GDD	Growth days	Soil moisture 10	Soil moisture 20	Soil moisture 30	LAI
2018/6/13	194	12	28.45	29.81	25.02	1.26
2018/6/14	212	13	26.88	28.45	24.39	1.40
2018/6/15	231	14	25.62	27.02	23.59	1.56
2018/6/16	245	15	24.65	25.59	22.93	1.74
2018/6/17	260	16	24.71	25.6	22.8	1.95
2018/6/18	278	17	24.15	24.89	22.41	2.18
2018/6/19	297	18	23.23	23.61	21.76	2.43
2018/6/20	318	19	22.15	22.11	21.01	2.71
2018/6/21	337	20	21.14	20.9	20.38	3.01

disappearance and information loss caused by covariate feature compression in long sequence data, we introduced an attention mechanism in the model to capture the long-term dependency relationship between input decoder covariates and feature variables (Figure 2D). The fully connected layer outputs the predicted results (Figure 2F). (3) During the training process, training batches containing mutation points are eliminated through pre-defined functions. This section focuses on the construction process of the TMEAD-BiLSTM model, the overall structure of which is shown in Figure 2.

2.3.2 Mutation point detection based on MOSUM

Time series data usually contains mutation points, and there is currently no widely recognized method to handle these mutation data points. The currently known processing methods include removing abnormal data or using linear interpolation for filling (Liguori et al., 2021). However, these methods may disrupt the time series properties of the data and affect the accuracy of the analysis. Time series prediction has made remarkable progress in recent years, evolving from traditional statistical methods and machine learning to the latest deep learning techniques, thereby advancing the field of time series analysis (Masini et al., 2023). However, there are still some basic problems in practice, and one of the main problems is the existence of lag difference. Lag differences are deviations or delays that occur in the predicted sequence, which can affect the accuracy and robustness of the model (Samanta et al., 2020). Time series data usually contain abrupt change, and the presence of abrupt change greatly reduces the prediction performance of the model when there is a lag in the prediction process, as shown in Figure 3. As a forage crop, alfalfa shows a positive correlation between leaf area index (LAI), and final yield throughout its growth cycle. During mowing, alfalfa LAI was generally reaches its maximum value, and LAI mutations occur afterward, significantly impacting the prediction performance of deep learning models.

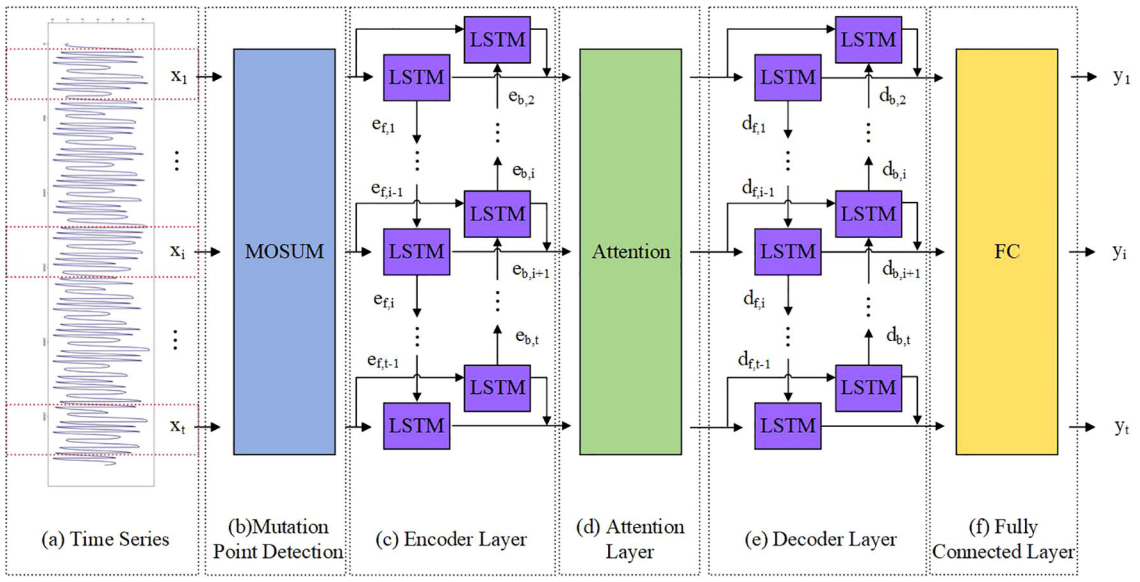


FIGURE 2
A hybrid model combining attention based BiLSTM encoder-decoder neural network and MOSUM (TMEAD-BiLSTM for short).

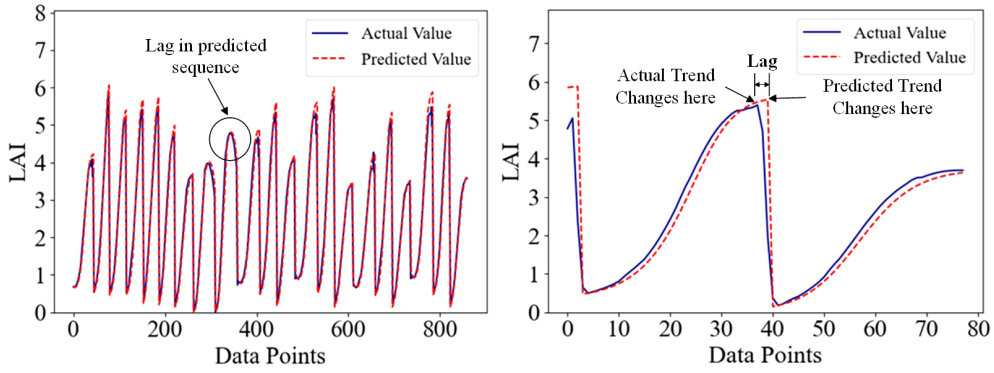


FIGURE 3
Demonstration of lag in time series forecasting.

The MOSUM method is a technique used to detect structural changes in time series, commonly used for detecting outliers (Eichinger and Kirch, 2018). The core principle is to monitor the evolution of data by calculating the cumulative sum within a sliding window. As the window slides, calculate the magnitude of the cumulative sum change within the current window and set a predetermined threshold. When the cumulative sum changes beyond this threshold, the corresponding window position is marked as an outlier. The MOSUM method utilizes the bandwidth parameter (G value) to adjust the size of the sliding window to adapt to different degrees of variation. The mathematical formula of the MOSUM method is as follows:

$$S_t = \sum_{i=t-k+1}^{t} X_i \tag{5}$$

$$R_t = S_t - S_{t-G} \tag{6}$$

Where, S_t represents the cumulative sum of the sliding window in the time series; $R_t$represents the amount of change in the time series at time t; X_i represents the data point in the time series; k represents the size of the sliding window; and G represents the bandwidth parameter.

As a forage crop aimed at obtaining plants, alfalfa has a positive correlation between leaf area index (LAI) and crop growth and final yield during its growth cycle. The maximum LAI of alfalfa during the growth period is generally observed during cutting, which leads to LAI mutations and significantly affects the predictive performance of deep learning models. Therefore, we consider using the MOSUM method to detect mutation points in alfalfa LAI data, as shown in Figure 4.

2.3.3 Attention with encoder and decoder

The encoder network is a BiLSTM that sequentially transforms the input time series $\{x_1,\ldots, x_i,\ldots, x_t\}$. The mapping from x_i to $e_{h,i}$ and $e_{b,i}$ can be obtained, as shown in Formula 6–8:

$$e_{h,i} = f_1(W_f[e_{h,i-1}, x_i]) \tag{7}$$

$$e_{b,i} = f_2(W_b[e_{b,i+1}, x_i]) \tag{8}$$

$$encoder_outputs_i = [e_{h,i}; e_{b,i}] \tag{9}$$

Where, $e_{h,i}$ and $e_{b,i}$ are the forward and backward hidden layer outputs of encoder i at time; $\boldsymbol{W}_f$ and $\boldsymbol{W}_b$ represent the weight matrices of forward and backward units, respectively; f_1 and f_2 are units of LSTM; $encoder_outputs_i$ represents the output of the encoder at time step i.

Using weighted attention to compute the weighted average of encoder outputs to generate context vectors:

$$attn_weights_i = softmax(e_{h,i} \cdot e_{b,i}) \tag{10}$$

$$context = \sum_{i=1}^{T} attn_Weights_i \cdot encoder_outputs_i \tag{11}$$

Where, $attn_weights_i$ represents the attention weights; $context$ represents the context vectors.

The decoder network is also a BiLSTM, $d_{h,i}$ and $d_{b,i}$ are the hidden layer outputs of the decoder at time i. As shown in Formula 11, 12, $d_{h,i}$ and $d_{b,i}$ can be obtained:

$$d_{h,i} = f_3(W_f[d_{h,i-1}, x_i, context]) \tag{12}$$

$$d_{b,i} = f_4(W_b[d_{b,i+1}, x_i, context]) \tag{13}$$

Where, $\boldsymbol{W}_f$ and $\boldsymbol{W}_b$ indicates the weight matrices of forward and backward units, respectively; f_3 and f_4 are units of LSTM; $d_{h,i}$ and $d_{b,i}$ respectively represent the forward and backward hidden layer outputs of the decoder at time step i.

BiLSTM can extract complex features of time series, and attention layer output and hidden state information $d_{f,i-1}$、$d_{b,i+1}$ can improve the prediction performance. Afterwards, the predicted values are output through the fully connected layer:

$$output_i = \mathrm{fc}(d_{h,i}, d_{b,i}) \tag{14}$$

Where, $output_i$ represents the prediction result; fc denotes the fully connected layer.

2.3.4 Predefined function

Before the start of training, the MOSUM method is used to detect potential mutation points in the data. When selecting a training batch, predefined functions are used to determine whether to include mutation points in the upcoming training batch. If a mutation point

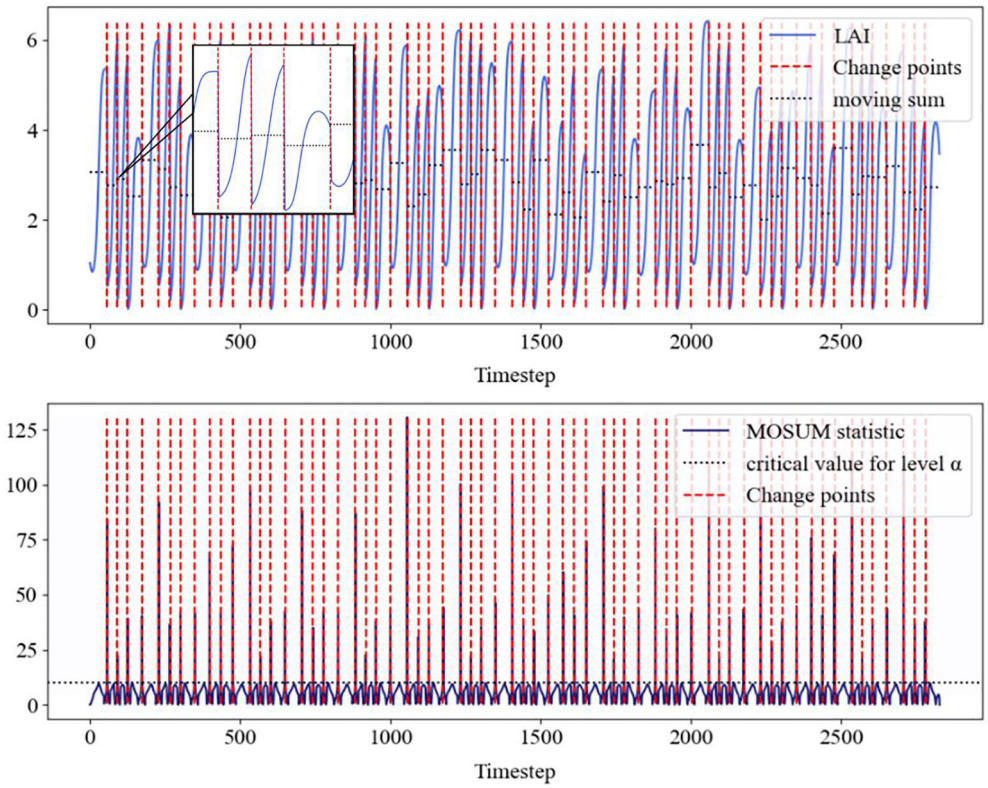


FIGURE 4
The above figure shows the LAI data of alfalfa detected by MOSUM (blue solid line), the mutation points (red dashed line), as well as the corresponding movement and (black dashed line). The following figure shows the MOSUM statistic, which identifies the mean, threshold (black dashed line), and corresponding change points (red dashed line).

exists, the batch will be skipped until a batch without the mutation point is found. The specific process is shown in Figure 5.

In detail, the batch size s is one of the most important control parameters of our method. We have to choose s in such a way that it could cover at most one change point. If s is larger than the distance between two change points c_i, c_j, the part of the time series between time index i and j would never be learned. The batches containing this part of the time series would always contain at least one change point and would, therefore, not be allowed. To avoid this, we develop a method that automatically selects the maximum batch size based on the change points. In this method we calculate the distance between all detected change points. We then set s_{max} to be half of the smallest distance between the change points. Taking half of this distance increases the probability of selecting the allowable batch between the change points, and the resulting number s_{max} is the maximum batch size that should be selected. See Algorithm 1 for details.

Input: training data d, change point detection method A
Output: maximal batch size s_{max}, change points $c_0, \ldots, c_{k-1}$
$c_0, \ldots, c_{k-1} \leftarrow$ MOSUM(d)
for $i, j \in \{0, \ldots, k-1\}$, $i \neq j$ **do**
  $diff_{i,j} \leftarrow ||c_i - c_j||_2$
end for
$s_{max} \leftarrow \lceil \min_{i,j}\{diff_{i,j}\}/2 \rceil$
return $c_0, \ldots, c_{k-1}, s_{max}$

Algorithm 1. Find change points and maximum batch size.

In the TMEAD-BiLSTM model, we pass a batch size $s \leq s_{max}$ and the detected change points $c_0, \ldots, c_{k-1}$ to the algorithm. Our goal is to find the start and end points of the batches that do not contain any change points. To do this, we select indices that lie between 0 and n-s-1. For each change point, we check whether the index of the respective change point lies between the previously selected start index and end index of the batch. If the change point lies in the batch, we repeat the method and select a new start index, and perform the same process for this, as shown in Algorithm 2. For example, a set of change points has been found through the MOSUM method:

$$C := \{c_0, \ldots, c_{k-1}\} \tag{15}$$

Select batch size $s \leq s_{max}$ in Algorithm 1. Select a random starting index *start*, that is in the range [0, n-s-1] of the time series. This gives us a batch with start point i_{start} and end point:

$$i_{end} := i_{start} + s - 1 \tag{16}$$

Assume that one of the change points lies in the batch, thus without loss of generality $\exists j \in \{0, \ldots, k\text{-}1\}$ such that $c_j \in [i_{start}, i_{end}]$. Then again a random start index $i_{start'}$ is chosen and:

$$i_{end'} = i_{start'} + s - 1 \tag{17}$$

If now ∀j∈{0,…, k − 1} holds:

$$c_j \notin [i_{start'}, i_{end'}] \tag{18}$$

In the following batch $[i_{start'}, i_{end'}]$ is the valid batch.

Input: training data d, training data d, batch size $s \le s_{max}$, change points $_{c0,\ldots,ck-1}$
Output: start and end points of the batch
n ← len(d)
repeat
  i_{start}←{0, …, n-s-1}
  valid←True
  for j∈{0, …, k-1}
    if $c_j \ge i_{start}$ **and** $c_j \le i_{start}$ +s-1 **then**
      Valid ← False
      break for
    end if
  end for
until valid = True
i_{end} ←i_{start}+s-1
return i_{start}, i_{end}

Algorithm 2. Find valid batch.

2.4 Model training

Train the hybrid model using the interpolated dataset, with a training-to-testing ratio of 8:2. The encoder and decoder are both BiLSTM with 12 hidden units, using the ReLU activation function. The attention mechanism generates context vectors to enhance output by calculating the correlation between queries and keys. The model uses a learning rate of 0.6e-3 and an Adam optimizer, with a batch size of 32 and a maximum of 300 rounds of training, combined with an early stopping strategy to prevent overfitting. All experiments in this paper were performed on Windows 10 Professional, version 21H2, on a computer with the following main parameters (CPU: AMD Ryzen 5 3600 6-core processor 3.60 GHz; GPU: NVIDIA GeForce GTX 1080 Ti GPU). The code was compiled using PyCharm version 2021.2.3 compiler, and the full syntax followed Python version 3.7.

2.5 Evaluation metrics

To prove the prediction performance of TMEAD-BiLSTM and non-linear model, three metrics are adopted to evaluate its

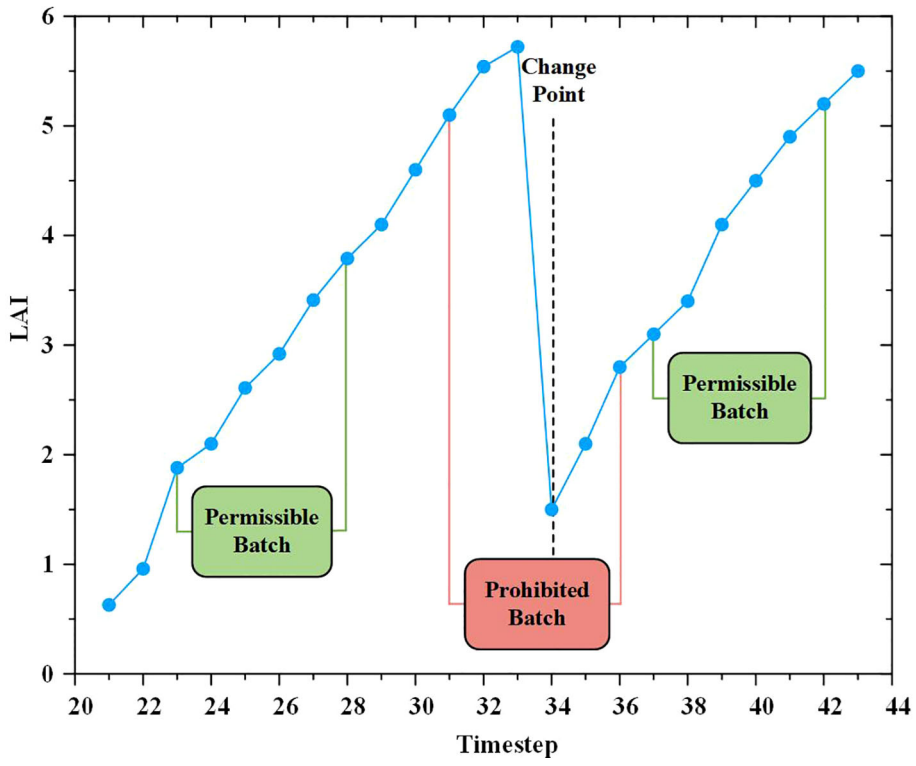


FIGURE 5
Example of batch processing selection around mutation points. The first green area shows a batch created by choosing t = 23 as the start index, and with a selected batch size of s = 6, the batch contains the points t = 23 to t = 28. The change point is located at t = 34 and is therefore outside the detected batch. The batch is therefore permissible. The red area next to it shows a prohibited batch. The start index of this batch is t = 31 and the end index at t = 36, so the change point is located inside the batch and is not allowed in training. A new batch must be found. The green area to the right indicates another permissible batch, since the change point lies outside of it.

prediction accuracy. Specifically, mean absolute error (MAE), coefficient (R^2) and root mean square (RMSE) are adopted. The R^2 signifies the agreement between estimated and measured values, with a value closer to 1 indicating a better fit of the model. The RMSE indicates the extent of deviation between estimated and measured values, with smaller values suggesting a better model fit. The MAE assesses the actual deviation between estimated and measured values, with smaller values indicating higher model accuracy. The calculation formula is as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{19}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{20}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - y)^2}{\sum_{i=1}^{n}(y_i - y)^2} \tag{21}$$

3 Experiment results

3.1 Prediction results annual alfalfa leaf area index

3.1.1 Prediction results of non-linear model

The relationship between the relative leaf area index and the number of growing days is illustrated in Figure 6. The Logistic, Richards, and Gompertz models were employed to fit the changes in the relative leaf area index over the growing days shown in Figure 6, respectively. The fitting results of the logistic (Formula 22), Richards (Formula 23), and Gompertz (Formula 24) models are as follows:

$$RLAI = \frac{LAI}{LAI_{\max}} = \frac{1}{1 + e^{2.4249 - 0.1699x + 0.0003x^2}} \tag{22}$$

$$RLAI = \frac{LAI}{LAI_{\max}} = \frac{0.9334}{(1 + e^{-0.2767 - 26.7844x})^{\frac{1}{2.7461}}} \tag{23}$$

$$RLAI = \frac{LAI}{LAI_{\max}} = 0.9971e^{-0.1137e^{-6.8693x}} \tag{24}$$

Table 2 illustrates the fitting and validation results of three non-linear models for the relative leaf area index. It can be observed that all three models exhibit acceptable fits (R^2>0.78), and the models were validated using test data. It is evident that there is no significant difference in the predictive performance of the three models for the alfalfa leaf area index.

3.1.2 Prediction results of deep learn model

Due to differences in alfalfa phenology under different meteorological conditions and soil moisture, growth days, and soil moisture at different depths (>0-10 cm, >10-20 cm, >20-30 cm) corresponding to LAI at time t were also used as input features. To ensure consistency in input features, the features were normalized and scaled to 0-1. Parallel control experiments were conducted between the TMEAD-BiLSTM model and four models: LSTM, BiLSTM, MLSTM and MBiLSTM. The prediction results of each model are compared using RMSE、 MAE and R^2 measures the degree of fitting of the model. The prediction results of each model are shown in Table 3.

As shown in Figure 7, the model has a high accuracy in predicting 1.5 ≤ LAI ≤ 5.5, with both predicted and labeled values near the 1:1 line. When LAI < 1.5 and LAI > 5.5, the prediction accuracy is slightly lower compared to the previous situation. This may be because under different growth conditions, the time for alfalfa to turn green and set pods is different, and there are significant differences in LAI changes during the corresponding stages of LAI increase and decrease.

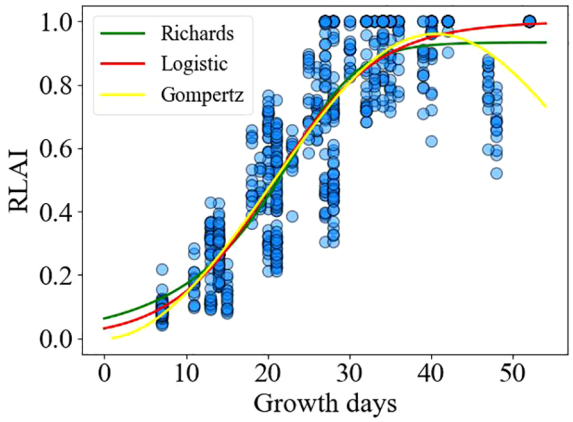


FIGURE 6
Fitting diagram of annual leaf area index of alfalfa.

TABLE 2 Modeling and validation results.

Metrics	Fitting results			Verification results		
	Richards	Logisitc	Gompertz	Richards	Logisitic	Gompertz
R^2	0.7944	0.7892	0.7968	0.8414	0.8362	0.8412
RMSE	0.1421	0.1439	0.1413	0.1191	0.1211	0.1192
MAE	0.1132	0.1112	0.1109	0.0952	0.0936	0.0947

TABLE 3 Comparison of prediction accuracy of different LAI prediction models.

Metrics	LSTM	BiLSTM	MLSTM	MBiLSTM	TMEAD-BiLSTM
R^2	0.8935	0.9179	0.9761	0.9889	0.9986
RMSE	0.5400	0.4857	0.2592	0.1810	0.0662
MAE	0.2857	0.2060	0.2134	0.1406	0.0397

The LSTM network performs well in temporal data processing due to its structural features, but the prediction results are not ideal under the influence of multiple feature factors. In contrast, BiLSTM networks have the ability to perceive forward and backward data, enabling stronger causality between forward and backward data. Based on the data presented in Table 3, it can be seen that the BiLSTM model has a 27.90% reduction in MAE compared to LSTM. In addition, after adding the MOSUM method, the prediction accuracy of all models was significantly improved, verifying that the MOSUM method can better preserve time series features and effectively reduce the impact of mutation points on model prediction accuracy. To further improve the model's capacity to mine feature factors and time, this study combines the MOSUM approach with the BiLSTM encoder decoder network model based on attention mechanism. The prediction results show that the R^2 and RMSE of the TMEAD-BiLSTM model are 0.9986 and 0.0662, respectively.

3.2 Prediction results of leaf area index of alfalfa in different cutting

3.2.1 Prediction results of non-linear model

Due to the influence of growth environment temperature, the growth rates of alfalfa vary between different cuttings (Lloveras et al., 1998). It is challenging to simulate the leaf area index changes for multiple cuttings within a year using nonlinear models. Therefore, we modeled the leaf area index of alfalfa for different cuttings separately. Although the variation trend of alfalfa leaf area index with growth days is basically consistent for the same cutting, there are significant differences in LAI values among the experimental plots. In order to analyze its intrinsic mechanism, the relative leaf area index was used to analyze its common growth characteristics.

Figure 8 shows the relationship between the relative leaf area index (RLAI) and growth days. The RLAI of alfalfa was fitted using the Logistic, Richards and Gompertz models. Tables 4, 5 presents the fitting effects and fitting coefficients of the models, respectively. The R^2, RMSE and MAE values of three models were acceptable. Additionally, the maximum LAI of the fourth cutting in the observed region was around 5, and the LAI of the fourth cutting exhibited a declining trend in the later growth period (Figure 8). The Logistic and Gompertz models effectively capture the decline process of the leaf area index of the fourth cutting of alfalfa (R^2 > 0.9). We utilized the Logistic, Richards, and Gompertz models to analyze the rate of change between the alfalfa leaf area index and growth days, computing the first derivative of the fitting curves for different cuttings. Setting growth days to 15, 20, 25, 30, 35, and 40, we averaged the slopes obtained from the three models. The results indicate that the fastest growth rates of alfalfa in different cutting occur at approximately 32, 23, 20, and 18 days after sowing for different cuttings, which is consistent with the results of Yang et al. (2024a)'s study.

The model obtained was validated using the test dataset, and the validation results are shown in Table 6. The simulated values of the leaf area index (LAI) based on the Logistic and Gompertz models

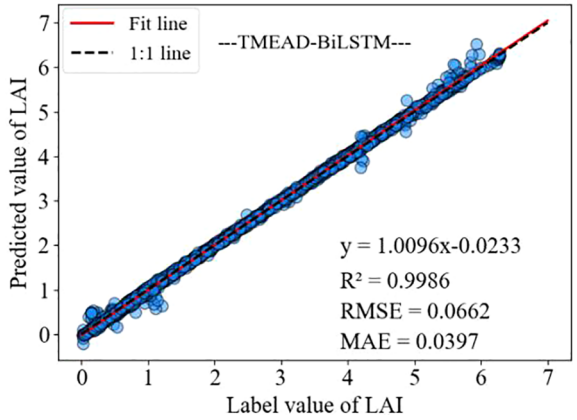


FIGURE 7
LAI prediction model prediction results.

for alfalfa exhibit a high degree of agreement with the measured values. The R^2 values for the simulated curves of different cutting times are all greater than 0.9, indicating strong goodness-of-fit. Compared to the Logistic and Gompertz models, the Richards model shows no significant advantage in fitting the data from the first three cuttings, and it performs poorly in fitting the data from the fourth cutting. In summary, the fitting of the Logistic and Gompertz models is satisfactory.

3.2.2 Prediction results of deep learn model

Alfalfa is harvested multiple times annually, and the publicly published dataset used in this study includes data from 3-4 harvests per year. During the growth process of alfalfa, there are differences in the changes in LAI of alfalfa in different cutting. Therefore, there are differences in the accuracy of LAI prediction models for alfalfa in different cutting. As shown in Figure 9, the alfalfa LAI dataset was divided into four batches, namely the first, second, third, and fourth batches. The accuracy analysis of these four batches was performed using the TMEAD-BiLSTM prediction model.

As shown in Table 7, the prediction accuracy of the TMEAD-BiLSTM model was higher than that of other experimental models in four different crop cycles. Due to the significant impact of temperature on the growth of alfalfa, the growth periods of the second and third batches of alfalfa are from May to September each year, and the LAI growth rate is relatively high. The model's ability to capture rapid changes in LAI is insufficient, resulting in a slight decrease in prediction accuracy.

4 Discussion

4.1 Comparative analysis of model accuracy

In this study, non-linear and deep learning models were used to predict alfalfa LAI. The results indicate that the prediction accuracy of the TMEAD-BiLSTM model is the highest, followed by the baseline time series forecasting models (LSTM, BiLSTM, MLSTM, MBiLSTM), while the logistic, Richards, and Gompertz models exhibit the poorest predictive performance ($R^2 > 0.78$), with no significant difference observed among the three nonlinear models in terms of predictive efficacy. In contrast to traditional nonlinear models, deep learning models have the capability to integrate multiple environmental factors (e.g., water and nitrogen treatments, soil moisture, meteorological data) for LAI prediction (Cheng et al., 2022). This may account for the superiority of deep learning models over other methods. Extensive studies have revealed that the growth of alfalfa is influenced by various environmental factors (Kaiwen et al., 2020; Liu et al., 2021). Based on our previous work on the dataset, we found that even under the same growth days, the leaf area index (LAI) of alfalfa varies under different water and nitrogen treatments. This observation is consistent with the findings of Ma et al. (2024) and Feng et al. (2016). The dataset we provide includes LAI data from the years 2017-2018 and 2022. The year 2022 marks the first year of alfalfa planting, and the planting period experienced higher

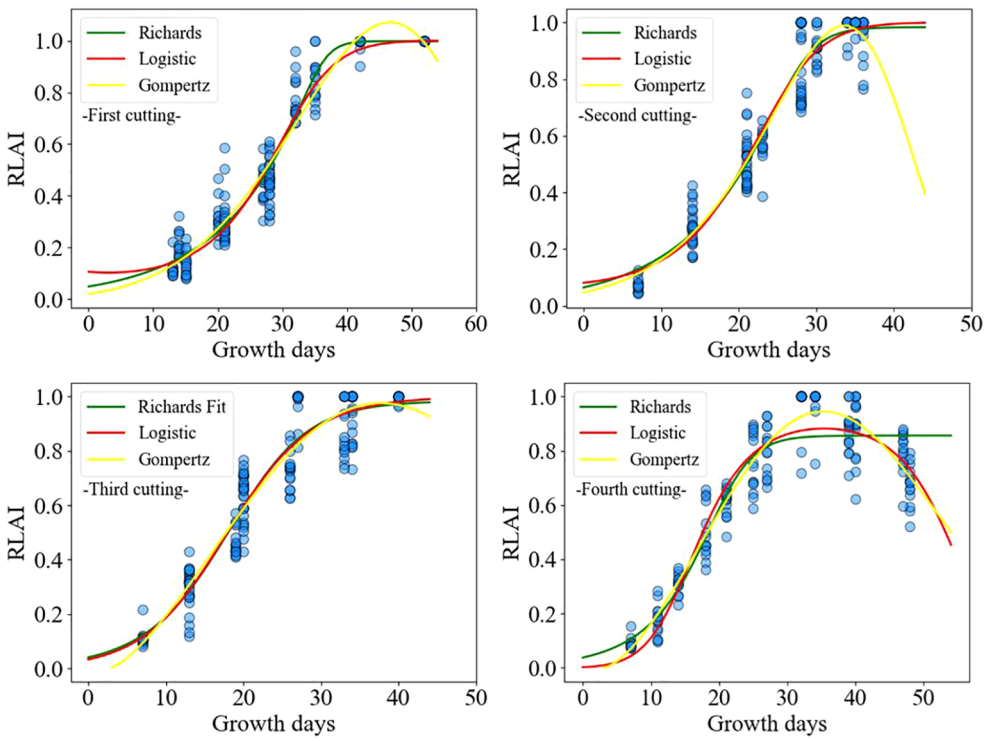


FIGURE 8
Relationships between relative leaf area index (RLAI) and growth days.

TABLE 4 Modeling results for different cutting.

Cutting time	Metrics	Richards	Logistic	Gompertz
First cutting	R^2	0.9451	0.9403	0.9424
	RMSE	0.0756	0.0800	0.0774
	MAE	0.0507	0.0576	0.0523
Second cutting	R^2	0.9331	0.9305	0.9353
	RMSE	0.0815	0.0831	0.0802
	MAE	0.0632	0.0649	0.0614
Third cutting	R^2	0.9091	0.9089	0.9051
	RMSE	0.0902	0.0903	0.0922
	MAE	0.0733	0.0723	0.0722
Fourth cutting	R^2	0.8698	0.9042	0.9210
	RMSE	0.1057	0.0907	0.0823
	MAE	0.0842	0.0765	0.0650

TABLE 6 Verification results of different cutting.

Cutting time	Metrics	Richards	Logistic	Gompertz
First cutting	R^2	0.9571	0.9546	0.9550
	RMSE	0.0676	0.0696	0.0693
	MAE	0.0473	0.0522	0.0474
Second cutting	R^2	0.9360	0.9371	0.9371
	RMSE	0.0770	0.0763	0.0763
	MAE	0.0593	0.0593	0.0575
Third cutting	R^2	0.9244	0.9245	0.9205
	RMSE	0.0766	0.0765	0.0785
	MAE	0.0662	0.0647	0.0656
Fourth cutting	R^2	0.8582	0.9206	0.9375
	RMSE	0.1044	0.0782	0.0693
	MAE	0.0829	0.0671	0.0552

temperatures in the region. Alfalfa growth is highly sensitive to temperature, leading to substantial variability in LAI across different growing years. This observation aligns with the studies by Ren et al. (2021) and Kim et al. (2023). These factors also contribute to the decreased accuracy in fitting non-linear models.

Different alfalfa cuttings exhibit significantly different growth rates, thus the use of nonlinear models significantly improves the accuracy of fitting the leaf area index (LAI) for different cuttings. The LAI of the fourth cutting shows a declining trend in the later stages of growth, consistent with the findings of Bai and Bao (2002) The Logistic and Gompertz models effectively captures this changing trend. We observe that the baseline time-series models (LSTM, BiLSTM) have the poorest predictive accuracy. This is attributed to the lagged effect in predictions and the presence of breakpoints in the LAI time-series data, consistent with the findings of Samanta et al. (2020). We use the MOSUM method to achieve mutation point detection and eliminate training batches containing mutation points through predefined functions, avoiding damage to the original dataset and the impact of data mutations on prediction accuracy. Despite the predictive accuracy of three nonlinear models for alfalfa LAI across different cuttings has significantly improved, the differences in growth environments across different years still constrain the predictive performance of nonlinear models, especially concerning temperature and soil moisture, consistent with the findings of Yang et al. (2024b) and Su et al. (2015).

TABLE 5 Change of leaf area index with growing days.

Non-liner model	Cutting time	a	b	c	d
Logistic	First cutting	2.13	0.02	-0.004	—
	Second cutting	2.42	-0.04	-0.004	—
	Third cutting	3.38	-0.19	0.001	—
	Fourth cutting	5.94	-0.45	0.01	—
Richards	First cutting	0.99	0.64	36.11	7.74
	Second cutting	0.98	0.44	27.78	4.48
	Third cutting	0.98	0.20	19.45	1.25
	Fourth cutting	0.86	0.36	21.17	2.44
Gompertz	First cutting	1.1653	12.9047	0.1768	—
	Second cutting	0.9986	11.6963	0.1919	—
	Third cutting	0.9889	12.3895	0.1152	—
	Fourth cutting	0.9756	12.4646	0.1807	—

4.2 Analysis of Model Complexity and Applicability

In the prediction of alfalfa leaf area index (LAI), deep learning models and nonlinear models each have their advantages and disadvantages. Deep learning models can handle multiple features, including effective accumulated temperature, growth days, meteorological data, soil moisture, and leaf area index, making them suitable for complex environments. However, they have high computational complexity $O(T\times n\times m^2)$ and space complexity $O(n\times m+m^2)$, with a large number of parameters (for instance, BiLSTM can reach $8\times(m\times d+m^2)$), leading to significant memory usage (Shah and Bhavsar, 2022; Maji and Mullins, 2018). Here, n represents the sequence length (i.e., number of time steps), m represents the number of hidden units, T represents the number of training iterations, and d represents the dimensionality of the input features (such as the total number of effective accumulated temperature, growth days, meteorological data, soil moisture, and leaf area index). In contrast, nonlinear models rely only on growth days and leaf area index, resulting in low complexity (time complexity $O(T)$ and space complexity $O(n)$, with fewer parameters (usually 3-5), and high computational efficiency (De Cock et al., 2017). For example, when processing thousands of samples, the training time for the Logistic model is only a few

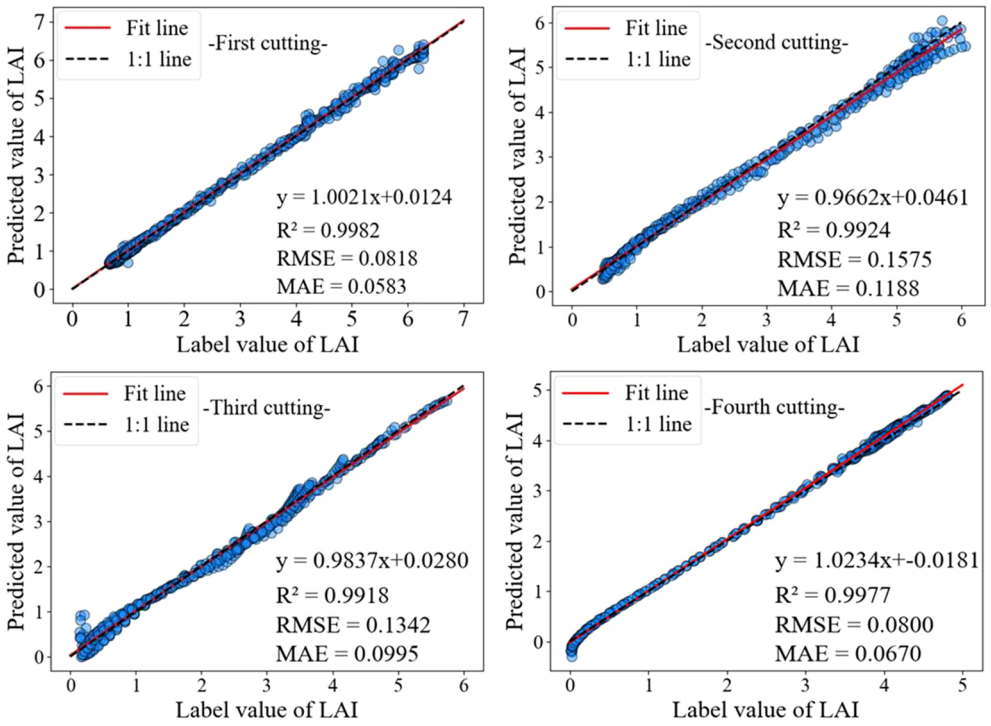


FIGURE 9
LAI prediction results in different crop cycles.

seconds, and the memory usage is less than one-tenth of that of deep learning models (Banerjee, 2007).

Deep learning models typically exhibit superior suitability for supporting offline predictions compared to nonlinear models (Jin et al., 2022). This preference stems from the inherent characteristics of deep learning architectures, which can leverage pre-trained weight parameters for offline inference without necessitating real-time computations or extensive computational resources. Once the training phase is completed, deep learning models can be seamlessly deployed on relatively modest hardware setups for prediction tasks, rendering them particularly advantageous in scenarios characterized by resource constraints or where continuous online computation is impractical. Conversely, nonlinear models often entail recalculating parameters with each prediction, making them better suited for online prediction scenarios or applications requiring real-time computations (Kalhor et al., 2010). This distinction underscores the versatility and efficiency of deep learning models in facilitating offline predictions, thereby catering

TABLE 7 Comparison of the precision of four LAI prediction results in different cutting.

Cutting time	Metrics	LSTM	BiLSTM	MLSTM	MBiLSTM	TMEAD-BiLSTM
First cutting	R^2	0.8733	0.9473	0.9857	0.9925	0.9982
	RMSE	0.6287	0.4295	0.2189	0.1638	0.0818
	MAE	0.3307	0.1864	0.1716	0.1235	0.0583
Second cutting	R^2	0.8342	0.8804	0.9552	0.9766	0.9924
	RMSE	0.6809	0.6173	0.3622	0.2678	0.1575
	MAE	0.3081	0.2699	0.2602	0.2080	0.1188
Third cutting	R^2	0.8409	0.8916	0.9633	0.9786	0.9918
	RMSE	0.5549	0.4724	0.2728	0.2127	0.1342
	MAE	0.2706	0.2249	0.2260	0.1723	0.0995
Fourth cutting	R^2	0.8843	0.9508	0.9818	0.9878	0.9977
	RMSE	0.4948	0.3566	0.2008	0.1726	0.0800
	MAE	0.2063	0.1837	0.1710	0.1458	0.0670

to a diverse array of research and practical applications within various scientific domains. Deep learning models can integrate multiple environmental factors for LAI prediction, including water and nitrogen treatment, soil moisture, and meteorological data. This ability significantly improves prediction accuracy, but also increases computational load and parameter complexity (Justus et al., 2018). In addition, deep learning models are suitable for predicting leaf area index (LAI) in different ecological environments due to their powerful feature extraction ability and the advantage of cross regional prediction through transfer learning. They can quickly deploy data from the source region in new areas and achieve good prediction results (Zeng et al., 2020). Although non-linear models perform well in specific regions, their applicability is mainly limited to local data due to the lack of generalization ability and adaptability to cross regional data, making it difficult to effectively meet the needs of cross regional prediction. However, the training of deep learning models requires the participation of large datasets to effectively capture complex patterns and interactions in the data (Najafabadi et al., 2015). The implementation of deep learning models may be challenging in resource constrained environments or with limited data availability (Liu et al., 2024). In this case, traditional nonlinear models provide a practical solution that balances the trade-off between computational efficiency and predictive performance.

5 Conclusion

To overcome the limitations of non-linear models in incorporating environmental factors and the inefficiency of LSTM in handling abrupt changes in LAI data, we proposed a TMEAD-BiLSTM model. This model was compared with classical non-linear models such as Logistic, Gompertz, and Richards. The TMEAD-BiLSTM model demonstrated superior predictive accuracy compared to the non-linear models. The strength of classic non-linear models lies in their simplicity, low computational demand, and suitability for low cost applications, particularly in resource constrained embedded devices. In contrast, the TMEAD-BiLSTM model offers high accuracy but requires high performance computing support, making it suitable for applications with high accuracy requirements and ample computational resources. Considering the potential of drones in monitoring alfalfa growth, we plan to lightweighten the model in future work and deploy it on drones to evaluate the method in real-world environments.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

SY: Methodology, Software, Validation, Writing – original draft, Writing – review & editing. YG: Funding acquisition, Resources, Writing – review & editing. JW: Data curation, Formal analysis, Visualization, Writing – review & editing. RL: Data curation, Project administration, Visualization, Writing – review & editing. LF: Formal analysis, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by National natural science foundation of China (62162052, 62262052), in part by Key Research and Development Project of Ningxia Province (2021BEB04016, 2022BDE03007), Natural Science Foundation of Ningxia (2022AAC03004).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Anastasiou, A., and Fryzlewicz, P. (2022). Detecting multiple generalized change-points by isolating single ones. *Metrika* 85, 141–174. doi: 10.1007/s00184-021-00821-6

Bai, W., and Bao, X. (2002). Simulation alfalfa growth in Wulanbuhe sandy region. *Ying Yong Sheng tai xue bao= J. Appl. Ecol.* 13, 1605–1609.

Banerjee, A. (2007). "An analysis of logistic models: Exponential family connections and online performance," in *Proceedings of the 2007 SIAM international conference on data mining* (Minneapolis, Minnesota, USA: Society for Industrial and Applied Mathematics), 204–215. doi: 10.1137/1.9781611972771

Baral, R., Lollato, R. P., Bhandari, K., and Min, D. (2022). Yield gap analysis of rainfed alfalfa in the United States. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.931403

Burczaniuk, M., and Jastrzębska, A. (2024). On the Improvements of Metaheuristic Optimization-Based Strategies for Time Series Structural Break Detection. *Informatica*, 1–33. doi: 10.15388/24-INFOR572

Brogi, C., Huisman, J. A., Herbst, M., Weihermüller, L., Klosterhalfen, A., Montzka, C., et al. (2020). Simulation of spatial variability in crop leaf area index and yield using agroecosystem modeling and geophysics-based quantitative soil information. *Vadose zone J.* 19, e20009. doi: 10.1002/vzj2.20009

Chaturvedi, R. K., Singh, S., Singh, H., and Raghubanshi, A. S. (2017). Assessment of allometric models for leaf area index estimation of Tectona grandis. *Trop. Plant Res.* 4, 274–285. doi: 10.1186/s40663-021-00284-1

Chen, Y., Wang, T., and Samworth, R. J. (2022). High-dimensional, multiscale online changepoint detection. *J. R. Stat. Soc. Ser. B: Stat. Method.* 84, 234–266. doi: 10.1111/rssb.12447

Cheng, Q., Xu, H., Fei, S., Li, Z., and Chen, Z. (2022). Estimation of maize LAI using ensemble learning and UAV multispectral imagery under different water and fertilizer treatments. *Agriculture* 12, 1267. doi: 10.3390/agriculture12081267

De Cock, M., Dowsley, R., Horst, C., Katti, R., Nascimento, A. C., Poon, W. S., et al. (2017). Efficient and private scoring of decision trees, support vector machines and logistic regression models based on pre-computation. *IEEE Trans. Dependable Secure Computing* 16, 217–230. doi: 10.1109/TDSC.2017.2679189

Desai, R. J., Wang, S. V., Vaduganathan, M., Evers, T., and Schneeweiss, S. (2020). Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes. *JAMA network Open* 3, e1918962. doi: 10.1001/jamanetworkopen.2019.18962

Duran, O., and Ünal, D. (2024). The proposed modified schnute model. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi* 28, 89–95. doi: 10.3390/en14020258

Eichinger, B., and Kirch, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli* 24, 526–564. doi: 10.3150/16-BEJ887

Feng, M., Yu, C., Lin, L., Wu, D., Song, R., and Liu, H. (2016). Effects of water and nitrogen fertilizer on biomass distribution and water use efficiency of alfalfa (Medicago sativa) in Hexi Corridor. *Chin. J. Eco-Agriculture* 24, 1623–1632. doi: 10.13930/j.cnki.cjea.160449

Guo, Y., Liu, Y., Wang, Q., Su, L., Zhang, J., and Wei, K. (2022). Regional growth model for summer maize based on a logistic model: Case study in China. *Int. J. Agric. Biol. Eng.* 15, 41–55. doi: 10.25165/j.ijabe.20221505.6584

Jin, H., Li, Y., Wang, B., Yang, B., Jin, H., and Cao, Y. (2022). Adaptive forecasting of wind power based on selective ensemble of offline global and online local learning. *Energy Conversion Manage.* 271, 116296. doi: 10.1016/j.enconman.2022.116296

Justus, D., Brennan, J., Bonner, S., and McGough, A. S. (2018). "Predicting the computational cost of deep learning models," in *2018 IEEE international conference on big data (Big Data)* (Seattle, WA, USA: IEEE), 3873–3882. doi: 10.1109/BigData.2018.8622396

Kaiwen, G., Zisong, X., Yuze, H., Qi, S., Yue, W., Yanhui, C., et al. (2020). Effects of salt concentration, pH, and their interaction on plant growth, nutrient uptake, and photochemistry of alfalfa (Medicago sativa) leaves. *Plant Signaling Behav.* 15, 1832373. doi: 10.1080/15592324.2020.1832373

Kalhor, A., Araabi, B. N., and Lucas, C. (2010). An online predictor model as adaptive habitually linear and transiently nonlinear model. *Evolving Syst.* 1, 29–41. doi: 10.1007/s12530-010-9004-z

Karadavut, U., Palta, Ç., Kökten, K., and Bakoğlu, A. (2010). Comparative study on some non-linear growth models for describing leaf growth of maize. *Int. J. Agric. Biol.* 12, 227–230. doi: 10.1186/s13717-021-00285-6

Kim, J. Y., Han, K. J., Sung, K. I., Kim, B. W., and Kim, M. (2023). Assessment of growing condition variables on alfalfa productivity. *J. Anim. Sci. Technol.* 65, 939. doi: 10.5187/jast.2023.e14

Liguori, A., Markovic, R., Dam, T. T. H., Frisch, J., van Treeck, C., and Causone, F. (2021). Indoor environment data time-series reconstruction using autoencoder neural networks. *Building Environ.* 191, 107623. doi: 10.1016/j.buildenv.2021.107623

Lin, S., Deng, M., Wei, K., Wang, Q., and Su, L. (2023). A new regional cotton growth model based on reference crop evapotranspiration for predicting growth processes. *Sci. Rep.* 13, 7368. doi: 10.1038/s41598-023-34552-7

Liu, H. I., Galindo, M., Xie, H., Wong, L. K., Shuai, H. H., Li, Y. H., et al. (2024). Lightweight deep learning for resource-constrained environments: A survey. *ACM Comput. Surv.* 56, 1–42. doi: 10.1145/36572

Liu, M., Wang, Z., Mu, L., Xu, R., and Yang, H. (2021). Effect of regulated deficit irrigation on alfalfa performance under two irrigation systems in the inland arid area of midwestern China. *Agric. Water Manage.* 248, 106764. doi: 10.1016/j.agwat.2021.106764

Liu, Y., Su, L., Wang, Q., Zhang, J., Shan, Y., and Deng, M. (2020). Comprehensive and quantitative analysis of growth characteristics of winter wheat in China based on growing degree days. *Adv. Agron.* 159, 237–273. doi: 10.1016/bs.agron.2019.07.007

Lloveras, J., Ferran, J., Alvarez, A., and Torres, L. (1998). Harvest management effects on alfalfa (Medicago sativaL.) production and quality in Mediterranean areas. *Grass forage Sci.* 53, 88–92. doi: 10.1046/j.1365-2494.1998.00100.x

Ma, H., Jiang, P., Zhang, X., Ma, W., Cai, Z., and Sun, Q. (2024). Effects of nitrogen fertilization combined with subsurface irrigation on alfalfa yield, water and nitrogen use efficiency, quality, and economic benefits. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1339417

Maji, P., and Mullins, R. (2018). On the reduction of computational complexity of deep convolutional neural networks. *Entropy* 20, 305. doi: 10.3390/e20040305

Masini, R. P., Medeiros, M. C., and Mendes, E. F. (2023). Machine learning advances for time series forecasting. *J. economic surveys* 37, 76–111. doi: 10.1111/joes.12429

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., and Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *J. big Data* 2, 1–21. doi: 10.1186/s40537-014-0007-7

Nhu, V. H., Shirzadi, A., Shahabi, H., Singh, S. K., Al-Ansari, N., Clague, J. J., et al. (2020). Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms. *Int. J. Environ. Res. Public Health* 17, 2749. doi: 10.3390/ijerph17082749

Qian, Z. H., Han, B. F., Liu, Z. T., Cai, W., Fu, B. Z., and Ma, H. B. (2020). Effects of infiltration irrigation on growth characters and water use efficiency of alfalfa in the ningxia yellow river diversion irrigation area. *Acta Pratacult. Sin.* 29, 147. doi: 10.11686/Cyxb2019254

Ren, L., Bennett, J. A., Coulman, B., Liu, J., and Biligetu, B. (2021). Forage yield trend of alfalfa cultivars in the Canadian prairies and its relation to environmental factors and harvest management. *Grass forage Sci.* 76, 390–399. doi: 10.1111/gfs.12513

Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J. M., et al. (2020). Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, China: February 13–23, 2020. *J. Clin. Med.* 9, 596. doi: 10.3390/jcm9020596

Samanta, S., Pratama, M., Sundaram, S., and Srikanth, N. (2020). "A dual network solution (DNS) for lag-free time series forecasting," in *2020 international joint conference on neural networks (IJCNN)* (Glasgow, UK: IEEE), 1–8. doi: 10.1109/IJCNN48605.2020.9207022

Shah, B., and Bhavsar, H. (2022). Time complexity in deep learning models. *Proc. Comput. Sci.* 215, 202–210. doi: 10.1016/j.procs.2022.12.023

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena* 404, 132306. doi: 10.1016/j.physd.2019.132306

Siami-Namini, S., Tavakoli, N., and Namin, A. S. (2019). "The performance of LSTM and BiLSTM in forecasting time series," in *2019 IEEE International conference on big data (Big Data)* (Los Angeles, CA, USA: IEEE), 3285–3292. doi: 10.1109/BigData47090.2019.9005997

Su, L., Tao, W., Sun, Y., Shan, Y., and Wang, Q. (2022). Mathematical models of leaf area index and yield for grapevines grown in the turpan area, xinjiang, China. *Agronomy* 12, 988. doi: 10.3390/agronomy12050988

Su, L., Wang, Q., Wang, C., and Shan, Y. (2015). Simulation models of leaf area index and yield for cotton grown with different soil conditioners. *PloS One* 10, e0141835. doi: 10.1371/journal.pone.0141835

Tooley, E. G., Nippert, J. B., and Ratajczak, Z. (2024). Evaluating methods for measuring the leaf area index of encroaching shrubs in grasslands: From leaves to optical methods, 3-D scanning, and airborne observation. *Agric. For. Meteorology* 349, 109964. doi: 10.1016/j.agrformet.2024.109964

Tripathi, A. M., Pohanková, E., Fischer, M., Orság, M., Trnka, M., Klem, K., et al. (2018). The evaluation of radiation use efficiency and leaf area index development for the estimation of biomass accumulation in short rotation poplar and annual field crops. *Forests* 9, 168. doi: 10.3390/f9040168

Vaghi, C., Rodallec, A., Fanciullino, R., Ciccolini, J., Mochel, J. P., Mastri, M., et al. (2020). Population modeling of tumor growth curves and the reduced Gompertz model improve prediction of the age of experimental tumors. *PloS Comput. Biol.* 16, e1007178. doi: 10.1371/journal.pcbi.1007178

Wang, W., Liang, Y., Ru, Z., Guo, H., and Zhao, B. (2023). World forage import market: competitive structure and market forces. *Agriculture* 13, 1695. doi: 10.3390/agriculture13091695

Wang, T., and Zhang, W. H. (2023). Priorities for the development of alfalfa pasture in northern China. *Fundam. Res.* 3, 225–228. doi: 10.1016/j.fmre.2022.04.017

Yang, S., Ge, Y., Wang, J., Liu, R., Tang, D., Li, A., et al. (2024a). A dataset for estimating alfalfa leaf area and predicting leaf area index. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1290920

Yang, T., Zhao, J., and Fu, Q. (2024b). Quantitative relationship of plant height and leaf area index of spring maize under different water and nitrogen treatments based on effective accumulated temperature. *Agronomy* 14, 1018. doi: 10.3390/agronomy14051018

Ze-hao, L., Qi-ming, Q., Tian-yuan, Z., and Wei, X. (2020). Prediction of continuous timeseries leaf area index based on long short-term memory network: A case study of winterwheat. *Spectrosc. Spectral Anal.* 40, 898–904. doi: 10.3964/j.issn.1000-0593(2020)03-0898-07

Zeng, Q., Sun, Q., Chen, G., Duan, H., Li, C., and Song, G. (2020). Traffic prediction of wireless cellular networks based on deep transfer learning and cross-domain data. *IEEE Access* 8, 172387–172397. doi: 10.1109/ACCESS.2020.3025210

Zhang, H., An, D., Cao, Y., Tian, Y., and He, J. (2021). Modeling the methane production kinetics of anaerobic co-digestion of agricultural wastes using sigmoidal functions. *Energies* 14, 258. doi: 10.3390/en14020258

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 19 November 2024
DOI 10.3389/fpls.2024.1495222

Check for updates

OPEN ACCESS

EDITED BY
Ting Sun,
China Jiliang University, China

REVIEWED BY
Hongwei Sun,
Hangzhou Dianzi University, China
Yaohua Hu,
Zhejiang Agriculture and Forestry University, China

*CORRESPONDENCE
Xin Jin
jx.771@163.com
Shunhao Qing
18392163586@163.com

RECEIVED 12 September 2024
ACCEPTED 30 October 2024
PUBLISHED 19 November 2024

CITATION
Qiu Z, Wang F, Wang W, Li T, Jin X, Qing S and Shi Y (2024) YOLO-SDL: a lightweight wheat grain detection technology based on an improved YOLOv8n model.
Front. Plant Sci. 15:1495222.
doi: 10.3389/fpls.2024.1495222

COPYRIGHT
© 2024 Qiu, Wang, Wang, Li, Jin, Qing and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

YOLO-SDL: a lightweight wheat grain detection technology based on an improved YOLOv8n model

Zhaomei Qiu[1], Fei Wang[1], Weili Wang[1], Tingting Li[1], Xin Jin[1,2]*, Shunhao Qing[1]* and Yi Shi[1]

[1]College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang, Henan, China, [2]Science and Technology Innovation Center for Completed Set Equipment, Longmen Laboratory, Luoyang, Henan, China

Wheat, being a crucial global food crop, holds immense significance for food safety and agricultural economic stability, as the quality and condition of its grains are critical factors. Traditional methods of wheat grain detection are inefficient, and the advancements in deep learning offer a novel solution for fast and accurate grain recognition. This study proposes an improved deep learning model based on YOLOv8n, referred to as YOLO-SDL, aiming to achieve efficient wheat grain detection. A high-quality wheat grain dataset was first constructed, including images of perfect, germinated, diseased, and damaged grains. Multiple data augmentation techniques were employed to enhance the dataset's complexity and diversity. The YOLO-SDL model incorporates the ShuffleNetV2 architecture in its backbone and combines depthwise separable convolutions (DWConv) with the large separable kernel attention (LSKA) mechanism in its neck structure, significantly improving detection speed and accuracy while ensuring the model remains lightweight. The results indicate that YOLO-SDL achieves superior performance in wheat grain detection, balancing lightweight design and performance optimization. The model achieved a P of 0.942, R of 0.903, mAP50 of 0.965, and mAP50-95 of 0.859, with low computational complexity, making it suitable for resource-constrained environments. These findings demonstrate the efficiency of the ShuffleNetV2, DWConv, and LSKA structures. The proposed YOLO-SDL model provides a new technical solution for agricultural automation and serves as a reliable reference for detecting other crops.

KEYWORDS

YOLOv8, wheat grain, lightweight model, object detection, computer vision

1 Introduction

Wheat is one of the most widely cultivated food crops globally, and its grain yield and quality directly impact food security, making it a cornerstone of agricultural economies (Guo et al., 2023; Peng et al., 2022). During growth, harvesting, and storage, wheat grains can be affected by environmental factors or improper handling, leading to germination, damage, or disease, which compromise their quality (Ma et al., 2021). Thus, precise identification and detection of wheat grains, particularly distinguishing between different types of grains, is crucial for ensuring food safety and promoting sustainable agriculture (Que et al., 2023; Zhou et al., 2020). Moreover, grain detection plays a vital role in wheat breeding, enhancing breeding efficiency and aiding in the development of high-quality varieties (Misra et al., 2022).

Manual methods of detecting wheat grains suffer from inefficiency and subjectivity, failing to meet the demands of modern agriculture (Gao et al., 2022; Li et al., 2024c; Zhao et al., 2022). While traditional machine vision technology, which relies on computer-based image recognition and analysis, improves detection efficiency and accuracy, its generalization capability is limited. It tends to perform well on specific tasks but struggles to adapt to new ones, requiring expert intervention for feature extraction (Ji et al., 2022).

Recently, deep learning has made significant strides in object detection, particularly with deep neural networks excelling in image recognition and classification tasks (Feng et al., 2023; Han et al., 2023). Deep learning models for object detection are broadly categorized into two-stage and one-stage models. Two-stage models, such as the R-CNN series and Mask R-CNN, first generate candidate regions and then classify and localize these regions for accurate object detection (Fang et al., 2023; Li et al., 2024a). Li et al. (2022a) used a Faster R-CNN model to detect wheat spike in RGB images, achieving rapid and accurate detection with an average accuracy of 86.7%. Similarly, Ye et al. (2023) applied the Mask R-CNN model to visible-light images captured by uncrewed aerial vehicle to achieve rapid and accurate detection of individual cabbage plants. The study results demonstrated the superior performance of the Mask R-CNN model, with a detection accuracy of 99.5% and an overall average F1 score of 97.63%. Wu et al. (2020) combined transfer learning with the Faster R-CNN model for the detection and counting of wheat grains under complex environmental conditions. The results showed that the mean average precision (mAP) of the Faster R-CNN model was 0.91, and the model exhibited strong robustness. Although two-stage models exhibit high accuracy, they tend to be slower (Li et al., 2022b). In contrast, one-stage models like YOLO and SSD are widely favored for their speed and practicality, directly extracting image features to predict object categories and locations (Bacea and Oniga, 2023; Wan et al., 2024).

The YOLO series, a quintessential one-stage model, stands out for its exceptional performance and efficiency in object detection tasks (Wu et al., 2023; Zhang et al., 2024). However, while YOLO models are fast, there is still room to improve their detection accuracy. To optimize the YOLO algorithm, researchers have worked to enhance its accuracy and computational efficiency while reducing complexity and parameter count, aiming for a lightweight yet efficient model. For example, Ma et al. (2024) proposed an improved YOLOv8 model for wheat seed detection, incorporating shared convolutional layers and vision transformer with deformable attention mechanism, which resulted in a lightweight and efficient model with a mAP of 99.3%, a 16.8% improvement over YOLOv8. Similarly, Yao et al. (2024) improved the YOLOv7 model by integrating an efficient channel attention mechanism and a bi-directional feature pyramid network for detecting the germination rate of wild rice seeds, achieving detection accuracies of 94% and 98.2% in hydroponic and petri dish environments, respectively. Qing et al. (2024) proposed an enhanced YOLO-FastestV2 model for the detection of wheat spike counts. The model incorporated three attention mechanisms and a SimConv structure to enhance the model's feature extraction capabilities and detection accuracy. The study demonstrated that the improved model achieved a precision (P) of 83.91%, recall (R) of 78.35%, average precision of 81.52%, and F1 score of 81.03%, indicating the overall best performance.

To enhance the accuracy and efficiency of wheat grain detection, this study proposes an improved YOLOv8n model, YOLO-SDL. The model adopts the ShuffleNetV2 architecture in its backbone, known for its computational efficiency, which not only boosts image processing speed but also significantly enhances detection accuracy. Additionally, depthwise separable convolutions (DWConv) replace the original convolutional modules in the neck of YOLOv8n, further reducing computational load, making the model more suitable for resource-constrained conditions. The large separable kernel attention (LSKA) mechanism is also integrated into the neck structure to improve the model's ability to capture key object features, thereby enhancing detection precision. The contributions of this study are as follows: (1) constructing a high-quality dataset of wheat grains with various conditions, (2) conducting ablation experiments to validate the effectiveness of each improvement in the YOLO-SDL model, and (3) comparing the performance of YOLO-SDL with YOLOv5n, YOLOv6n, YOLOv8n, and YOLOv10n.

2 Materials and methods

2.1 Wheat grain dataset

This study constructed a comprehensive and diverse wheat grain dataset, including images of perfect, germinated, diseased, and damaged grains, captured using high-resolution imaging equipment to ensure quality. After capturing the images, detailed bounding box annotations were made using the LabelImg software to ensure accurate identification of different grain conditions during model training. The dataset contains 9400 perfect grains, 7260 germinated grains, 835 diseased grains, and 1890 damaged grains, accounting for 48.49%, 37.45%, 4.31%, and 9.75% of the total, respectively. To further enhance the complexity and diversity of the dataset, various data augmentation techniques were employed, such as random scaling, rotation, flipping, noise addition, and

adjustment of lighting intensity. These methods expanded the dataset's size and richness, improving the model's stability and generalization ability in recognizing wheat grains under different environments and conditions. The final dataset comprises 1170 images, divided into training, testing, and validation sets in an 8:1:1 ratio. Examples of wheat grain images are shown in Figure 1.

2.2 Improvement strategies

2.2.1 YOLO-SDL model

The YOLOv8n algorithm builds upon the advantages of the YOLO series in terms of both accuracy and speed, with further optimizations to enhance performance. The YOLOv8n model introduces the C2f structure, which integrates feature maps from various depths, enabling the network to leverage both shallow and deep features simultaneously. This significantly enhances the model's ability to recognize and capture object features (Jiang et al., 2024; Lu et al., 2024). The cross stage partial connections in the C2f structure optimize information flow through the network. By partially connecting different stages, the model not only improves feature representation but also strengthens its ability to detect objects in complex scenes. The neck of the YOLOv8n model employs the path aggregation network - feature pyramid network structure, which builds multi-scale feature fusion paths, allowing the model to handle objects of various scales more flexibly and efficiently. The detection head uses a decoupled structure that separates the classification and localization tasks, enabling independent optimization of both and improving detection accuracy and efficiency (Chen et al., 2024).

To further enhance detection accuracy while reducing computational complexity and parameter count, making the model more lightweight and efficient, this study proposes improvements to the YOLOv8n model. The structure of the improved model is shown in Figure 2. First, the ShuffleNetV2 architecture is introduced into the backbone of YOLOv8n, ensuring high precision while enabling fast object detection with lower computational cost. This is particularly crucial for tasks such as wheat grain detection, which often need to be conducted in environments like fields or warehouses where computational resources are limited. Additionally, we integrate DWConv and LSKA mechanisms into the neck of YOLOv8n. The DWConv structure significantly reduces the computational load, making the model more suitable for deployment on resource-constrained devices. The LSKA mechanism provides a more flexible feature extraction process, effectively enhancing the model's detection accuracy. This aids the model in better capturing detailed characteristics of wheat grain under various environments to achieve accurate detection. By integrating improvements such as ShuffleNetV2 and DWConv, the model proposed in this study achieves structural lightweight and efficiency, and is functionally particularly suited for wheat grain detection tasks. It is capable of maintaining high detection accuracy while adapting to resource-constrained environments and meeting the demand for processing large volumes of image data.

2.2.2 DWConv

DWConv reduces the number of parameters in the model by decomposing a standard convolution into two processes: depthwise convolution and pointwise convolution, significantly lowering

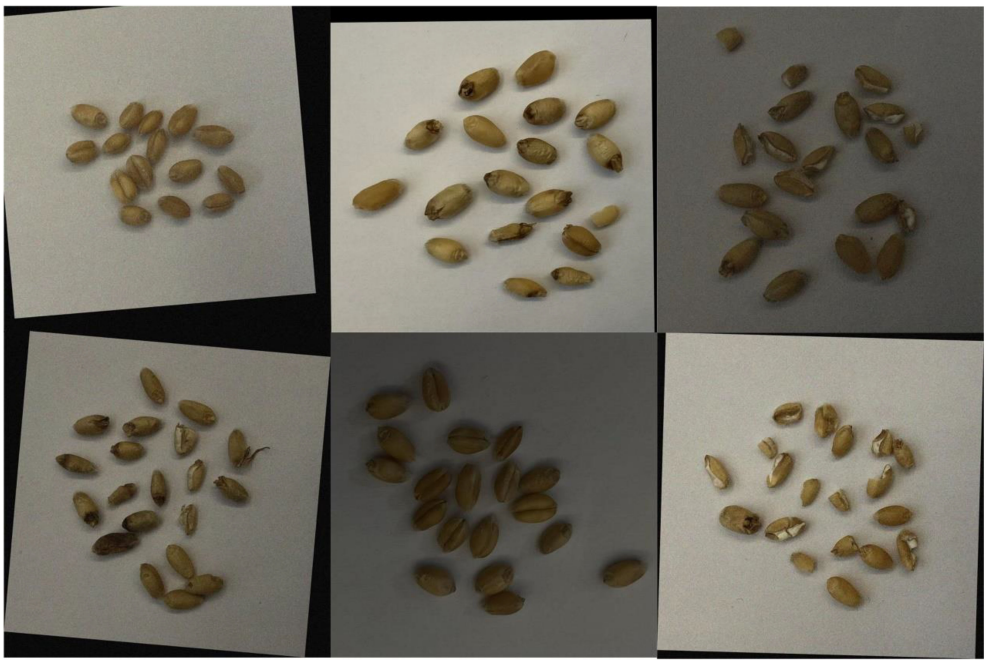


FIGURE 1
Sample images from the wheat grain dataset.

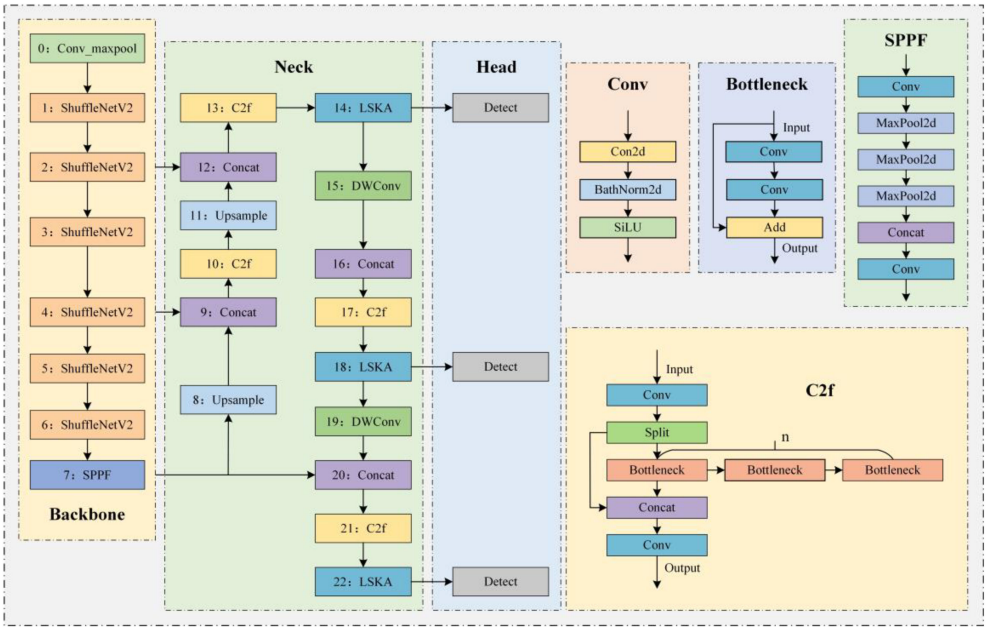


FIGURE 2
Structure of the YOLO-SDL model.

computational complexity (Balasundaram et al., 2023). During the depthwise convolution stage, each channel undergoes independent convolution operations, generating feature maps corresponding to the number of channels. This process enables the model to precisely recognize and capture the local features of each channel without a significant increase in the number of parameters. Pointwise convolution, which uses a 1×1 convolution kernel, integrates the feature maps produced by the depthwise convolution across channels, generating the final output feature map. The DWConv structure preserves the model's ability to accurately capture spatial features while dramatically reducing its computational complexity and parameter count, resulting in a more lightweight model (Li et al., 2024b; Shi et al., 2024). DWConv replaces the original convolution structure in the YOLOv8n model, reducing model complexity and achieving further lightweight optimization. Figure 3 illustrates the DWConv structure.

2.2.3 ShuffleNetV2

ShuffleNetV2 is an efficient and lightweight convolutional neural network architecture that balances model complexity with accuracy (Thakuria and Erkinbaev, 2024). It employs group convolution to divide the input channels into groups, where

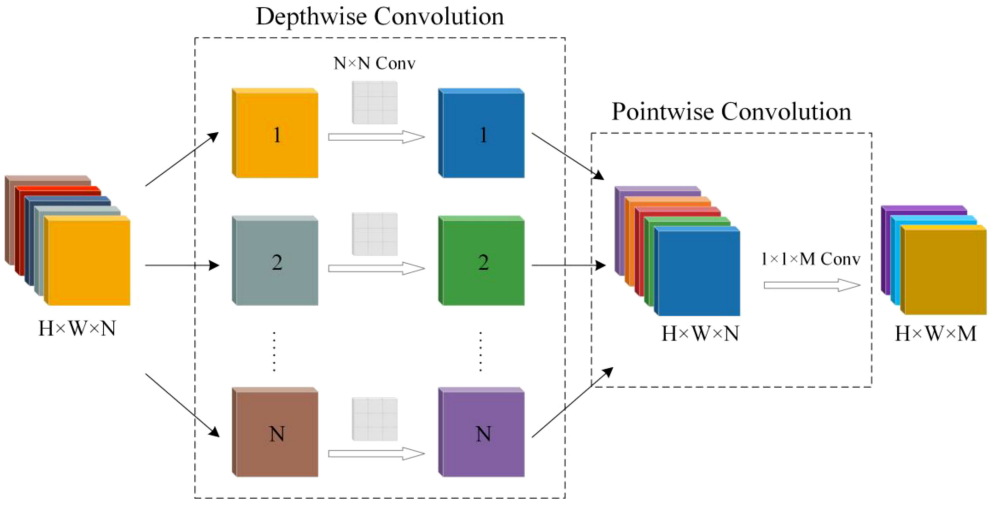


FIGURE 3
Structure of DWConv.

independent convolution operations are performed within each group. This technique significantly reduces the number of parameters in the convolutional layers, thereby lowering the computational cost of the model. ShuffleNetV2 also incorporates an innovative channel shuffle mechanism, which enables effective feature fusion after the grouping process, enhancing the model's ability to integrate features and improving detection accuracy. Moreover, the integration of DWConv further reduces the computational complexity of the model. Due to its lightweight and efficient nature, ShuffleNetV2 offers an effective solution for resource-constrained environments. ShuffleNetV2 exhibits a reduced computational parameter count relative to EfficientNet-B0, which is the smallest model within the EfficientNet series (Ding et al., 2023). It also demonstrates superior performance in both speed and accuracy compared to MobileNetV2, providing faster inference capabilities. These attributes render ShuffleNetV2 a highly suitable choice for applications within resource-limited environments (Ma et al., 2018). The channel shuffle mechanism and the design of the DWConv structure in ShuffleNetV2 provide better robustness when dealing with small objects, which is crucial for wheat grain identification and detection tasks. In this study, the ShuffleNetV2 architecture was applied to the backbone of the YOLO-SDL model, enabling it to adapt to environments with limited resources while maintaining fast and accurate object detection capabilities. Figure 4 shows the structure of ShuffleNetV2.

2.2.4 LSKA

LSKA is an innovative attention mechanism that decomposes a 2D convolutional kernel in the deep convolutional layers into two 1D convolutional kernels, performing horizontal and vertical convolution operations separately. This reduces the number of parameters and computational load while enhancing the model's ability to capture image features (Lau et al., 2024). LSKA generates an attention map through a 1×1 convolution layer based on the 1D convolutions, which is used to adjust and enhance key information in the feature maps. The attention map is then multiplied pointwise with the original feature map to achieve adaptive optimization of the features (Zhong et al., 2024). Compared to traditional large kernel attention mechanisms, LSKA overcomes the computational and memory efficiency challenges associated with large convolutional kernels. It significantly reduces computational complexity, making it more efficient and flexible in object detection tasks, while providing more precise detection results. In this study, the LSKA mechanism was introduced into the neck network of the YOLOv8n model to further optimize its performance. Figure 5 shows the structure of LSKA mechanisms.

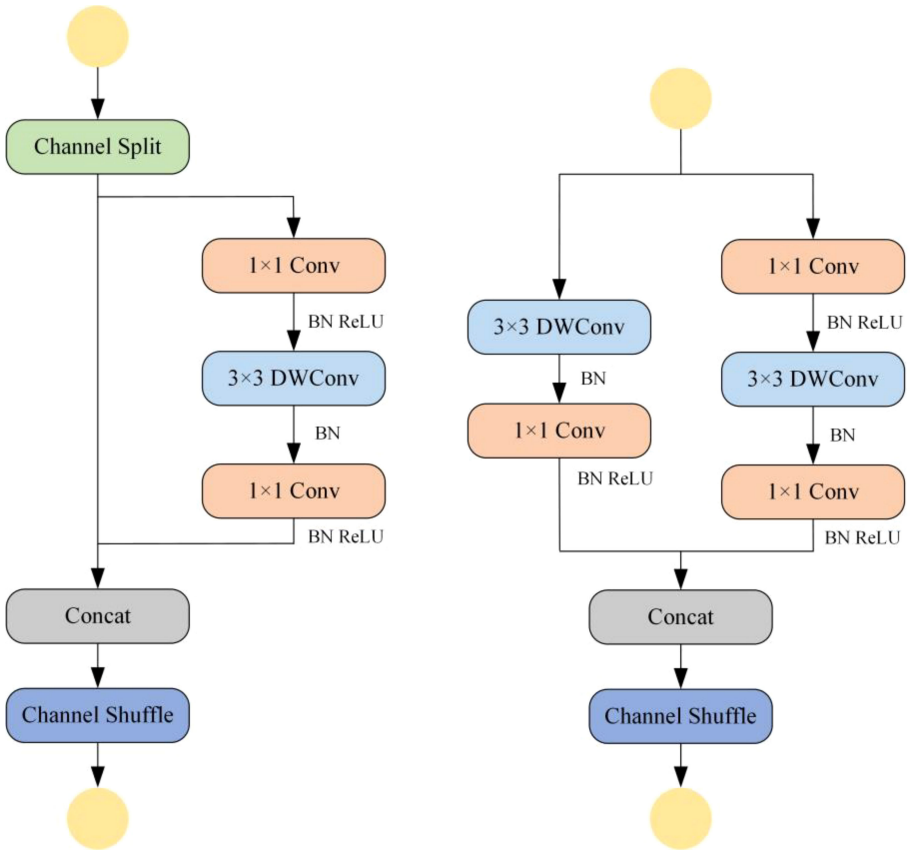


FIGURE 4
Structure of ShuffleNetV2.

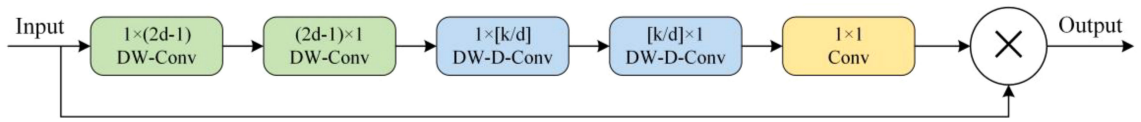


FIGURE 5
Structure of LSKA mechanisms.

2.3 Evaluation metrics

To assess the performance of different models in detecting wheat grains, this study employs P, R, mean average precision (mAP50), extended mean average precision (mAP50-95), and giga floating-point operations per second (GFLOPs) as evaluation metrics. P measures the proportion of correctly predicted positive samples out of all samples predicted as positive, as shown in Equation 1. R evaluates the proportion of actual positive samples that were correctly predicted by the model, as shown in Equation 2. mAP50 represents the mean average precision when the intersection over union (IoU) threshold is set at 50%, serving as a general indicator of the model's overall performance, as shown in Equation 3. mAP50-95 is an extended version of mAP50 that considers a range of IoU thresholds (from 50% to 95% in increments of 5%), providing a more detailed performance evaluation across various levels of IoU, as shown in Equation 4. GFLOPs quantifies the capability of performing ten billion floating-point operations per second and serves as a pivotal metric for assessing the efficiency of deep learning models. It indicates the computational velocity of the models during both training and inference; a higher GFLOPs value signifies a model's enhanced data processing capacity. As illustrated in Equation 5. The Frames per second (FPS) is an important indicator of the speed of model recognition and detection, which reflects the number of images processed by the model per second, the larger the FPS, the faster the processing speed of the model.

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$mAP_{50} = \frac{1}{m}\sum_{i=1}^{m} AP_i \tag{3}$$

$$mAP_{50-95} = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{10}\sum_{j=1}^{10} AP_{i,j} \tag{4}$$

$$GFLOPs = \frac{2*(K \times K \times C_{in} \times C_{out} \times H \times W + C_{out})}{10^9} \tag{5}$$

$$FPS = \frac{1}{Processing\ \ time\ \ per\ \ frame} \tag{6}$$

In this context, TP refers to the number of true positives, TN refers to the number of true negatives, FP refers to the number of false positives, and FN refers to the number of false negatives. m represents the number of classes in the dataset, AP_i represents the average precision for class i when the IoU threshold is 0.5, and $AP_{i,j}$ represents the average precision for class i at IoU thresholds of $0.5 + 0.05 \times (j-1)$. $K \times K$ denotes the size of the convolutional kernel, C_{in} represents the number of input channels in the convolutional layer, C_{out} represents the number of output channels in the convolutional layer, and $H \times W$ indicates the height and width of the feature map output by the convolutional layer.

3 Results

3.1 Experimental environment

The experiments were conducted on the Ubuntu 20.04 operating system, with Python 3.10 as the programming language, CUDA version 11.8, and PyTorch 2.0.1 as the deep learning framework. Jupyter was used as the IDE for developing the experiments. The CPU was an Intel(R) Xeon(R) Gold 5318Y @ 2.10GHz, and the GPU used was an NVIDIA A16 with 15GB of memory. During the network training process, this study sets the learning rate to 0.01 and the batch size to 200, with the loss function chosen as IoU. We employ a stochastic gradient descent optimizer with a momentum of 0.937, along with weight decay of 0.0005 to regularize the model and prevent overfitting. To more smoothly initiate the training process, we set up 3 warmup epochs, during which the initial momentum is set to 0.8 and the initial bias learning rate is set to 0.1. Such settings help to gradually enhance the model's stability and convergence rate at the initial stage of training.

3.2 Ablation experiments

This study utilized the YOLOv8n model as a baseline and made several improvements by incorporating ShuffleNetV2, DWConv, and LSKA structures to enhance the accuracy of wheat grain detection. Ablation experiments were conducted to analyze the performance impact of each modification, with the results summarized in Table 1. In the ablation experiments, the results of different improved models on wheat grain feature recognition are shown in Figure 6.

The baseline YOLOv8n model achieved a detection performance of P = 0.898, R = 0.903, mAP50 = 0.937, and mAP50-95 = 0.816. The model had 3006623 parameters and a GFLOP count of 8.1. After introducing the ShuffleNetV2 architecture, the YOLO-S model improved its P to 0.909 and R to

TABLE 1 Accuracy results of ablation experiments.

Model	ShuffleNetV2	DWConv	LSKA	P	R	mAP50	mAP50-95
YOLOv8n	–	–	–	0.898	0.903	0.937	0.816
YOLO-S	√	–	–	0.909	0.917	0.954	0.834
YOLO-D	–	√	–	0.904	0.904	0.944	0.823
YOLO-L	–	–	√	0.904	0.926	0.952	0.832
YOLO-SD	√	√	–	0.911	0.917	0.960	0.848
YOLO-SDL	√	√	√	0.942	0.903	0.965	0.859

√ indicates that the module was used in this set of experiments; - indicates that the module was not used in this set of experiments.

0.917, with mAP50 and mAP50-95 reaching 0.954 and 0.834, respectively. The number of parameters and GFLOPs were reduced to 1835343 and 5.2, indicating that ShuffleNetV2 not only enhanced detection performance but also achieved significant model lightweighting. By incorporating the DWConv structure, the YOLO-D model achieved P = 0.904 and R = 0.904, with mAP50 and mAP50-95 improving to 0.944 and 0.823, respectively. The parameter count was 2824031, and GFLOPs were 7.9. While DWConv contributed to improved precision and target recognition ability, it offered slightly less performance improvement compared to ShuffleNetV2, suggesting potential for further optimization. With the introduction of the LSKA mechanism, the YOLO-L model saw increased complexity, with parameters growing to 3100255 and GFLOPs to 8.3. However, detection performance improved, with P = 0.904, R = 0.926, and mAP50 and mAP50-95 reaching 0.952 and 0.832, respectively, demonstrating the significant impact of LSKA on enhancing detection accuracy.

The YOLO-SD model, combining the strengths of ShuffleNetV2 and DWConv, achieved P = 0.911, R = 0.917, mAP50 = 0.960, and mAP50-95 = 0.848, with the parameter count reduced to 1652751 and GFLOPs to 4.9. These results indicate that the model achieved a high level of performance while further reducing its complexity. Finally, the YOLO-SDL model, which integrates ShuffleNetV2, DWConv, and LSKA, achieved the best overall performance with P = 0.942, mAP50 = 0.965, and mAP50-95 = 0.859. Although recall dropped slightly to 0.903, the parameter count was 1755215, and GFLOPs were 5.1. These results demonstrate the model's ability to balance lightweight design with high accuracy and efficiency. According to the results demonstrated in Figure 6, it can be seen that the YOLO-SDL model focuses on the image region closer to the wheat grain, which has better recognition prediction effect.

This study significantly improved the accuracy of wheat grain detection by incorporating ShuffleNetV2, DWConv, and LSKA into the YOLOv8n model, achieving a balance between model lightweighting and computational efficiency. The results show

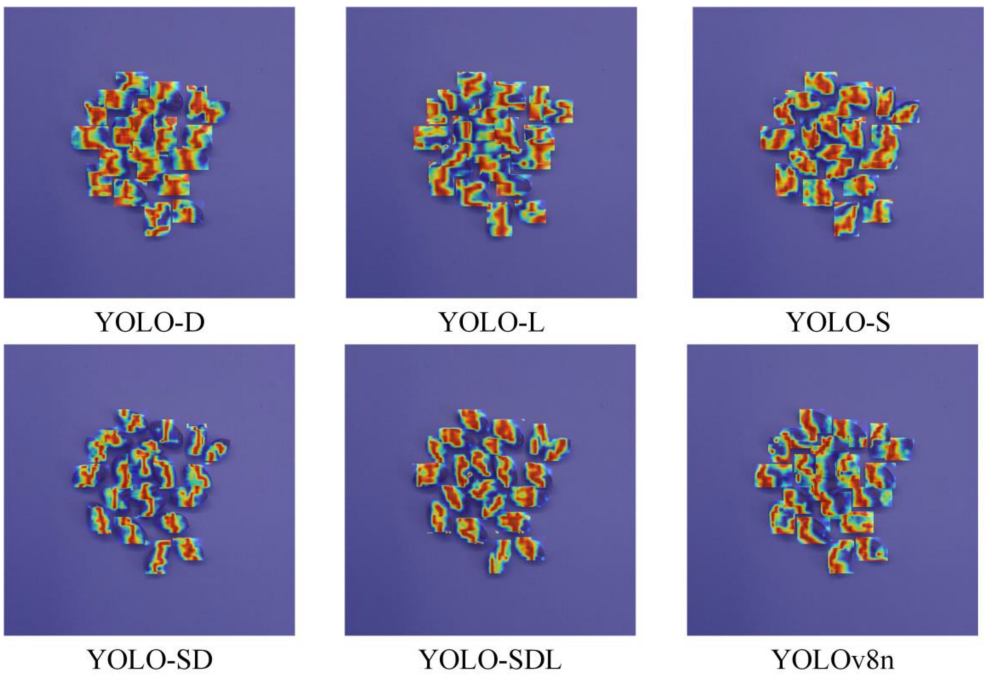


FIGURE 6
Heat map of ablation experiment results.

that the ShuffleNetV2 architecture enhanced detection accuracy while significantly reducing model parameters and computational load. The DWConv structure provided modest accuracy improvements and reduced computational costs, though it still has room for further optimization. While the LSKA structure increased model complexity, it substantially improved detection accuracy. Among the various improved models, YOLO-SDL strikes the best balance between accuracy, parameter count, and computational efficiency, making it the optimal model for wheat grain detection.

3.3 Comparison of different models

This study compares the YOLO-SDL model with three other models from the YOLO series—YOLOv5n, YOLOv6n, and YOLOv10n—evaluating their accuracy and performance in wheat grain detection. The detection accuracy of the different models is summarized in Table 2, with the accuracy comparison trends shown in Figure 7, and the detection results comparison presented in Figure 8.

As shown in Table 2, the YOLOv5n model achieved relatively high P and R, with P = 0.900 and R = 0.887. Its mAP50 reached 0.940, while mAP50-95 was 0.817. The model also demonstrated a processing speed of 156.250 FPS, indicating a balance between high accuracy and fast processing. The YOLOv6n model had slightly higher P than YOLOv5n at 0.902, with R = 0.888, mAP50 = 0.938, and mAP50-95 improving to 0.827. It showed a real-time processing advantage with a FPS of 106.383. The YOLOv8n model had slightly lower P (P = 0.898), but R increased to 0.903, with mAP50 = 0.937 and mAP50-95 = 0.816. Its processing speed was 140.845 FPS, faster than YOLOv6n. The YOLOv10n model had the lowest P (P = 0.895) and R (R = 0.887), although its mAP50 and mAP50-95 were 0.940 and 0.834, respectively. It had the highest

TABLE 2 Accuracy of detection results for different models.

Model	P	R	mAP50	mAP50-95	Parameters	GFLOPs	Speed (FPS)
YOLOv5n	0.900	0.887	0.940	0.817	2503919	7.1	156.250
YOLOv6n	0.902	0.888	0.938	0.827	4234239	11.8	106.383
YOLOv8n	0.898	0.903	0.937	0.816	3006623	8.1	140.845
YOLOv10n	0.895	0.887	0.940	0.834	2696366	8.2	175.439
YOLO-SDL	0.942	0.903	0.965	0.859	1755215	5.1	161.290

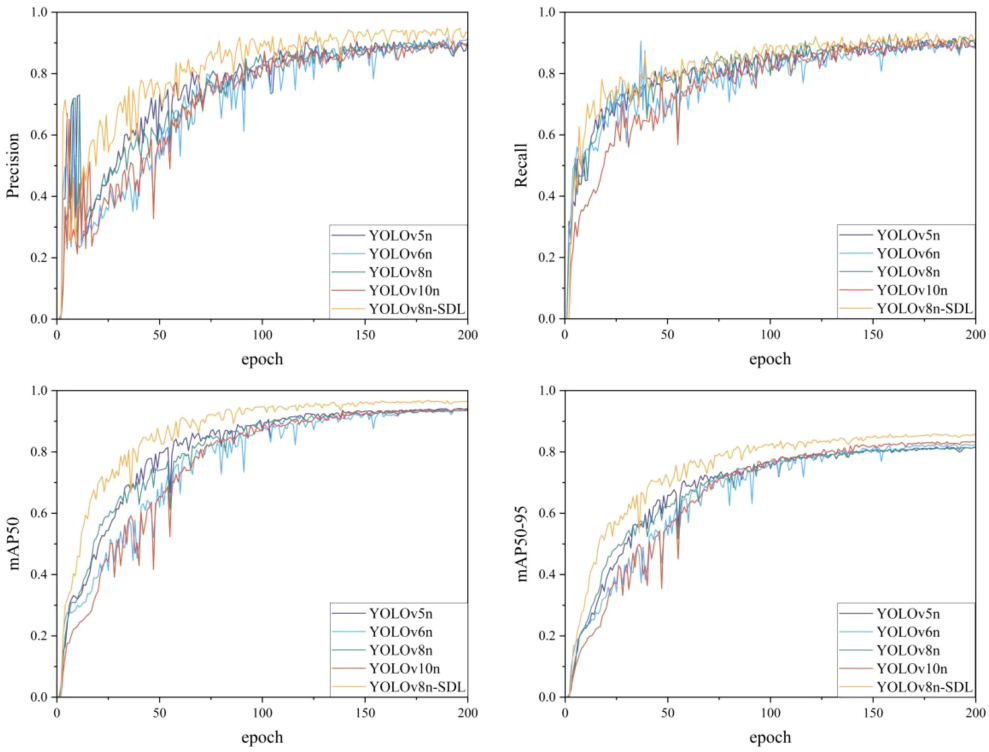


FIGURE 7
Comparison of detection accuracy variations across different models.

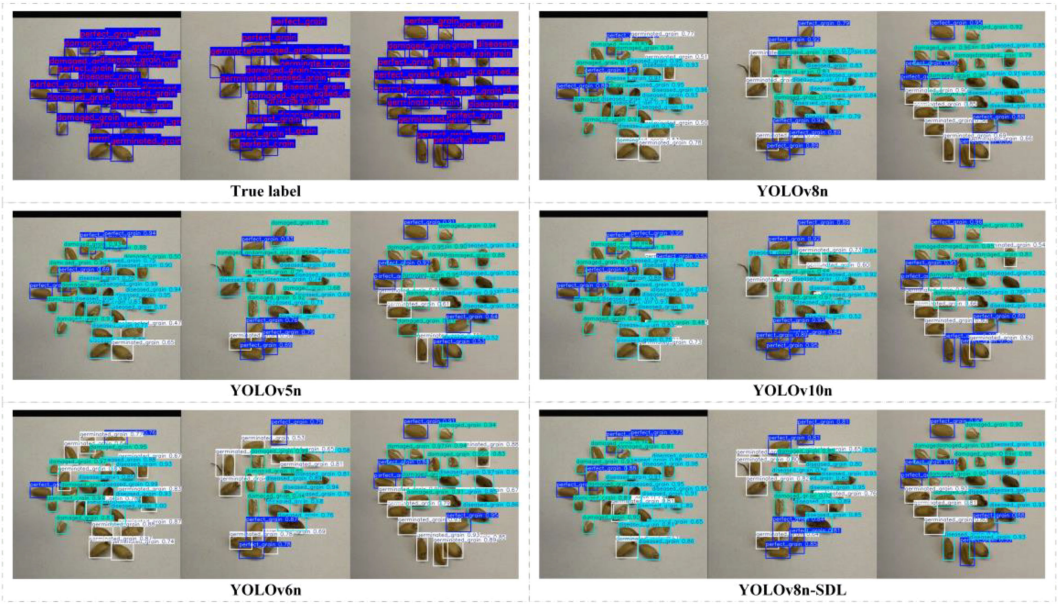


FIGURE 8
Comparison of detection results of different models.

processing speed among all models at 175.439 FPS. In contrast, the YOLO-SDL model demonstrated a significant accuracy advantage in the wheat detection task, with P reaching 0.942, R at 0.903, and mAP50 and mAP50-95 of 0.965 and 0.859, respectively, making it the most accurate model among all tested. The YOLO-SDL also had a processing speed of 161.290 FPS, second only to YOLOv10n, indicating a fast image processing capability. From the perspective of computational complexity and efficiency, YOLO-SDL was the most lightweight model, with the smallest parameter count (1755215) and GFLOPs (5.1). In comparison, YOLOv6n had the highest computational cost, with 4234239 parameters and 11.8 GFLOPs. The YOLOv5n model was less complex, with 2503919 parameters and 7.1 GFLOPs, while the YOLOv8n and YOLOv10n models had intermediate levels of complexity.

The Figure 6 shows the changes in accuracy for each model as the number of iterations increased. All models demonstrated an upward trend in accuracy over time, with the YOLO-SDL model maintaining the highest accuracy curve, consistently outperforming the other models. Additionally, as shown in Figure 7, the detection results indicate that the YOLO-SDL model was able to more accurately identify various types of wheat grains compared to the other models. In summary, the analysis demonstrates that appropriate adjustments and optimizations to the network structure can significantly enhance model performance. By integrating the strengths of ShuffleNetV2, DWConv, and LSKA, the YOLO-SDL model achieves the goal of reducing model complexity while maintaining high accuracy. This combination of lightweight design and high precision not only improves the model's adaptability and practicality but also provides new insights for the fields of agricultural automation and smart agriculture.

4 Discussion

This study introduces improvements to the YOLOv8n model by incorporating ShuffleNetV2, DWConv, and LSKA structures, with the goal of optimizing detection accuracy, reducing the model's parameter count, and lowering computational complexity. These enhancements aim to improve the model's performance in wheat grain detection tasks while achieving a balance between model lightweighting and computational efficiency. Through ablation experiments and comparative analysis with different models, this research demonstrates that the YOLO-SDL model is the optimal solution for wheat grain detection, offering high-precision results while maintaining a lightweight design.

The YOLO-SDL model leverages the ShuffleNetV2 architecture as its backbone, significantly improving detection accuracy while greatly reducing both the model's parameter count and computational load. ShuffleNetV2, an efficient lightweight network architecture, uses grouped convolution, channel shuffle mechanisms, and DWConv to capture rich feature information while minimizing computational complexity. These lightweight features are especially valuable in resource-constrained environments. Shang et al. (2023) demonstrated in their study on apple flower detection that ShuffleNetv2 possesses significant advantages in terms of lightweight design, accuracy, and adaptability, making it an ideal choice for object detection in resource-constrained environments. Additionally, the introduction of DWConv and LSKA structures in the neck of the YOLO-SDL model further enhances its performance. While the DWConv structure does not provide as significant a precision boost as ShuffleNetV2, it plays a crucial role in reducing computational costs. By lowering the number of parameters and

computational load in convolution operations, DWConv improves both efficiency and accuracy. Guo et al. (2024) showed that replacing the convolutional modules in the YOLOv8s network with DWConv modules can reduce the complexity of the network, proving to be an effective lightweight strategy. Although LSKA increases model complexity, it also delivers significant accuracy improvements. By decomposing large kernels into smaller, more manageable parts, LSKA can effectively model global information, which is crucial for distinguishing between wheat grains and other components such as husks and stalks. The LSKA mechanism enhances the model's ability to recognize features in wheat grains by adaptively optimizing local spatial attention within the feature maps. Gao et al. (2024) pointed out in their research aimed at improving the efficiency and accuracy of young citrus fruit detection that the LSKA mechanism enhances the model's ability to fuse features across different scales, thereby improving detection accuracy. The inclusion of DWConv and LSKA in YOLO-SDL boosts detection accuracy, thereby improving overall model performance.

In the model comparison analysis, YOLO-SDL demonstrated clear performance advantages in wheat grain detection tasks. The model achieved the highest values for P, R, mAP50, and mAP50-95 compared to other YOLO series models. From a computational complexity and efficiency standpoint, YOLO-SDL is the most lightweight model, outperforming the others in terms of parameter count and GFLOPs. These findings suggest that appropriate network adjustments and optimizations can significantly enhance overall detection performance while maintaining a lightweight design.

In conclusion, the YOLO-SDL model introduced in this study has demonstrated superior performance in wheat grain detection tasks. By combining the strengths of ShuffleNetV2, DWConv, and LSKA, the model achieves significant improvements in accuracy while balancing lightweight design and performance optimization. These results offer new approaches and methods for agricultural automation and smart farming. Future research could explore further reductions in model complexity, optimize the network structure, and enhance the model's adaptability in resource-limited environments, advancing the development and application of agricultural automation technologies.

5 Conclusion

This study developed an improved YOLOv8n model, termed YOLO-SDL, for detecting various types of wheat grains. Through in-depth analysis and enhancement of the YOLOv8n model, a lightweight, efficient, and highly accurate wheat grain detection model was constructed. By introducing ShuffleNetV2, DWConv, and LSKA structures, the YOLO-SDL model significantly improves the capture of features across different wheat grain types while maintaining high precision, lightweight design, and computational efficiency. The results show that the YOLO-SDL model excels in detection accuracy, achieving a P of 0.942, R of 0.903, mAP50 of 0.965, and mAP50-95 of 0.859, outperforming YOLOv8n and other YOLO series models in both lightweight design and computational efficiency. The YOLO-SDL model's advantages in parameter count and computational complexity make it highly suitable for deployment in resource-constrained environments, which is particularly relevant for real-world agricultural automation and intelligent systems. The outstanding performance of the YOLO-SDL model is attributed to the introduction of ShuffleNetV2, DWConv, and LSKA structures. The ShuffleNetV2 architecture, through its use of grouped convolution and channel shuffle mechanisms, maintains high accuracy while reducing computational costs. The DWConv structure, by decomposing standard convolutions into depthwise and pointwise convolutions, further reduces parameter count and computational complexity. Finally, the LSKA mechanism, by decomposing 2D convolutional kernels into 1D kernels, achieves adaptive feature optimization, enhancing the model's ability to capture image features and improving detection accuracy.

The YOLO-SDL model demonstrates excellent performance in wheat grain detection tasks, improving both accuracy and efficiency while achieving a lightweight design. This provides a novel technical solution for agricultural automation and intelligent farming. Future research will continue to explore and optimize the model's structure to extend its applicability to a wider range of agricultural scenarios, driving further advancements in agricultural technology.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

ZQ: Writing – review & editing, Supervision, Resources, Investigation, Conceptualization. FW: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. WW: Writing – review & editing, Formal analysis, Data curation. TL: Writing – review & editing, Visualization, Formal analysis, Data curation. XJ: Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. SQ: Writing – review & editing, Software, Methodology. YS: Writing – review & editing, Software, Project administration, Funding acquisition.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by National Key Research and Development

Program Sub-project (No. 2022YFD2001205C), and Henan Province Major Science and Technology Special Project (No. 231100110200), and Key R&D and Promotion Projects in Henan Province (Science and Technology Development) (No.222102110452), and Key Scientific Research Projects of Colleges and Universities in Henan Province (No. 22B416002).

Acknowledgments

We extend our sincere gratitude to the farmers of Shenqiu County, Zhoukou City, Henan Province, for their invaluable support in providing experimental materials of wheat grains. Their contributions have been instrumental in facilitating the successful conduct of our research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bacea, D.-S., and Oniga, F. (2023). Single stage architecture for improved accuracy real-time object detection on mobile devices. *Image. Vision Computing.* 130, 104613. doi: 10.1016/j.imavis.2022.104613

Balasundaram, A., Shaik, A., Pradeep, K., Vijayakumar, K. P., and Kavitha, M. S. (2023). Zero-DCE++ Inspired object detection in less illuminated environment using improved YOLOv5. *Computers. Mater. Continua.* 77, 2751–2769. doi: 10.32604/cmc.2023.044374

Chen, G., Hou, Y., Cui, T., Li, H., Shangguan, F., and Cao, L. (2024). YOLOv8-CML: A lightweight target detection method for Color-changing melon ripening in intelligent agriculture. *Sci. Rep.* 14, 14400. doi: 10.1038/s41598-024-65293-w

Ding, Y., Huang, H., Cui, H., Wang, X., and Zhao, Y. (2023). A non-destructive method for identification of tea plant cultivars based on deep learning. *Forests* 14, 728. doi: 10.3390/f14040728

Fang, S., Zhang, B., and Hu, J. (2023). Improved mask R-CNN multi-target detection and segmentation for autonomous driving in complex scenes. *Sensors* 23, 3853. doi: 10.3390/s23083853

Feng, Y., Liu, C., Han, J., Lu, Q., and Xing, X. (2023). IRB-5-CA net: A lightweight, deep learning-based approach to wheat seed identification. *IEEE Access.* 11, 119553–119559. doi: 10.1109/ACCESS.2023.3326824

Gao, A., Tian, Z., Ma, W., Song, Y., Ren, L., Feng, Y., et al. (2024). Fruits hidden by green: an improved YOLOV8n for detection of young citrus in lush citrus trees. *Front. Plant Sci.* 15, 1375118. doi: 10.3389/fpls.2024.1375118

Gao, H., Zhen, T., and Li, Z. (2022). Detection of wheat unsound kernels based on improved ResNet. *IEEE Access.* 10, 20092–20101. doi: 10.1109/ACCESS.2022.3147838

Guo, Y., Chen, Y., Wang, Y., Wu, X., Zhang, X., Mao, W., et al. (2023). The translational landscape of bread wheat during grain development. *Plant Cell.* 35, 1848–1867. doi: 10.1093/plcell/koad075

Guo, P., Diao, Z., Zhao, C., Li, J., Zhang, R., Yang, R., et al. (2024). Navigation line extraction algorithm for corn spraying robot based on YOLOv8s-CornNet. *J. Field Robotics.* 41, 1887–1899. doi: 10.1002/rob.22360

Han, K., Zhang, N., Xie, H., Wang, Q., and Ding, W. (2023). An improved strategy of wheat kernel recognition based on deep learning. *DYNA-Ingeniería. e. Industria.* 98, 91–97. doi: 10.6036/DYNAII

Ji, W., Pan, Y., Xu, B., and Wang, J. (2022). A real-time apple targets detection method for picking robot based on ShufflenetV2-YOLOX. *Agriculture* 12, 856. doi: 10.3390/agriculture12060856

Jiang, P., Qi, A., Zhong, J., Luo, Y., Hu, W., Shi, Y., et al. (2024). Field cabbage detection and positioning system based on improved YOLOv8n. *Plant Methods* 20, 96. doi: 10.1186/s13007-024-01226-y

Lau, K. W., Po, L.-M., and Rehman, Y. A. U. (2024). Large separable kernel attention: Rethinking the large kernel attention design in cnn. *Expert Syst. Applications.* 236, 121352. doi: 10.1016/j.eswa.2023.121352

Li, L., Hassan, M. A., Yang, S., Jing, F., Yang, M., Rasheed, A., et al. (2022a). Development of image-based wheat spike counter through a Faster R-CNN algorithm and application for genetic studies. *Crop J.* 10, 1303–1311. doi: 10.1016/j.cj.2022.07.007

Li, F., Yan, H., and Shi, L. (2024a). Multi-scale coupled attention for visual object detection. *Sci. Rep.* 14, 11191. doi: 10.1038/s41598-024-60897-8

Li, M., Zhang, J., Liu, H., Yuan, Y., Li, J., and Zhao, L. (2024b). A lightweight method for apple-on-tree detection based on improved YOLOv5. *SIViP* 18, 6713–6727. doi: 10.1007/s11760-024-03346-3

Li, W., Zhu, T., Li, X., Dong, J., and Liu, J. (2022b). Recommending advanced deep learning models for efficient insect pest detection. *Agriculture* 12, 1065. doi: 10.3390/agriculture12071065

Li, Z., Zhu, Y., Sui, S., Zhao, Y., Liu, P., and Li, X. (2024c). Real-time detection and counting of wheat ears based on improved YOLOv7. *Comput. Electron. Agricult.* 218, 108670. doi: 10.1016/j.compag.2024.108670

Lu, Y., Yu, J., Zhu, X., Zhang, B., and Sun, Z. (2024). YOLOv8-Rice: a rice leaf disease detection model based on YOLOv8. *Paddy Water Environ.* 22, 695–710. doi: 10.1007/s10333-024-00990-w

Ma, M., Li, Y., Xue, C., Xiong, W., Peng, Z., Han, X., et al. (2021). Current situation and key parameters for improving wheat quality in China. *Front. Plant Sci.* 12, 638525. doi: 10.3389/fpls.2021.638525

Ma, N., Su, Y., Yang, L., Li, Z., and Yan, H. (2024). Wheat seed detection and counting method based on improved YOLOv8 model. *Sensors* 24, 1654. doi: 10.3390/s24051654

Ma, N., Zhang, X., Zheng, H.-T., and Sun, J. (2018). "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, vol 11218.* eds. V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss. (Cham: Springer). doi: 10.1007/978-3-030-01264-9_8

Misra, T., Arora, A., Marwaha, S., Ranjan Jha, R., Ray, M., Kumar, S., et al. (2022). Yield-SpikeSegNet: An extension of SpikeSegNet deep-learning approach for the yield estimation in the wheat using visual images. *Appl. Artif. Intelligence.* 36, 2137642. doi: 10.1080/08839514.2022.2137642

Peng, Y., Zhao, Y., Yu, Z., Zeng, J., Xu, D., Dong, J., et al. (2022). Wheat quality formation and its regulatory mechanism. *Front. Plant Sci.* 13, 834654. doi: 10.3389/fpls.2022.834654

Qing, S., Qiu, Z., Wang, W., Wang, F., Jin, X., Ji, J., et al. (2024). Improved YOLO-FastestV2 wheat spike detection model based on a multi-stage attention mechanism with a LightFPN detection head. *Front. Plant Sci.* 15, 1411510. doi: 10.3389/fpls.2024.1411510

Que, H., Zhao, X., Sun, X., Zhu, Q., and Huang, M. (2023). Identification of wheat kernel varieties based on hyperspectral imaging technology and grouped convolutional neural network with feature intervals. *Infrared. Phys. Technol.* 131, 104653. doi: 10.1016/j.infrared.2023.104653

Shang, Y., Xu, X., Jiao, Y., Wang, Z., Hua, Z., and Song, H. (2023). Using lightweight deep learning algorithm for real-time detection of apple flowers in natural environments. *Comput. Electron. Agricult.* 207, 107765. doi: 10.1016/j.compag.2023.107765

Shi, L., Wei, Z., You, H., Wang, J., Bai, Z., Yu, H., et al. (2024). OMC-YOLO: A lightweight grading detection method for oyster mushrooms. *Horticulturae* 10, 742. doi: 10.3390/horticulturae10070742

Thakuria, A., and Erkinbaev, C. (2024). Real-time canola damage detection: An end-to-end framework with semi-automatic crusher and lightweight ShuffleNetV2_YOLOv5s. *Smart. Agric. Technol.* 7, 100399. doi: 10.1016/j.atech.2024.100399

Wan, D., Lu, R., Hu, B., Yin, J., Shen, S., and Lang, X. (2024). YOLO-MIF: Improved YOLOv8 with Multi-Information fusion for object detection in Gray-Scale images. *Adv. Eng. Informatics.* 62, 102709. doi: 10.1016/j.aei.2024.102709

Wu, H., Wang, Y., Zhao, P., and Qian, M. (2023). Small-target weed-detection model based on YOLO-V4 with improved backbone and neck structures. *Precis. Agricult.* 24, 2149–2170. doi: 10.1007/s11119-023-10035-7

Wu, W., Yang, T., Li, R., Chen, C., Liu, T., Zhou, K., et al. (2020). Detection and enumeration of wheat grains based on a deep learning method under various scenarios and scales. *J. Integr. Agricult.* 19, 1998–2008. doi: 10.1016/S2095-3119(19)62803-0

Yao, Q., Zheng, X., Zhou, G., and Zhang, J. (2024). SGR-YOLO: a method for detecting seed germination rate in wild rice. *Front. Plant Sci.* 14, 1305081. doi: 10.3389/fpls.2023.1305081

Ye, Z., Yang, K., Lin, Y., Guo, S., Sun, Y., Chen, X., et al. (2023). A comparison between Pixel-based deep learning and Object-based image analysis (OBIA) for individual detection of cabbage plants based on UAV Visible-light images. *Comput. Electron. Agricult.* 209, 107822. doi: 10.1016/j.compag.2023.107822

Zhang, Y., Zhang, H., Huang, Q., Han, Y., and Zhao, M. (2024). DsP-YOLO: An anchor-free network with DsPAN for small object detection of multiscale defects. *Expert Syst. Applications.* 241, 122669. doi: 10.1016/j.eswa.2023.122669

Zhao, W., Liu, S., Li, X., Han, X., and Yang, H. (2022). Fast and accurate wheat grain quality detection based on improved YOLOv5. *Comput. Electron. Agricult.* 202, 107426. doi: 10.1016/j.compag.2022.107426

Zhong, C., Wu, H., Jiang, J., Zheng, C., and Song, H. (2024). YOLO-DLHS-P: A lightweight behavior recognition algorithm for captive pigs. *IEEE Access.* 12, 104445–104462. doi: 10.1109/ACCESS.2024.3414859

Zhou, L., Zhang, C., Taha, M. F., Wei, X., He, Y., Qiu, Z., et al. (2020). Wheat kernel variety identification based on a large near-infrared spectral dataset and a novel deep learning-based feature selection method. *Front. Plant Sci.* 11, 575810. doi: 10.3389/fpls.2020.575810

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 22 November 2024
DOI 10.3389/fpls.2024.1492372

Check for updates

OPEN ACCESS

EDITED BY
Ting Sun,
China Jiliang University, China

REVIEWED BY
Baodong Xu,
Huazhong Agricultural University, China
Meichen Feng,
Shanxi Agricultural University, China
Jiayi Zhang,
Jiangsu Academy of Agricultural Sciences Wuxi Branch (JAASWB), China

*CORRESPONDENCE
Huifang Wang
whf428@126.com
Haikuan Feng
fenghk@nercita.org.cn

†These authors have contributed equally to this work

RECEIVED 06 September 2024
ACCEPTED 06 November 2024
PUBLISHED 22 November 2024

CITATION
Fan J, Liu Y, Fan Y, Yao Y, Chen R, Bian M, Ma Y, Wang H and Feng H (2024) Estimation of potato leaf area index based on spectral information and Haralick textures from UAV hyperspectral images.
Front. Plant Sci. 15:1492372.
doi: 10.3389/fpls.2024.1492372

COPYRIGHT
© 2024 Fan, Liu, Fan, Yao, Chen, Bian, Ma, Wang and Feng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Estimation of potato leaf area index based on spectral information and Haralick textures from UAV hyperspectral images

Jiejie Fan[1,2†], Yang Liu[1,3†], Yiguang Fan[1], Yihan Yao[4], Riqiang Chen[1], Mingbo Bian[1], Yanpeng Ma[1], Huifang Wang[5*] and Haikuan Feng[1,6,7*]

[1]Key Laboratory of Quantitative Remote Sensing in Agriculture of Ministry of Agriculture and Rural Affairs, Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China, [2]College of Geomatics, Xi'an University of Science and Technology, Xi'an, China, [3]Key Lab of Smart Agriculture System, Ministry of Education, China Agricultural University, Beijing, China, [4]College of Information and Management Science, Henan Agricultural University, Zhengzhou, China, [5]Ecological Meteorology and Satellite Remote Sensing, Beijing Municipal Climate Center, Beijing, China, [6]College of Agriculture, Nanjing Agricultural University, Nanjing, China, [7]Collaborative Innovation Center for Modern Crop Production Co-sponsored by Province and Ministry, Nanjing Agricultural University, Nanjing, Jiangsu, China

The Leaf Area Index (LAI) is a crucial parameter for evaluating crop growth and informing fertilization management in agricultural fields. Compared to traditional methods, UAV-based hyperspectral imaging technology offers significant advantages for non-destructive, rapid monitoring of crop LAI by simultaneously capturing both spectral information and two-dimensional images of the crop canopy, which reflect changes in its structure. While numerous studies have demonstrated that various texture features, such as the Gray-Level Co-occurrence Matrix (GLCM), can be used independently or in combination with crop canopy spectral data for LAI estimation, limited research exists on the application of Haralick textures for evaluating crop LAI across multiple growth stages. In this study, experiments were conducted on two early-maturing potato varieties, subjected to different treatments (e.g., planting density and nitrogen levels) at the Xiaotangshan base in Beijing, during three key growth stages. Crop canopy spectral reflectance and Haralick textures were extracted from ultra-low-altitude UAV hyperspectral imagery, while LAI was measured using ground-based methods. Three types of spectral data—original spectral reflectance (OSR), first-order differential spectral reflectance (FDSR), and vegetation indices (VIs)—along with three types of Haralick textures—simple, advanced, and higher-order—were analyzed for their correlation with LAI across multiple growth stages. A model for LAI estimation in potato at multiple growth stages based on spectral and textural features screened by the successive projection algorithm (SPA) was constructed using partial least squares regression (PLSR), random forest regression (RFR) and gaussian process regression (GPR) machine learning methods. The results indicated that: (1) Spectral data demonstrate greater sensitivity to LAI than Haralick textures, with sensitivity decreasing in the order of VIs, FDSR and OSR; (2) spectral data alone provide more accurate LAI estimates than Haralick textures, with VIs achieving an

accuracy of R^2 = 0.63, RMSE = 0.38, NRMSE = 28.36%; and (3) although Haralick textures alone were not effective for LAI estimation, they can enhance LAI prediction when combined with spectral data, with the GPR method achieving R^2 = 0.70, RMSE = 0.30, NRMSE = 20.28%. These findings offer a valuable reference for large-scale, accurate monitoring of potato LAI.

KEYWORDS

potato, Unmanned Aerial Vehicle, hyperspectral, Haralick textures, leaf area index

1 Introduction

The potato, ranked fourth in the global list of food crops, is irreplaceable in achieving development goals such as the eradication of hunger, poverty and malnutrition (Ahmed et al., 2024). Its production in developing countries has grown significantly over the past two decades, making the monitoring of its growth critical to the security and stability of the global food supply (Badr et al., 2022). The Leaf Area Index (LAI) represents the total one-sided leaf area of a crop per unit ground area and is closely linked to key physiological processes such as respiration, transpiration, productivity, and photosynthesis (Li et al., 2016; Ilniyaz et al., 2022). Rapid and accurate estimation of LAI is essential for diagnosing crop growth and predicting yield (Cohrs et al., 2020; Kaplan and Rozenstein, 2021). While traditional manual measurement methods are precise, they are also time-consuming, destructive to crops, and incapable of providing large-scale spatiotemporal data, thereby limiting the effectiveness of crop management. Remote sensing, which involves detecting and obtaining information about objects from a distance without direct contact (Kokubu et al., 2020; Raj et al., 2021), has become increasingly important in monitoring crop LAI due to its non-destructive nature and high-throughput capabilities. Among the various remote sensing technologies available, UAV (Unmanned Aerial Vehicle) remote sensing has emerged as the preferred method for estimating crop LAI, owing to its ability to configure multiple sensors, its flexibility in data collection, and the high spatial and temporal resolution of the images it captures (Dong et al., 2019; Revill et al., 2020).

There are three primary approaches to crop LAI estimation using UAV remote sensing imagery: physical models, hybrid models, and empirical statistical models (Verrelst et al., 2015). Physical models are based on light-crop interactions, such as reflection and absorption, providing a mechanistic and generalizable framework (Fu et al., 2022). These models simulate crop canopy spectral reflectance, which can indirectly estimate LAI. However, despite their mechanistic nature, physical models require numerous input parameters, making them costly to implement. Hybrid models leverage simulation data from physical models to train machine learning algorithms, offering an effective method for estimating LAI without extensive ground-truthing data (Verrelst et al., 2015). On the other hand, empirical statistical models establish a statistical relationship between crop characteristics and LAI (Cao et al., 2020). This approach is user-friendly, easy to implement, and has gained popularity. UAV hyperspectral remote sensing has high spectral resolution and strong band continuity, and its original spectral reflectance (OSR) contains a large amount of spectral information, which can be effectively used for crop physicochemical parameter estimation (Chen et al., 2024). In order to improve the accuracy of hyperspectral remote sensing in the estimation of crop physicochemical parameters, the OSR need to be preprocessed to reduce noise and improve accuracy. First-order differential spectral reflectance (FDSR) not only eliminates the effect of background, but also resolves the overlapping signals and improves the sensitivity of spectral features to crop physicochemical parameters. Additionally, vegetation indices (VIs) are also an effective form of extracting spectral features of crop canopies, which are mainly formed by combinations of reflectance in two or more bands in the visible to near-infrared region (Fan et al., 2022; Liu et al., 2024b). Optical VIs can mitigate the effects of canopy structure and soil on crop reflectance spectra, thereby enhancing and highlighting relevant crop information (Liu et al., 2024a). Numerous VIs are currently used to estimate crop LAI. For example, Lu et al. (2022) evaluated the performance of several VIs in estimating wheat LAI at multiple growth stages and found that the normalized red-edge index exhibited a significant linear relationship with LAI (R^2 = 0.53). Similarly, Gong et al. (2021) assessed the LAI of rice throughout its growth stages using different VIs and discovered that the Enhanced Vegetation Index 2 produced higher estimation accuracy (R^2 = 0.38). These studies show a gradual decrease in the sensitivity of VIs to LAI over different growth stages, thus challenging the use of VIs for multiple growth stages LAI estimation.

The band reflectance used to form VIs is primarily influenced by leaf morphology and canopy structure (Croft et al., 2014; Zhang et al., 2021b). Compared to maize and wheat, potatoes exhibit more complex changes in canopy structure throughout their growth phases (Li et al., 2020; Yang et al., 2021). During the nutrient reserve stage, potato stems grow randomly, and leaves increase in size. As the crop enters the pre-mid reproductive stage, the canopy

becomes denser, leading to VI saturation under high LAI conditions. As underground tubers begin to form, nutrients accumulated during the earlier growth stages are transferred to these tubers. At this stage, potato leaves start to yellow and wither, significantly affecting the reflectance spectrum of the potato canopy. Consequently, estimating LAI without considering the growth stage of the potato, which accounts for changes in canopy structure, is challenging. To achieve accurate LAI estimation, the differences in potato canopy structure across growth stages must be quantified. Texture features of an image, which define the variability between neighboring pixels within a given window, can describe variations in canopy structure (Wulder et al., 1998; Yue et al., 2019; Zhang et al., 2022c). Recent studies have incorporated texture features when estimating LAI across multiple crop growth stages (Li et al., 2019). For instance, Qiao et al. (2024) improved multi-growth stage LAI estimation in peanut by combining VIs with image texture features extracted using the Gray-Level Co-occurrence Matrix (GLCM). Similarly, Zhou et al. (2022) demonstrated that texture features extracted using discrete wavelet transform, when combined with VIs, provided more accurate LAI estimates for rice compared to using VIs alone. While combining texture features with VIs is expected to enhance LAI estimation across different crop growth stages, it remains uncertain whether existing texture feature extraction methods can accurately capture the complex changes in potato canopy structure at various growth stages. Therefore, it is essential to explore texture features that effectively reflect these structural changes to assist VIs in improving LAI estimation accuracy across multiple growth stages in potatoes.

Among texture features, the Haralick texture feature is the most prominent, relying on the GLCM. Typically, one or more summary statistics, known as Haralick features, are used to summarize the GLCM for ease of interpretation (Haralick et al., 1973; Das and Naskar, 2024). Different types of texture features, including simple, advanced, and high-order features, can be extracted using the Haralick texture extraction method based on the GLCM (Jiang et al., 2021). Among these features, it is possible to filter out those that are particularly sensitive to changes in potato canopy structure, which, when combined with VIs, can enable accurate estimation of LAI across multiple growth stages. Compared to digital or multispectral images, UAV hyperspectral images can provide richer spectral information. When captured under ultra-low-altitude flight conditions, these high-spatial-resolution hyperspectral images can also extract canopy texture features that respond to crop LAI (Duan et al., 2019). Consequently, UAV-based hyperspectral imaging technology offers significant potential for improving LAI estimation. However, extracting sensitive features from hyperspectral image-derived variables is challenging due to the high dimensionality and complexity of the data. As a result, sensitive feature selection methods, such as genetic algorithms, successive projection algorithms (SPA), and stepwise regression algorithms, have been employed to identify the optimal combination of spectral and texture features (Gu et al., 2022; Ji et al., 2022; Ma et al., 2022). Zhang et al. (2021a) demonstrated that features extracted using SPA combined with machine learning can yield more accurate LAI estimation results. Although machine learning methods are increasingly powerful tools for analyzing spatial variations in LAI from multiple remotely sensed feature datasets, not all machine learning approaches are equally effective for crop LAI estimation at multiple growth stages. The choice of appropriate regression methods is also crucial for improving the accuracy of LAI estimation. To the best of our knowledge, few studies have investigated the use of hyperspectral imagery to extract spectral and Haralick texture features for estimating LAI across multiple growth stages in potato.

To enhance the accuracy of potato LAI estimation across multiple growth stages, this study aimed to: (1) evaluate the performance of spectral information and Haralick texture features extracted from UAV hyperspectral imagery for LAI estimation independently; (2) determine whether combining spectral information with Haralick texture features improves LAI estimation accuracy; and (3) identify the optimal combination of variables and the most effective regression method for estimating potato LAI at various growth stages. The findings of this study offer insights into utilizing UAV hyperspectral imaging technology to obtain both canopy spectral and structural information, thereby improving the precision of LAI estimation and establishing a foundation for monitoring potato growth at the field scale.

2 Materials and methods

2.1 Experiment design

Potato trials were conducted at the National Precision Agriculture Experimental Base in Changping District, Beijing, China (Figure 1). Two early-maturing potato varieties, Zhongshu 3 (Z3) and Zhongshu 5 (Z5), were planted. To increase the spatial variability of LAI, three treatments were performed, including different planting densities, nitrogen gradients, and potassium gradients. Planting density was categorized into three levels: 60,000 (P1), 72,000 (P2), and 84,000 (P3) tubers/hm^2. Nitrogen treatments were categorized into four levels: 0 (N0), 112.5 (N1), 225 (N2), and 337.5 (N3) kg/hm^2 of pure nitrogen. Potash treatments were categorized into three levels: 0 (K0), 495 (K1, applied to the planting density and nitrogen test areas), and 990 (K2) kg/hm^2 of K_2O. In this experiment, each plot size of 32.5 m^2 was applied with 90 kg/hm2 P2O5 and each treatment was replicated three times, totalling 48 plots. Field management practices, including weeding, fertilization, and irrigation, were consistent with local agricultural conditions. To calibrate the UAV hyperspectral images acquired during the study, 11 ground control points were evenly distributed around the experimental field. The spatial locations of these points were determined using the Qianxun SR2 high-precision RTK measurement system.

2.2 Hyperspectral images acquisition and processing

In 2019, UAV hyperspectral images were acquired during three critical stages of potato development: tuber formation (28 May), tuber growth (10 June), and starch accumulation (20 June). These stages correspond to BBCH codes 41, 44, and 47 (referred to as BBCH41, BBCH44, and BBCH47), respectively. The images were captured using an M600 UAV (SZ DJI Technology Co., Ltd.,

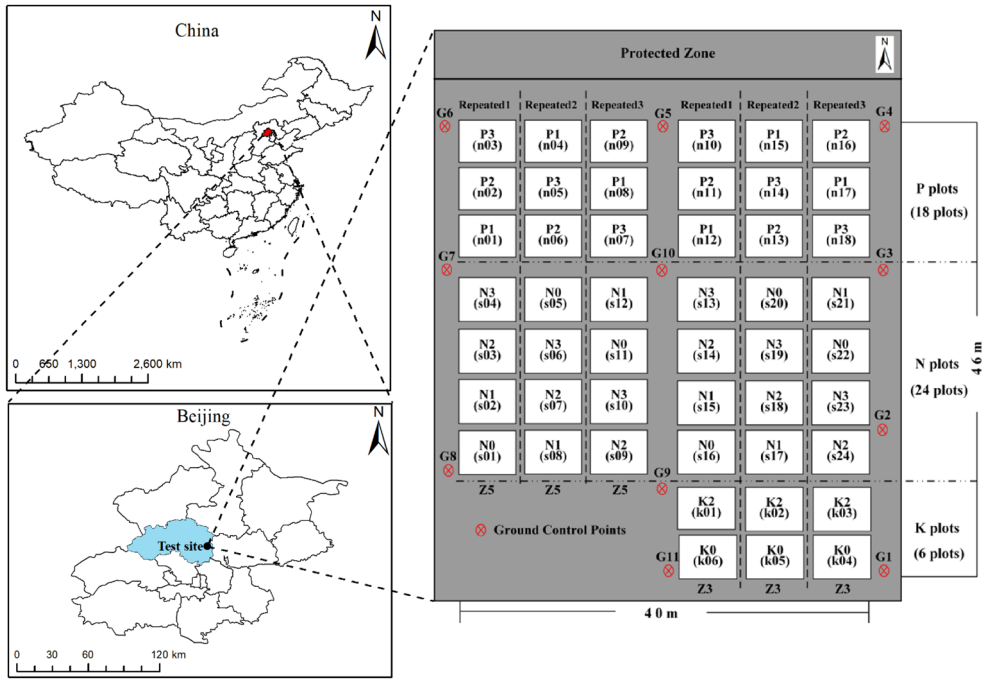


FIGURE 1
The spatial planting distribution of potatoes in this study.

Shenzhen, China) equipped with a UHD 185-Firefly imaging spectrometer (Cubert GmbH, Ulm, Germany). The UHD185 has 125 spectral channels covering a wavelength range of 450–950 nm, with a sampling interval of 4 nm. The UAV followed identical flight paths across all three growth stages, maintaining consistent take-off positions. The image acquisition was conducted under clear, cloudless, and windless conditions, between 11:30 am and 13:30 pm, when direct sunlight reached the ground. Flights were carried out at an altitude of 20 meters, resulting in a spatial resolution of 1.43 cm, with a flight speed of 1.5 m/s. To facilitate the efficient stitching of orthophotos, the heading and sidelap overlaps were manually set to 85% and 93%, respectively. Given the small size of the field, each hyperspectral image acquisition took approximately 12 minutes. Before each take-off, dark current and radiometric calibration were performed on the ground using a black-and-white calibration plate.

The processing of UAV hyperspectral images involves several key steps: panchromatic image stitching, fusion of panchromatic and hyperspectral images, and extraction of canopy spectral reflectance. First, Agisoft PhotoScan software (Agisoft LLC, St. Petersburg, Russia) was used to correct the topography of the panchromatic images by integrating the positional data from the ground control points. Next, the panchromatic images were mosaicked using the structure-from-motion algorithm, and orthophotos of the test field were generated. The panchromatic orthophoto was then fused with the hyperspectral image using Cube-Pilot software (Cubert GmbH, Ulm, Germany) to create a new hyperspectral image. Finally, the average spectral reflectance of the potato canopy was extracted from the vector files of the 48 plots using ENVI software, and this data was used for subsequent analysis.

2.3 LAI measurement

To obtain LAI data, three uniformly growing plants were destructively sampled at each growth stage of tuber formation, tuber growth, and starch accumulation. Plants were transported quickly to the laboratory to prevent water loss leading to leaf curling. After manual separation of the stems and leaves, 60 leaf discs were punched out with a 0.8 cm diameter punch and their wet weights were measured using a high-precision balance (W1). The remaining leaves of the three plants were also measured for their wet weight (W2). Finally, the LAI of each plot was calculated using the weight ratio and planting density. The formula for LAI was as follows:

$$LAI = \frac{(W1 + W2)}{3 \times W1} \times S \times M \quad (1)$$

Where WI and W2 were the wet weight of 60 leaf discs and the remaining leaves of three plants respectively. S was the area of leaf discs. M was the number of potato plants per unit area.

2.4 The extraction of spectral and textural feature

The spectral information extracted from UAV hyperspectral images in this study primarily includes canopy original spectral reflectance (OSR), first-order differential spectral reflectance (FDSR), and vegetation indices (VIs). Additionally, simple, advanced, and higher-order texture features were derived using Haralick texture extraction methods.

2.4.1 The extraction of FDSR

A first-order differential transformation of the raw canopy spectra can eliminate the interference of linear background noise on the canopy spectral signals and enhance the spectral absorption and reflectance features in the visible to near-infrared region, which is conducive to the improvement of the spectral reflectance response to LAI (Imanishi et al., 2004; Li et al., 2014). Hyperspectral data are well-suited for first-order differential processing due to their large number of continuous bands. The first-order differentiation formula used in this study is as follows:

$$\mathrm{FDSR}_{\lambda(i)} = \frac{R_{\lambda(i-1)} - R_{\lambda(i+1)}}{\Delta\lambda} \quad (2)$$

where $\mathrm{FDSR}_{\lambda(i)}$ is the first-order differential spectral reflectance at a central wavelength i between waveband $i-1$ and $i+1$. $R_{\lambda(i-1)}$ and $R_{\lambda(i+1)}$ are the reflectance in the waveband $i-1$ and $i+1$, respectively. $\Delta\lambda$ is the sampling interval.

2.4.2 The extraction of VIs

VIs are effective spectral signals used to enhance vegetation information (Sone et al., 2009; Hu et al., 2021; Wengert et al., 2021). In this study, thirty VIs were selected for estimating LAI, with their specific expressions and names provided in Table 1.

2.4.3 The extraction of Haralick texture features

Haralick texture features are computed from a GLCM, which records the co-occurrence of neighboring gray levels in an image (Brynolfsson et al., 2017). The GLCM is a square matrix, where the number of rows and columns corresponds to the number of gray levels in the region of interest. Each texture feature is derived from the elements of the GLCM and represents a specific relationship between neighboring voxels. In this study, we utilized the Python package and Orfeo Toolbox 7.2.0 (https://www.orfeo-toolbox.org/download/) to extract the Haralick features.

In order to reduce the difficulty of extracting texture features, principal component analysis (PCA) was performed on the raw hyperspectral images using ENVI software. It was found that the first principal component image could explain 90% of the image information. Consequently, various Haralick texture features were extracted from this first principal component image. Simple, advanced, and higher-order textures were derived using a 5×5 window size in the 45° direction. The names and quantities of the extracted Haralick texture features are listed in Table 2.

2.5 Statistical analysis and methodology

The optimal combination of variables for estimating LAI was determined using the SPA. SPA is a forward variable selection algorithm based on the projection analysis of vectors (Sun et al., 2019). Initially, a set of vectors is randomly selected as the starting variables. These vectors are then projected onto the unselected vectors, and the magnitudes of the projected vectors are compared. The variable with the largest projected vector is selected, and this process is repeated through several iterations. Ultimately, the optimal combination of variables is chosen based on the minimum root mean square error (RMSE) of cross-validation.

Partial Least Squares Regression (PLSR) is a statistical method used to develop a linear regression model by projecting independent and dependent variables into a new space. This method involves both the extraction of principal components from the independent and dependent variables and the maximization of the correlation between these components during the regression modeling process (Tao et al., 2020). Random Forest Regression (RFR) is a machine learning technique that integrates a large number of decision trees (Yue et al., 2018). The accuracy of the RFR model is primarily influenced by the number of decision trees (ntree) and the number of nodes (mtry), with mtry typically set to one-third of the input variables and ntree set to 500. Gaussian Process Regression (GPR) is a method based on Bayes' theorem that models the relationship between dependent and independent variables using a kernel function (Liu et al., 2022b). Compared to other popular machine learning methods, GPR models are simpler to optimize and particularly suitable for training on small sample datasets. In this study, MATLAB software was used for training and validating LAI models with the aforementioned machine learning methods. A total of 96 samples (replicates 2 and 3 for each growth stage) were used for model training, while 48 samples (replicate 1 for each growth stage) were reserved for model validation.

Our flowchart from different types of spectral and texture extraction to model construction and evaluation in this study is shown in Figure 2.

The coefficient of determination (R^2), root mean square error (RMSE), and normalized root mean square error (NRMSE) were used to evaluate the accuracy of the models. The corresponding equations are shown in Equations 3–5:

$$R^2 = \frac{(\sum_{i=1}^{n} y_i - \bar{y})^2}{(\sum_{i=1}^{n} x_i - \bar{y})^2} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1,j=1}^{n} (x_i - y_i)^2}{n}} \quad (4)$$

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1,j=1}^{n} (x_i - y_j)^2}{n}}}{\bar{y}} \quad (5)$$

Where x_i is the measured potato LAI; y_i is the estimated potato LAI; $\bar{y}$ is the mean value of the measured potato LAI; n is the number of samples.

3 Results and analysis

3.1 Analysis of measured LAI

In the field experiment, LAI values of potato at each growth stage in 48 experimental plots were determined. A total of 144 LAI

TABLE 1 The VIs used in the study.

VIs	Number	Formula	Reference
Difference vegetation index (DVI)	1	R_{890}-R_{670}	(Richardson and Wiegand, 1977)
Enhanced vegetation index (EVI)	2	$2.5\times(R_{800}\text{-}R_{670})/(R_{800}+6\times R_{670}\text{-}7.5\times R_{450}+1)$	(Ahamed et al., 2011)
Enhanced vegetation index2 (EVI2)	3	$2.5\times(R_{800}\text{-}R_{670})/(R_{800}+2.4\times R_{670}+1)$	(Jiang et al., 2008)
Green normalized-difference vegetation index (GNDVI)	4	$(R_{750}\text{-}R_{550})/(R_{750}+R_{550})$	(Gitelson et al., 1996)
Greenness index (GI)	5	$R_{554}/R6_{77}$	(Zarco-Tejada et al., 2005)
Linear combination index (LCI)	6	$(R_{850}\text{-}R_{710})/(R_{850}+R_{670})^{1/2}$	(Datt, 1999)
Modified chlorophyll-absorption ratio index (MCARI)	7	$((R_{700}\text{-}R_{670})\text{-}0.2\times(R_{700}\text{-}R_{550}))(R_{700}/R_{670})$	(Haboudane et al., 2002)
Modified simple ratio index (MSR)	8	$(R_{800}/R_{670}\text{-}1)/(R_{800}/R_{670}+1)^{1/2}$	(Chen, 1996)
Modified soil adjusted vegetation index (MSAVI)	9	$0.5\times[2\times R_{800}+1-((2\times R_{800}+1)^2\text{-}8\times(R_{800}\text{-}R_{670}))^{1/2}]$	(Haboudane et al., 2002)
Modified triangular vegetation index 1 (MTVI1)	10	$1.2\times[1.2\times(R_{800}\text{-}R_{550})\text{-}2.5(R_{670}\text{-}R_{550})]$	(Haboudane et al., 2004)
Modified triangular vegetation index 2 (MTVI2)	11	$1.5\times(1.2\times(R_{800}\text{-}R_{500})\text{-}2.5\times(R_{670}\text{-}R_{550}))/(2\times(R_{800}+1)^2\text{-}(6\times R_{800}\text{-}5\times(R_{670})^{1/2})\text{-}0.5)^{1/2}$	(Haboudane et al., 2004)
Normalized-difference vegetation index (NDVI)	12	$(R_{800}\text{-}R_{680})/(R_{800}+R_{680})$	(Rouse et al., 1974)
Normalized difference red edge (NDRE)	13	$(R_{790}\text{-}R_{720})/(R_{790}+R_{720})$	(Fitzgerald et al., 2010)
Normalized pigment chlorophyll ratio index (NPCI)	14	$(R_{670}\text{-}R_{460})/(R_{670}+R_{460})$	(Peñuelas et al., 1994)
Normalized difference index (NDI)	15	$(R_{850}\text{-}R_{710})/(R_{850}+R_{680})$	(Apan et al., 2003)
Nitrogen reflectance index (NRI)	16	$(R_{570}\text{-}R_{670})/(R_{570}+R_{670})$	(Barnes et al., 1992)
Optimizing soil-adjusted vegetation index) (OSAVI)	17	$1.16\times(R_{800}\text{-}R_{670})/(R_{800}+R_{670}+0.16)$	(Li et al., 2016)
Plant senescence reflectance index (PSRI)	18	$(R_{680}\text{-}R_{500})/R_{750}$	(Merzlyak et al., 1999)
Pigment-specific normalized difference (PSND)	19	$(R_{800}\text{-}R_{470})/(R_{800}+R_{470})$	(Blackburn, 1998)
Plant biochemical index (PBI)	20	R_{810}/R_{560}	(Rama Rao et al., 2008)
Ratio vegetation index (RVI)	21	R_{810}/R_{660}	(Xue et al., 2004)
Renormalized-difference vegetation index (RDVI)	22	$(R_{800}\text{-}R_{670})/(R_{800}+R_{670})^{1/2}$	(Roujean and Breon, 1995)
Ratio analysis of reflectance spectra (RASI)	23	R_{760}/R_{500}	(Chappelle et al., 1990)
Red-edge vegetation stress index (RVSI)	24	$[(R_{712}+R_{752})/2]\text{-}R_{732}$	(Merton and Huntington, 1999)
Soil-adjusted vegetation index (SAVI)	25	$1.5\times(R_{800}\text{-}R_{670})/(R_{800}+R_{670}+0.5)$	(Huete, 1988)
Spectral-polygon vegetation index (SPVI)	26	$0.4\times[3.7\times(R_{800}\text{-}R_{670})\text{-}1.2\times\lvert R_{550}\text{-}R_{670}\rvert]$	(Vincini et al., 2006)
Triangular vegetation index (TVI)	27	$0.5\times[120\times(R_{750}\text{-}R_{550})\text{-}200\times(R_{670}\text{-}R_{550})]$	(Broge and Leblanc, 2001)
Transformed chlorophyll absorption ratio index (TCARI)	28	$3\times[(R_{710}\text{-}R_{680})\text{-}0.2\times(R_{700}\text{-}R_{560})(R_{710}/R_{680})]$	(Haboudane et al., 2002)
Visible atmospherically resistance index (VARI)	29	$(R_{555}\text{-}R_{680})/(R_{555}+R_{680}\text{-}R_{480})$	(Gitelson et al., 2003)
Wodified wide dynamic range vegetation index (WDRVI)	30	$(0.1\times R_{800}\text{-}R_{670})/(0.1\times R_{800}+R_{670})$	(Stamatiadis et al., 2006)

samples were obtained at three growth stages of potato in 2019. The box line plot of potato LAI is shown in Figure 3. With the dynamic change of potato growth stage, the mean LAI value first decreased and then levelled off. During the tuber formation stage, the LAI ranged from 0.43 to 3.29, with an average of 1.50. During the tuber growth stage, the LAI ranged from 0.34 to 3.00, with an average of 1.32. During the starch accumulation stage, the LAI varied from 0.29 to 3.75, with an average of 1.28.

3.2 Correlation of LAI against hyperspectral image-derived features

The correlation coefficients between different types of characterization variables and the LAI of potatoes across multiple growth stages are presented in Figures 4, 5. LAI exhibited a highly significant correlation with OSR in the wavelength ranges of 454–702 nm and 718–950 nm ($p< 0.01$), with a particularly strong

10.3389/fpls.2024.1492372

TABLE 2 The Haralick texture used in the study.

Type of textures	Number	Names
simple	1	Energy
	2	Entropy
	3	Correlation
	4	Inverse difference moment
	5	Inertia
	6	Cluster shade
	7	Cluster prominence
	8	Haralick correlation
advanced	9	Mean
	10	Sum of squares: variance
	11	Dissimilarity
	12	Sum average
	13	Sum variance
	14	Sum entropy
	15	Difference variance
	16	Difference entropy
	17	Information measures of correlation IC1
	18	Information measures of correlation IC2
higher order	19	Short run emphasis
	20	Long run emphasis
	21	Grey-level nonuniformity
	22	Run length nonuniformity
	23	Run percentage
	24	Low grey-Level run emphasis
	25	High grey-Level run emphasis
	26	Short run low grey-level emphasis
	27	Short run high grey-level emphasis
	28	Long run low grey-level emphasis

correlation observed at 750–922 nm (Figure 4A). The correlation between LAI and FDSR showed greater variability after the first-order derivation of OSR, with highly significant correlations occurring at 454–506 nm, 550–642 nm, 690–774 nm, and 790–950 nm ($p< 0.01$). The highest correlation between LAI and FDSR was noted at 734–758 nm (Figure 4B). The VIs selected in this study also demonstrated highly significant correlations with LAI across multiple growth stages of potatoes (Figure 5A). Additionally, the extracted Haralick textures, including Inverse Difference Moment, Cluster Shade, Haralick Correlation, Mean, Sum Average, Low Grey-Level Run Emphasis, Short Run High Grey-Level Emphasis, and Long Run Low Grey-Level Emphasis, showed highly significant correlations with LAI (Figure 5B).

The variables with the highest correlation to LAI across multiple growth stages of potatoes were 886 nm for OSR (r = 0.55, $p< 0.01$), 742 nm for FDSR (r = 0.67, $p< 0.01$), NDRE for VIs (r = 0.65, $p< 0.01$), and the Mean for Haralick texture (r = -0.50, $p< 0.01$). The mean absolute correlation coefficients of OSR (450–950 nm), FDSR (450–950 nm), VIs, and Haralick texture features with LAI were 0.48, 0.35, 0.53, and 0.17, respectively. These results suggest that VIs may better capture the spatial variability of potato LAI, whereas Haralick texture features may be less effective.

3.3 Sensitive feature acquisition using SPA method

The results of filtering five feature datasets using the SPA method are presented in Table 3, including OSR, FDSR, VIs, Haralick textures, and their combined features. From a total of 125 OSR features, 125 FDSR features, 30 VIs, 28 textures, and 308 combined features, 10, 7, 8, 10, and 17 features were selected, respectively. This selection indicates that the unselected features were either uninformative or contained redundant information for estimating LAI across multiple growth stages of potatoes.

Although the selected wavelengths for LAI estimation using OSR and FDSR differ, they share a common characteristic: these wavelengths span the visible to near-infrared regions and exhibit a high correlation with LAI. Typically, VIs containing near-infrared or red-edge wavelengths are chosen for LAI estimation. Three features were selected from simple textures, five from advanced textures, and two from higher-order textures; however, not all of these features achieved a high correlation with LAI. For instance, features such as Inertia, Sum of squares: variance, Difference entropy, and Information measures of correlation IC1 did not fully reach the highly significant correlation level with LAI. This suggests that the SPA method helps to prevent the omission of features that significantly contribute to LAI estimation. In total, four OSR, seven FDSR, three VIs, and three textures were selected from the combined features, all of which remained highly sensitive to LAI, except for Inertia, which did not achieve a highly significant correlation. These findings indicate that the integration of different types of features is crucial for accurately estimating LAI across multiple growth stages in potatoes.

3.4 Estimation of LAI in multiple growth stages of potato

The calibration and validation results obtained using PLSR, RFR, and GPR based on the variables screened in Table 3 are presented in Figure 6. X1, X2, X3, X4 and X5 in Figure 6 represent OSR, FDSR, VIs, Haralick textures and All. Among the models, the LAI estimation model constructed using the GPR method demonstrated higher fitting accuracy and lower error compared

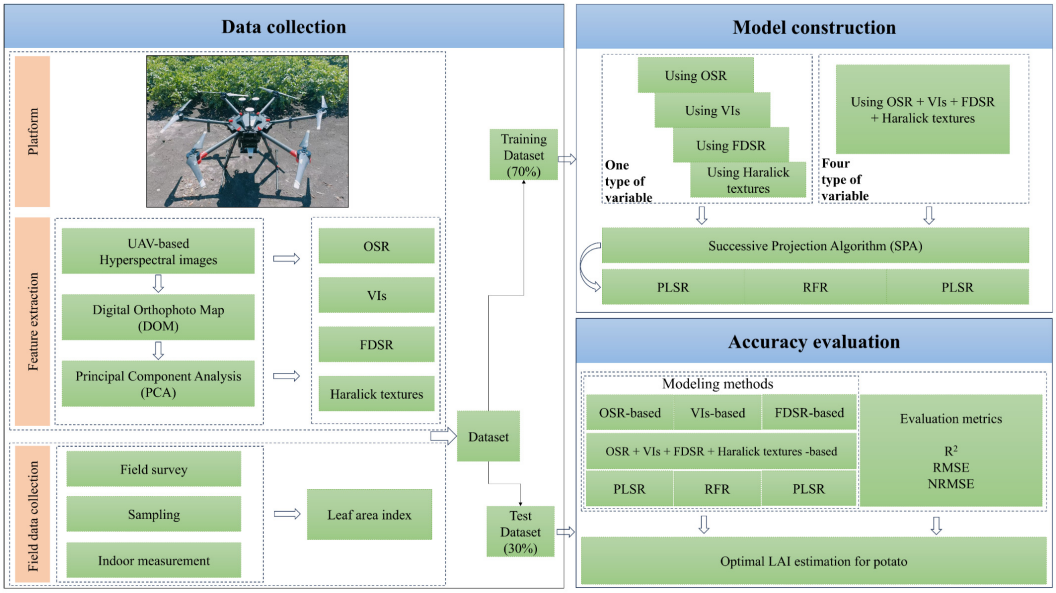


FIGURE 2
Flowchart of the experimental research.

to the others. When using the same regression method, LAI estimation based on different types of spectral information was more accurate than that based on Haralick texture features, with the best results achieved using the GPR method with VIs (Calibration: R^2 = 0.63, RMSE = 0.41, NRMSE = 31.32%; Validation: R^2 = 0.63, RMSE = 0.34, NRMSE = 22.84%). Although using Haralick texture features alone to estimate LAI across multiple growth stages of potatoes is not ideal, combining texture features with different forms of spectral information can enhance the accuracy of LAI estimation, suggesting that Haralick texture can compensate for the limitations of spectral information. Compared to the PLSR and RFR methods, the GPR method was the most accurate for estimating multi-stage LAI in potatoes using 17 features that included both spectral and texture information (Calibration: R^2 = 0.68, RMSE = 0.38, NRMSE = 29.04%; Validation: R^2 = 0.70, RMSE = 0.30, NRMSE = 20.28%).

The fitted relationship between the estimated and measured LAI values of potatoes across multiple growth stages, obtained using the best regression method based on four types of spectral information, one type of texture, and a combination of spectral and texture features, is illustrated in Figure 7. The results indicate the following: (1) Estimating LAI using only spectral information

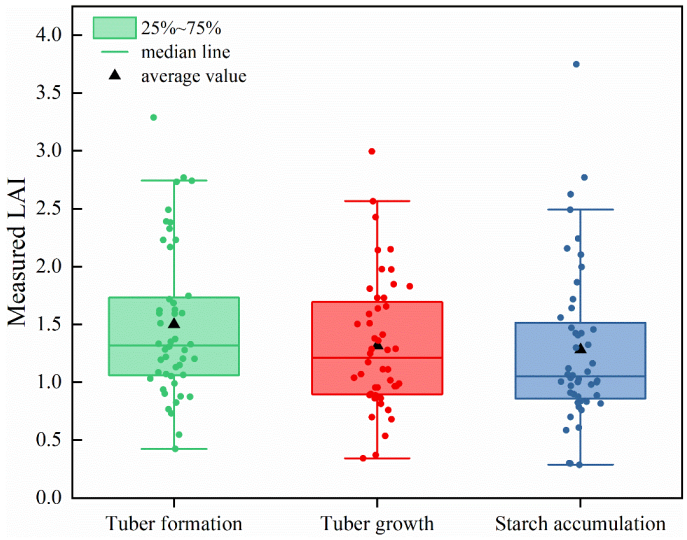


FIGURE 3
Box line plot of measured leaf area index at each growth stage.

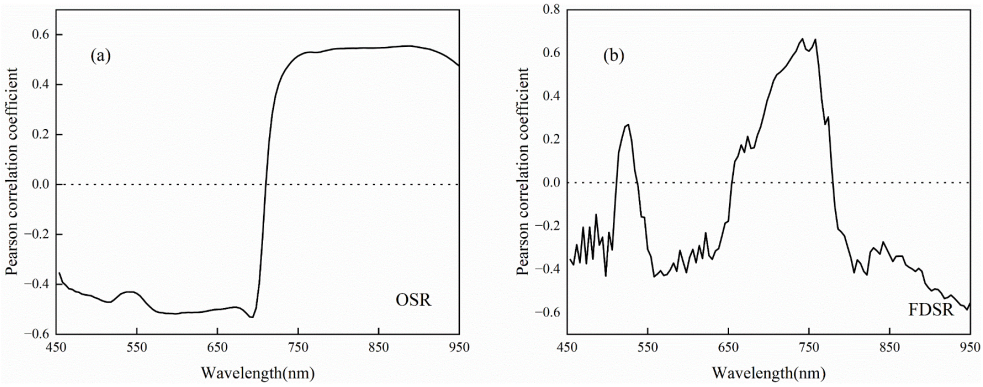


FIGURE 4
Pearson's correlation coefficient of LAI with OSR and FOD: **(A)** OSR **(B)** FOD.

or Haralick textures yields significant errors, particularly with Haralick textures, leading to low model accuracy (R^2 = 0.45, RMSE = 0.41, NRMSE = 27.60%). (2) Among the spectral information, VIs demonstrated the best estimation performance (R^2 = 0.63, RMSE = 0.34, NRMSE = 22.84%). (3) The combination of spectral information and texture features reduced the error in LAI estimation across multiple growth stages and enhanced the accuracy of the model. Notably, when LAI was greater than 2, the estimates based on the fused sensitive features were closer to the 1:1 line with the measured values (R^2 = 0.70, RMSE = 0.30, NRMSE = 20.28%), suggesting that integrating multiple types of features can mitigate the inaccuracies associated with using a single type of feature for LAI estimation.

3.5 Applicability of estimated LAI for different growth stages

The study combines three types of spectra and Haralick textures with the GPR machine learning method to estimate LAI in two potato varieties. Figure 8 illustrates the effect of this estimation to assess the model's applicability across different growth stages. The results showed that the performance of the method proposed in this study differed significantly in the estimation of LAI for the three growth stages, as evidenced by the poorer performance of the model in terms of accuracy in tuber growth (Figure 8B). This is due to the high canopy cover of potato in tuber growth compared to the other two growth stages. Spectral information saturation is more severe and textural information contribution is reduced. In contrast, the accuracy of the model was satisfactory in all three growth stages, although it still varied, so the method of combining different types of spectral information with Haralick textures still has potential for predicting potato LAI in different growth stages.

3.6 Mapping potato LAI

The optimal estimation model of potato LAI for multiple growth stages was constructed based on the three spectral information and Haralick textures, and the model was used to generate the spatial distribution map of potato LAI in the study area, as shown in Figure 9. This method helps to provide a more comprehensive understanding of the growth conditions and spatial distribution pattern in the study area. Analyzing LAI spatial distribution maps aids in assessing plant water use efficiency, optimizing irrigation and fertilizer strategies, and enhancing

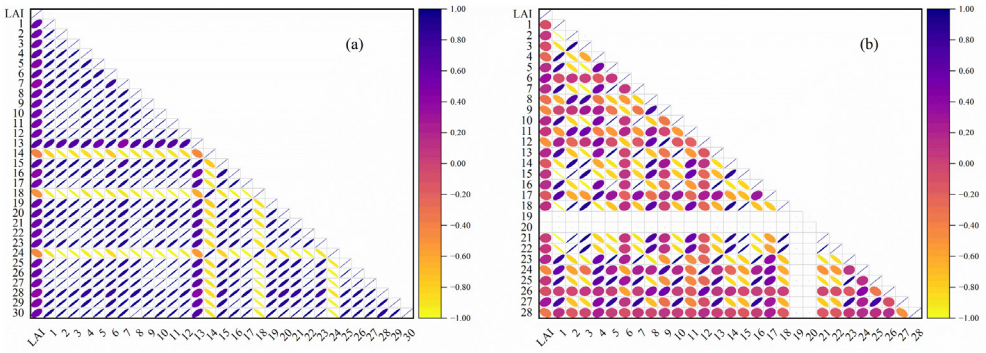


FIGURE 5
Pearson's correlation coefficient of LAI with VIs and Haralick textures. **(A)** VIs, **(B)** Haralick textures. The numbers in Panel **(A)** match the Number column in Table 1, and the numbers in Panel **(B)** match the Number column in Table 2.

TABLE 3 Sensitive features were obtained based on different datasets using the SPA method.

Type of features	Features
OSR	462, 474, 514, 542, 594, 674, 694, 714, 798, and 914 nm
FDSR	466, 486, 502, 506, 670, 778, and 946 nm
VIs	EVI, GNDVI, NDRE, NRI, PSRI, PSND, PBI, RVSI
Haralick textures	Inverse difference moment, Inertia, Cluster shade, Mean, Sum of squares: variance, Sum average, Difference entropy, Information measures of correlation IC1, Low grey-level run emphasis, Long run low grey-level emphasis
All	650, 710, 886, and 922 nm (OSR); 478, 530, 586, 630, 702, 738, and 770 nm (FDSR); GNDVI, NDRE, TCARI; Inertia, Cluster shade, Short run high grey-level emphasis

overall water and fertilizer efficiency. Additionally, it supports early detection of pests and diseases, minimizing yield loss, and serves as a preliminary tool for predicting crop yields, ultimately improving agricultural production and management.

4 Discussion

Miniature UAV-imaging hyperspectral systems not only acquire spectral reflectance information of the crop canopy but also capture image data of the canopy structure, enabling rapid and non-destructive monitoring of crop LAI (Yan et al., 2019; Verma et al., 2022). Estimating crop LAI across multiple growth stages using only the optical information of the crop canopy is challenging, as the performance of optical data can be inconsistent due to the influence of phenology and crop variety. Given the variations in potato canopy structure and the advantages of hyperspectral imaging, this study explored the applicability and potential of combining Haralick texture features with canopy spectral information for LAI estimation.

4.1 Response of spectral and texture features to LAI

The spectral reflectance of the potato canopy reflects the crop's physiological and geometrical characteristics, while VIs enhance the contrast between crop and soil, allowing for the quantitative description of crop growth. This spectral information is frequently used to estimate crop growth parameters such as LAI (Kayad et al., 2022). The three types of spectra used in this study varied significantly in their sensitivity to LAI across multiple growth stages of potatoes (Figures 4A, B, 5A). After differential transformation, FDSR reduced the effects of soil background and overlapping signal interference compared to OSR, thereby enhancing the sensitivity of spectral signals to LAI across different growth stages (Figures 4A, B). This finding is consistent with the results obtained by Zhang et al. (2022a) using differentially transformed spectra to estimate the LAI of rice. The VIs selected in this study exhibited higher sensitivity to LAI across multiple growth stages than OSR and FDSR. This increased sensitivity is due to the combination of multiple

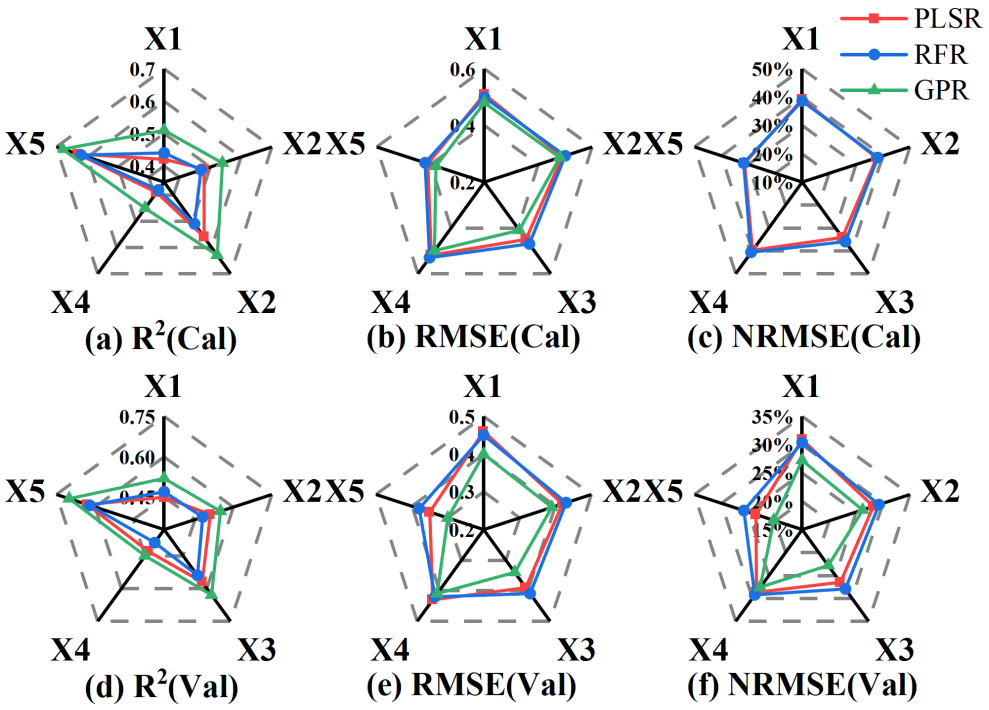


FIGURE 6
The Accuracy of LAI estimation based on different types of variables. **(A)** R^2(Cal), **(B)** RMSE(Cal), **(C)** NRMSE(Cal), **(D)** R^2(Val), **(E)** RMSE(Val), **(F)** NRMSE (Val). X1: OSR, X2: FDSR, X3: VIs, X4: Haralick textures and X5: All features.

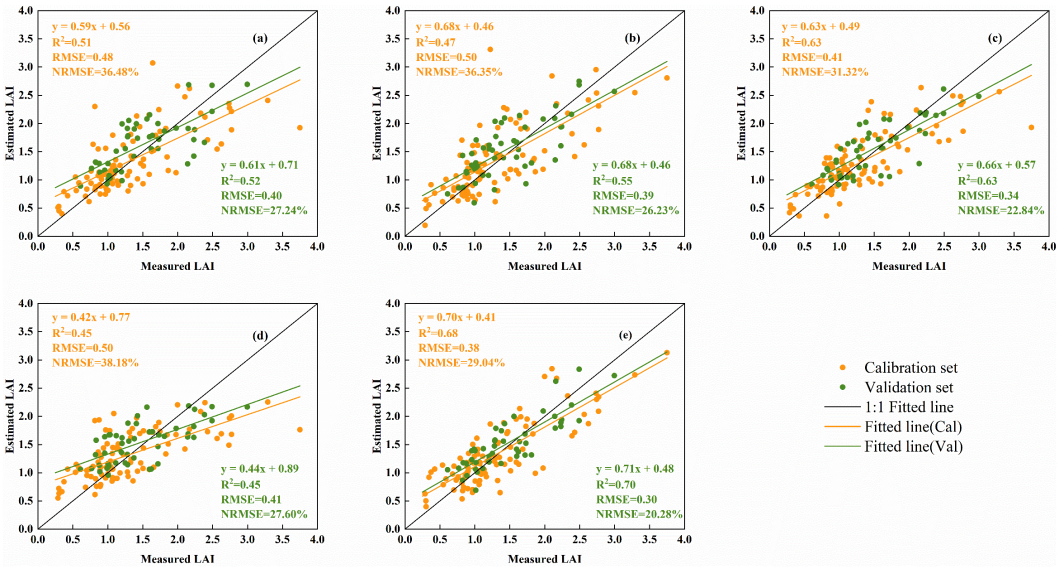


FIGURE 7
Fitting relationship between measured and estimated LAI based on different types of spectral information and Haralick textures. **(A)** OSR, **(B)** FDSR, **(C)** VIs, **(D)** Haralick textures, and **(E)** spectral information combined with texture features.

bands, which can mitigate the interference of external environmental signals on LAI response (Figure 5A). However, previous studies by Liu et al. (2022a) and Shi et al. (2022) have shown that the sensitivity of VIs to LAI across different growth stages can be affected by climatic variations and varietal differences, potentially limiting the accuracy of LAI estimation models. Spatial information such as texture of high-resolution UAV remote sensing data correlates with the spatial structure of the crop, reflecting more effectively the shading and internal structural information of the crop canopy. Given the importance of texture information in describing changes in canopy structure, this study attempted to use Haralick texture features to estimate LAI. The application of Haralick texture in the estimation of crop physicochemical parameters has not been extensively investigated. It remains unclear whether the sensitive bands of texture features align with the spectral bands. To address this, we employed the PCA downscaling method, which preserves the majority (>90%) of the image information and offers a robust foundation for subsequent studies. However, compared to canopy spectral information, the simple, advanced, and higher order textures did not effectively capture changes in LAI across multiple growth stages due to the low sensitivity of Haralick texture to potato LAI (Figure 5B), which is similar to the previous work of Qiao et al. (2024). The extraction of texture features from an image is often affected by a variety of factors, such as noise in the image and changes in lighting conditions, which may interfere with the stability of the texture features. This limitation may be related to the size of the window used for texture extraction and the calculation method employed. Consequently, we conclude that Haralick texture alone does not adequately reflect changes in LAI across multiple growth stages of potatoes.

4.2 Optimization of model parameters

To reduce the number of input parameters for the LAI estimation model and minimize errors due to subjective variable selection, the feature variables obtained using the SPA algorithm are presented in Table 3. For both OSR and FDSR, the characteristic

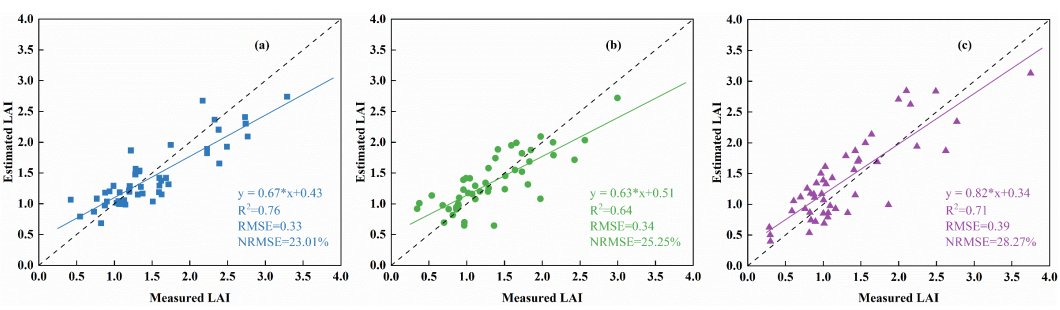


FIGURE 8
Applicability of estimated LAI for different growth stages based on the optimal model. **(A)** Tuber formation **(B)** Tuber growth **(C)** Starch accumulation.

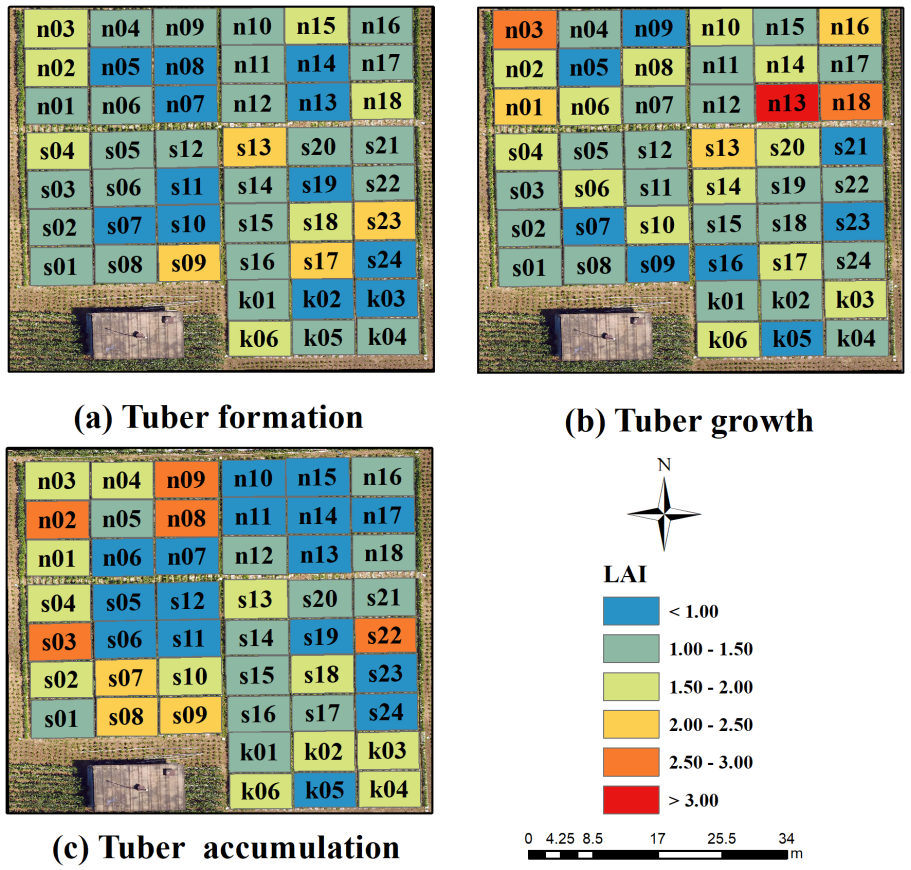


FIGURE 9
Spatial distribution of LAI in potato at different growth stages. **(A)** Tuber formation, **(B)** Tuber growth, **(C)** Tuber accumulation.

wavelengths identified after SPA screening, although not identical, fall within the visible to near-infrared regions. Since LAI is determined by plant leaves, and the absorption and reflection of pigments within these leaves, along with scattering between leaf tissues, influence the crop canopy spectra across the visible to near-infrared ranges, the differences in the sensitive spectral positions of various transformed spectral data and LAI are expected to be minimal (Zhou et al., 2022). The eight VIs selected through screening all include red-edge or near-infrared wavelengths, as these spectral regions are more responsive to changes in LAI (Figures 2A, B), a finding consistent with studies by Tao et al. (2020) and Liu et al. (2022a) in monitoring winter wheat LAI. The fact that not all of the 10 variables selected from the 28 Haralick textures achieved a highly significant correlation with potato LAI suggests that the variables identified by the SPA algorithm are objectively determined, allowing for the retention of those that significantly contribute to LAI estimation and thereby avoiding the large estimation errors that can result from subjective variable selection. To evaluate the performance of combining spectral and texture information for LAI estimation, four OSR, seven FDSR, three VIs, and three textures were selected from 308 combined features, underscoring the importance of feature fusion in accurately estimating LAI in potatoes across multiple growth stages (Table 3).

4.3 Performance evaluation of LAI estimation model

Compared to traditional algorithms, machine learning regression models can make effective use of data when dealing with complex data and obtain higher model prediction accuracy. The modeling and validation results obtained using the GPR method for the variables identified in Table 3 were superior to those achieved with the PLSR and RFR methods. This finding aligns with the results reported by Caballero et al. (2022) for estimating LAI in winter wheat. The superior performance of GPR may be attributed to its ability to effectively handle nonlinear relationships between independent and dependent variables, as well as its suitability for small sample datasets (Upreti et al., 2019; Zhang et al., 2022b). In this study, the modeling and validation datasets were small, and the spectral and texture information extracted from UAV hyperspectral imagery likely exhibited a stronger nonlinear relationship with LAI, making GPR a fitting choice for constructing the LAI estimation model. Analyzing the results in Table 3, the accuracy of LAI estimation across multiple growth stages of potatoes was higher when using the three forms of spectral information compared to Haralick texture alone. This is consistent with the findings of Zhang et al. (2022b) in estimating LAI for maize, where the grey-scale covariance matrix texture was

less effective than crop canopy spectral information. Among the spectral information types, the order of performance in estimating LAI was VIs, FDSR, and OSR, which corresponds to the sensitivity of these spectral types to LAI (Figures 4A, B, 5A). Although Haralick texture alone was not effective for estimating potato LAI (Table 3, Figure 3), combining it with crop canopy spectral features improved the accuracy of LAI estimation, consistent with Zhou et al (Zhou et al., 2022), who combined wavelet texture with spectral features for estimating rice LAI. Further illustrating that different types of feature coupling provide unique and complementary information with great potential in crop monitoring (Ochiai et al., 2024). Yu et al. (2023) also demonstrated that combining different types of features can improve the prediction performance of LAI in potato. This study highlights the limitations of Haralick texture, but also demonstrates its value as auxiliary information when combined with spectral features for LAI estimation. The combination of spectral and Haralick texture information in this study explained 69% of the variability in LAI across multiple growth stages in potatoes, which is lower than the 85% of spatial variability in maize LAI explained by Zhang et al. (2022b) using fusion information. This discrepancy may be due to differences in crop variety and canopy structure.

4.4 Advantages and limitations of research

Haralick texture differs from spectral information in response to potato canopy structure, and their complementary information helps to improve LAI estimation accuracy. This complementary information can enhance the accuracy of LAI estimation. In this study, models for LAI estimation that integrated both spectral and texture data outperformed those relying on a single type of data. Thus, combining spectral information with Haralick textures offers valuable insights for LAI estimation across various growth stages of potatoes. The study derived the spatial and temporal distribution of potato LAI at the field scale using the optimal estimation model (Figure 9), facilitating the assessment of the model's potential for practical field applications. Despite achieving satisfactory results, some limitations remain. Future research should consider incorporating additional crop structural characteristics, such as vegetation cover, plant height, and volume, to improve model construction. This study compared the potential of three machine learning methods for UAV estimation of LAI in wheat, but found that deep learning performed better for potato biomass prediction (Liu et al., 2024c), suggesting that deep learning has greater potential for application in the field of UAV spectral monitoring of crop growth. Therefore, future research will deeply explore the application of methods such as deep learning in this field. In addition, since the UAV and ground data used in this study were limited to a single location, within one year, and two varieties, we will further explore the impact of multi-year and multi-variety potato data on improving the performance of Haralick texture features. Future work will focus on different varieties and multiple years of data to analyze the robustness of canopy spectra and Haralick texture features in estimating potato LAI.

5 Conclusions

To accurately and timely estimate potato LAI, this study utilized UAV hyperspectral image data to extract three types of spectral information and Haralick textures, and analyzed their combined ability to estimate potato LAI. The effectiveness of this method was validated using plot test data, yielding the following results:

1. Using the SPA feature selection method, 17 sensitive features were identified as significant contributors to potato LAI, including four OSR wavelengths (650, 710, 886, and 922 nm), seven FDSR wavelengths (478, 530, 586, 630, 702, 738, and 770 nm), three VIs (GNDVI, NDRE, TCARI), and three Haralick textures (Inertia, Cluster shade, Short run high grey-level emphasis).
2. Among the single-type feature models, VIs demonstrated the best performance. In contrast, the regression performance of Haralick textures was less effective when used independently.
3. The accuracy of potato LAI estimation improved when combining features as input parameters, achieving an R^2 of 0.70 and an RMSE of 0.30. This model outperformed the optimal VIs model based on univariate parameters (R^2 = 0.63, RMSE = 0.34), suggesting that the combination of spectral information and Haralick textures provides better accuracy for LAI estimation and offers a feasible approach for monitoring potato growth.

Combining different types of spectral information and Haralick textures from UAV hyperspectral imagery can improve the estimation of LAI for potatoes with multiple growth stages to some extent compared to using a single type of feature. However, UAV multispectral data can provide higher spatial resolution and may be more suitable for texture feature extraction. Therefore, an attempt was made to validate the feasibility of this study's method using low-cost UAV multispectral in future studies.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

JF: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Methodology, Software, Validation. YL: Methodology, Software, Writing – review & editing. YF: Data curation, Software, Writing – review & editing. YY: Project administration, Writing – review & editing. RC: Data curation, Investigation, Writing – review & editing. MB: Data curation, Project administration, Writing – review & editing.

YM: Data curation, Visualization, Writing – review & editing. HF: Conceptualization, Data curation, Funding acquisition, Resources, Supervision, Writing – original draft, Writing – review & editing. HW: Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Key scientific and technological projects of Heilongjiang province (2021ZXJ05A05), the National 433 Natural Science Foundation of China (41601346), and the Key Field Research and Development Program of Guangdong Province 434 (2019B020216001).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahamed, T., Tian, L., Zhang, Y., and Ting, K. (2011). A review of remote sensing methods for biomass feedstock production. *Biomass Bioenergy* 35, 2455–2469. doi: 10.1016/j.biombioe.2011.02.028

Ahmed, M., Ahmad, S., Abbas, G., Hussain, S., and Hoogenboom, G. (2024). "Potato-potato system," in *Cropping Systems Modeling Under Changing Climate*. Eds. M. Ahmed, S. Ahmad, G. Abbas, S. Hussain and G. Hoogenboom (Springer Nature Singapore, Singapore), 271–306. doi: 10.1007/978-981-97-0331-9_10

Apan, A., Held, A., Phinn, S., and Markley, J. (2003). "Formulation and assessment of narrow-band vegetation indices from EO-1 Hyperion imagery for discriminating sugarcane disease," in *Proceedings of the 2003 Spatial Sciences Institute Biennial Conference (SSC2003): Spatial Knowledge Without Boundaries*. (Canberra, Australia: Spatial Sciences Institute of Australia), 1-13. doi: 10.1016/j.fecs.2024.100239

Badr, M. A., El-Tohamy, W. A., Salman, S. R., and Gruda, N. (2022). Yield and water use relationships of potato under different timing and severity of water stress. *Agric. Water Manage.* 271, 107793. doi: 10.1016/j.agwat.2022.107793

Barnes, J. D., Balaguer, L., Manrique, E., Elvira, S., and Davison, A. (1992). A reappraisal of the use of DMSO for the extraction and determination of chlorophylls a and b in lichens and higher plants. *Environ. Exp. Bot.* 32, 85–100. doi: 10.1016/0098-8472(92)90034-Y

Blackburn, G. A. (1998). Quantifying chlorophylls and caroteniods at leaf and canopy scales: an evaluation of some hyperspectral approaches. *Remote Sens. Environ.* 66, 273–285. doi: 10.1016/S0034-4257(98)00059-5

Broge, N. H., and Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sens. Environ.* 76, 156–172. doi: 10.1016/S0034-4257(00)00197-8

Brynolfsson, P., Nilsson, D., Torheim, T., Asklund, T., Karlsson, C. T., Trygg, J., et al. (2017). Haralick texture features from apparent diffusion coefficient (ADC) MRI images depend on imaging and pre-processing parameters. *Sci. Rep.* 7, 4041. doi: 10.1038/s41598-017-04151-4

Caballero, G., Pezzola, A., Winschel, C., Casella, A., Sanchez Angonova, P., Orden, L., et al. (2022). Quantifying irrigated winter wheat LAI in Argentina using multiple sentinel-1 incidence angles. *Remote Sens.* 14, 5867. doi: 10.3390/rs14225867

Cao, Y., Li, G. L., Luo, Y. K., Pan, Q., and Zhang, S. Y. (2020). Monitoring of sugar beet growth indicators using wide-dynamic-range vegetation index (WDRVI) derived from UAV multispectral images. *Comput. Electron. Agric.* 171, 105331. doi: 10.1016/j.compag.2020.105331

Chappelle, E. W., Kim, M. S., and Mcmurtrey, J. E. (1990). "Ratio analysis of reflectance spectra," in *10th Annual International Symposium on Geoscience and Remote Sensing*. New York, USA: IEEE (Institute of Electrical and Electronics Engineers), 1605–1608.

Chen, J. M. (1996). Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Can. J. Remote Sens.* 22, 229–242. doi: 10.1080/07038992.1996.10855178

Chen, R., Liu, W., Yang, H., Jin, X., Yang, G., Zhou, Y., et al. (2024). A novel framework to assess apple leaf nitrogen content: Fusion of hyperspectral reflectance and phenology information through deep learning. *Comput. Electron. Agric.* 219, 108816. doi: 10.1016/j.compag.2024.108816

Cohrs, C. W., Cook, R. L., Gray, J. M., and Albaugh, T. J. (2020). Sentinel-2 leaf area index estimation for pine plantations in the Southeastern United States. *Remote Sens.* 12, 1406. doi: 10.3390/rs12091406

Croft, H., Chen, J. M., and Zhang, Y. (2014). The applicability of empirical vegetation indices for determining leaf chlorophyll content over different leaf and canopy structures. *Ecol. Complexity* 17, 119–130. doi: 10.1016/j.ecocom.2013.11.005

Das, D., and Naskar, R. (2024). Image splicing detection using low-dimensional feature vector of texture features and Haralick features based on Gray Level Co-occurrence Matrix. *Signal Processing: Image Communication* 125, 117134. doi: 10.1016/j.image.2024.117134

Datt, B. (1999). A new reflectance index for remote sensing of chlorophyll content in higher plants: tests using eucalyptus leaves. *J. Plant Physiol.* 154, 30–36. doi: 10.1016/S0176-1617(99)80314-9

Dong, T., Liu, J., Shang, J., Qian, B., Ma, B., Kovacs, J. M., et al. (2019). Assessment of red-edge vegetation indices for crop leaf area index estimation. *Remote Sens. Environ.* 222, 133–143. doi: 10.1016/j.rse.2018.12.032

Duan, B., Liu, Y., Gong, Y., Peng, Y., Wu, X., Zhu, R., et al. (2019). Remote estimation of rice LAI based on Fourier spectrum texture from UAV image. *Plant Methods* 15, 124. doi: 10.1186/s13007-019-0507-8

Fan, Y., Feng, H., Jin, X., Yue, J., Liu, Y., Li, Z., et al. (2022). Estimation of the nitrogen content of potato plants based on morphological parameters and visible light vegetation indices. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1012070

Fitzgerald, G., Rodriguez, D., and O'Leary, G. (2010). Measuring and predicting canopy nitrogen nutrition in wheat using a spectral index—The canopy chlorophyll content index (CCCI). *Field Crops Res.* 116, 318–324. doi: 10.1016/j.fcr.2010.01.010

Fu, B., Sun, J., Wang, Y., Yang, W., He, H., Liu, L., et al. (2022). Evaluation of LAI estimation of mangrove communities using DLR and ELR algorithms with UAV, hyperspectral, and SAR images. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.944454

Gitelson, A. A., Gritz, Y., and Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* 160, 271–282. doi: 10.1078/0176-1617-00887

Gitelson, A. A., Merzlyak, M. N., and Lichtenthaler, H. K. (1996). Detection of red edge position and chlorophyll content by reflectance measurements near 700 nm. *J. Plant Physiol.* 148, 501–508. doi: 10.1016/S0176-1617(96)80285-9

Gong, Y., Yang, K., Lin, Z., Fang, S., Wu, X., Zhu, R., et al. (2021). Remote estimation of leaf area index (LAI) with unmanned aerial vehicle (UAV) imaging for different rice cultivars throughout the entire growing season. *Plant Methods* 17, 88. doi: 10.1186/s13007-021-00789-4

Gu, J., Su, Z., Gao, R., Wang, Y., Meng, Y., and Kong, Q. (2022). Improving QGA-ELM inversion model of rice leaf area index based on UAV remote sensing image. *Mobile Inf. Syst.* 2022, 9658966. doi: 10.1155/2022/9658966

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., and Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* 90, 337–352. doi: 10.1016/j.rse.2003.12.013

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., and Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ.* 81, 416–426. doi: 10.1016/S0034-4257(02)00018-4

Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Trans. Systems Man Cybernetics SMC* 3, 610–621. doi: 10.1109/TSMC.1973.4309314

Hu, P., Chapman, S. C., Jin, H., Guo, Y., and Zheng, B. (2021). Comparison of modelling strategies to estimate phenotypic values from an unmanned aerial vehicle with spectral and temporal vegetation indexes. *Remote Sens.* 13, 2827. doi: 10.3390/rs13142827

Huete, A. (1988). A soil vegetation adjusted index (SAVI). *Remote Sens. Environ.* 25, 295–309. doi: 10.1080/07038992.1996.10855178

Ilniyaz, O., Kurban, A., and Du, Q. (2022). Leaf area index estimation of pergola-trained vineyards in arid regions based on UAV RGB and multispectral data using machine learning methods. *Remote Sens.* 14, 415. doi: 10.3390/rs14020415

Imanishi, J., Sugimoto, K., and Morimoto, Y. (2004). Detecting drought status and LAI of two Quercus species canopies using derivative spectra. *Comput. Electron. Agric.* 43, 109–129. doi: 10.1016/j.compag.2003.12.001

Ji, S., Gu, C., Xi, X., Zhang, Z., Hong, Q., Huo, Z., et al. (2022). Quantitative monitoring of leaf area index in rice based on hyperspectral feature bands and ridge regression algorithm. *Remote Sens.* 14, 2777. doi: 10.3390/rs14122777

Jiang, Z., Huete, A. R., Didan, K., and Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* 112, 3833–3845. doi: 10.1016/j.rse.2008.06.006

Jiang, X., Zhen, J., Miao, J., Zhao, D., Wang, J., and Jia, S. (2021). Assessing mangrove leaf traits under different pest and disease severity with hyperspectral imaging spectroscopy. *Ecol. Indic.* 129, 107901. doi: 10.1016/j.ecolind.2021.107901

Kaplan, G., and Rozenstein, O. (2021). Spaceborne estimation of leaf area index in cotton, tomato, and wheat using sentinel-2. *Land* 10, 505. doi: 10.3390/land10050505

Kayad, A., Rodrigues, F. A., Naranjo, S., Sozzi, M., Pirotti, F., Marinello, F., et al. (2022). Radiative transfer model inversion using high-resolution hyperspectral airborne imagery – Retrieving maize LAI to access biomass and grain yield. *Field Crops Res.* 282, 108449. doi: 10.1016/j.fcr.2022.108449

Kokubu, Y., Hara, S., and Tani, A. (2020). Mapping seasonal tree canopy cover and leaf area using worldview-2/3 satellite imagery: A megacity-scale case study in Tokyo urban area. *Remote Sens.* 12, 1505. doi: 10.3390/rs12091505

Li, B., Xu, X., Zhang, L., Han, J., Bian, C., Li, G., et al. (2020). Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *ISPRS J. Photogrammetry Remote Sens.* 162, 161–172. doi: 10.1016/j.isprsjprs.2020.02.013

Li, S., Yuan, F., Ata-UI-Karim, S. T., Zheng, H., Cheng, T., Liu, X., et al. (2019). Combining color indices and textures of UAV-based digital imagery for rice LAI estimation. *Remote Sens.* 11, 1763. doi: 10.3390/rs11151763

Li, X., Zhang, Y., Bao, Y., Luo, J., Jin, X., Xu, X., et al. (2014). Exploring the best hyperspectral features for LAI estimation using partial least squares regression. *Remote Sens.* 6, 6221–6241. doi: 10.3390/rs6076221

Li, X., Zhang, Y., Luo, J., Jin, X., Xu, Y., and Yang, W. (2016). Quantification winter wheat LAI with HJ-1CCD image features over multiple growing seasons. *Int. J. Appl. Earth Observation Geoinformation* 44, 104–112. doi: 10.1016/j.jag.2015.08.004

Liu, Y., Fan, Y., Feng, H., Chen, R., Bian, M., Ma, Y., et al. (2024a). Estimating potato above-ground biomass based on vegetation indices and texture features constructed from sensitive bands of UAV hyperspectral imagery. *Comput. Electron. Agric.* 220, 108918. doi: 10.1016/j.compag.2024.108918

Liu, Y., Fan, Y., Yue, J., Jin, X., Ma, Y., Chen, R., et al. (2024b). A model suitable for estimating above-ground biomass of potatoes at different regional levels. *Comput. Electron. Agric.* 222, 109081. doi: 10.1016/j.compag.2024.109081

Liu, Y., Feng, H., Yue, J., Fan, Y., Jin, X., Zhao, Y., et al. (2022b). Estimation of potato above-ground biomass using UAV-based hyperspectral images and machine-learning regression. *Remote Sens.* 14, 5449. doi: 10.3390/RS14215449

Liu, Y., Feng, H., Yue, J., Jin, X., Fan, Y., Chen, R., et al. (2024c). Improving potato AGB estimation to mitigate phenological stage impacts through depth features from hyperspectral data. *Comput. Electron. Agric.* 219, 108808. doi: 10.1016/j.compag.2024.108808

Liu, S., Hu, Z., Han, J., Li, Y., and Zhou, T. (2022a). Predicting grain yield and protein content of winter wheat at different growth stages by hyperspectral data integrated with growth monitor index. *Comput. Electron. Agric.* 200, 107235. doi: 10.1016/j.compag.2022.107235

Lu, Z., Deng, L., and Lu, H. (2022). An improved LAI estimation method incorporating with growth characteristics of field-grown wheat. *Remote Sens.* 14, 4013. doi: 10.3390/rs14164013

Ma, Y., Zhang, Q., Yi, X., Ma, L., Zhang, L., Huang, C., et al. (2022). Estimation of cotton leaf area index (LAI) based on spectral transformation and vegetation index. *Remote Sens.* 14, 136. doi: 10.3390/rs14010136

Merton, R., and Huntington, J. (1999). "Early simulation results of the ARIES-1 satellite sensor for multi-temporal vegetation research derived from AVIRIS," in *Proceedings of the eighth annual JPL airborne earth science workshop* (Pasadena, CA, USA: Jet Propulsion Lab., California Inst. of Tech), 9–11.

Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., and Rakitin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum* 106, 135–141. doi: 10.1034/j.1399-3054.1999.106119.x

Ochiai, S., Kamada, E., and Sugiura, R. (2024). Comparative analysis of RGB and multispectral UAV image data for leaf area index estimation of sweet potato. *Smart Agric. Technol.* 9, 100579. doi: 10.1016/j.atech.2024.100579

Peñuelas, J., Gamon, J. A., Fredeen, A. L., Merino, J., and Field, C. B. (1994). Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. *Remote Sens. Environ.* 48, 135–146. doi: 10.1016/0034-4257(94)90136-8

Qiao, D., Yang, J., Bai, B., Li, G., Wang, J., Li, Z., et al. (2024). Non-destructive monitoring of peanut leaf area index by combing UAV spectral and textural characteristics. *Remote Sens.* 16, 2182. doi: 10.3390/rs16122182

Raj, R., Walker, J. P., Pingale, R., Nandan, R., Naik, B., and Jagarlapudi, A. (2021). Leaf area index estimation using top-of-canopy airborne RGB images. *Int. J. Appl. Earth Observation Geoinformation* 96, 102282. doi: 10.1016/j.jag.2020.102282

Rama Rao, N., Garg, P. K., Ghosh, S. K., and Dadhwal, V. K. (2008). Estimation of leaf total chlorophyll and nitrogen concentrations using hyperspectral satellite imagery. *J. Agric. Sci.* 146, 65–75. doi: 10.1017/S0021859607007514

Revill, A., Florence, A., MacArthur, A., Hoad, S., Rees, R., and Williams, M. (2020). Quantifying uncertainty and bridging the scaling gap in the retrieval of leaf area index by coupling sentinel-2 and UAV observations. *Remote Sens.* 12, 1843. doi: 10.3390/rs12111843

Richardson, A. J., and Wiegand, C. (1977). Distinguishing vegetation from soil background information. *Photogrammetric Eng. Remote Sens.* 43, 1541–1552.

Roujean, J.-L., and Breon, F.-M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.* 51, 375–384. doi: 10.1016/0034-4257(94)00114-3

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA Spec. Publ* 351, 309.

Shi, Y., Gao, Y., Wang, Y., Luo, D., Chen, S., Ding, Z., et al. (2022). Using unmanned aerial vehicle-based multispectral image data to monitor the growth of intercropping crops in tea plantation. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.820585

Sone, C., Saito, K., and Futakuchi, K. (2009). Comparison of three methods for estimating leaf area index of upland rice cultivars. *Crop Sci.* 49, 1438–1443. doi: 10.2135/cropsci2008.09.0520

Stamatiadis, S., Taskos, D., Tsadilas, C., Christofides, C., Tsadila, E., and Schepers, J. S. (2006). Relation of ground-sensor canopy reflectance to biomass production and grape color in two merlot vineyards. *Am. J. Enology Viticulture* 57, 415–422. doi: 10.5344/ajev.2006.57.4.415

Sun, J., Zhou, X., Hu, Y., Wu, X., Zhang, X., and Wang, P. (2019). Visualizing distribution of moisture content in tea leaves using optimization algorithms and NIR hyperspectral imaging. *Comput. Electron. Agric.* 160, 153–159. doi: 10.1016/j.compag.2019.03.004

Tao, H., Feng, H., Xu, L., Miao, M., Long, H., Yue, J., et al. (2020). Estimation of crop growth parameters using UAV-based hyperspectral remote sensing data. *Sensors* 20, 1296. doi: 10.3390/s20051296

Upreti, D., Huang, W., Kong, W., Pascucci, S., Pignatti, S., Zhou, X., et al. (2019). A comparison of hybrid machine learning algorithms for the retrieval of wheat biophysical variables from sentinel-2. *Remote Sens.* 11, 481. doi: 10.3390/rs11050481

Verma, B., Prasad, R., Srivastava, P. K., Yadav, S. A., Singh, P., and Singh, R. K. (2022). Investigation of optimal vegetation indices for retrieval of leaf chlorophyll and leaf area index using enhanced learning algorithms. *Comput. Electron. Agric.* 192, 106581. doi: 10.1016/j.compag.2021.106581

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G. P. W., et al. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. *ISPRS J. Photogrammetry Remote Sens.* 108, 273–290. doi: 10.1016/j.isprsjprs.2015.05.005

Vincini, M., Frazzi, E., and D'Alessio, P. (2006). "Angular dependence of maize and sugar beet Vis from directional CHRIS/PROBA data," in *Fourth ESA CHRIS PROBA Workshop* (Frascati, Italy: ESRIN), 19–21.

Wengert, M., Piepho, H.-P., Astor, T., Graß, R., Wijesingha, J., and Wachendorf, M. (2021). Assessing spatial variability of barley whole crop biomass yield and leaf area index in silvoarable agroforestry systems using UAV-borne remote sensing. *Remote Sens.* 13, 2751. doi: 10.3390/rs13142751

Wulder, M. A., LeDrew, E. F., Franklin, S. E., and Lavigne, M. B. (1998). Aerial image texture information in the estimation of northern deciduous and mixed wood forest leaf area index (LAI). *Remote Sens. Environ.* 64, 64–76. doi: 10.1016/S0034-4257(97)00169-7

Xue, L., Cao, W., Luo, W., Dai, T., and Zhu, Y. (2004). Monitoring leaf nitrogen status in rice with canopy spectral reflectance. *Agron. J.* 96, 135–142. doi: 10.2134/agronj2004.1350

Yan, G., Hu, R., Luo, J., Weiss, M., Jiang, H., Mu, X., et al. (2019). Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives. *Agric. For. Meteorology* 265, 390–411. doi: 10.1016/j.agrformet.2018.11.033

Yang, H., Li, F., Wang, W., and Yu, K. (2021). Estimating above-ground biomass of potato using random forest and optimized hyperspectral indices. *Remote Sens.* 13, 2339. doi: 10.3390/rs13122339

Yu, T., Zhou, J., Fan, J., Wang, Y., and Zhang, Z. (2023). Potato leaf area index estimation using multi-sensor unmanned aerial vehicle (UAV) imagery and machine learning. *Remote Sens.* 15, 4108. doi: 10.3390/rs15164108

Yue, J., Feng, H., Jin, X., Yuan, H., Li, Z., Zhou, C., et al. (2018). A comparison of crop parameters estimation using images from UAV-mounted snapshot hyperspectral sensor and high-definition digital camera. *Remote Sens.* 10, 1138. doi: 10.3390/rs10071138

Yue, J., Yang, G., Tian, Q., Feng, H., Xu, K., and Zhou, C. (2019). Estimate of winter-wheat above-ground biomass based on UAV ultrahigh-ground-resolution image textures and vegetation indices. *ISPRS J. Photogrammetry Remote Sens.* 150, 226–244. doi: 10.1016/j.isprsjprs.2019.02.022

Zarco-Tejada, P. J., Berjón, A., López-Lozano, R., Miller, J. R., Martín, P., Cachorro, V., et al. (2005). Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sens. Environ.* 99, 271–287. doi: 10.1016/j.rse.2005.09.002

Zhang, J., Cheng, T., Guo, W., Xu, X., Qiao, H., Xie, Y., et al. (2021a). Leaf area index estimation model for UAV image hyperspectral data based on wavelength variable selection and machine learning methods. *Plant Methods* 17, 49. doi: 10.1186/s13007-021-00750-5

Zhang, T., Jiang, X., Jiang, L., Li, X., Yang, S., and Li, Y. (2022a). Hyperspectral reflectance characteristics of rice canopies under changes in diffuse radiation fraction. *Remote Sens.* 14, 285. doi: 10.3390/rs14020285

Zhang, J., Qiu, X., Wu, Y., Zhu, Y., Cao, Q., Liu, X., et al. (2021b). Combining texture, color, and vegetation indices from fixed-wing UAS imagery to estimate wheat growth parameters using multivariate regression methods. *Comput. Electron. Agric.* 185, 106138. doi: 10.1016/j.compag.2021.106138

Zhang, Y., Ta, N., Guo, S., Chen, Q., Zhao, L., Li, F., et al. (2022c). Combining spectral and textural information from UAV RGB images for leaf area index monitoring in kiwifruit orchard. *Remote Sens.* 14, 1063. doi: 10.3390/rs14051063

Zhang, X., Zhang, K., Sun, Y., Zhao, Y., Zhuang, H., Ban, W., et al. (2022b). Combining spectral and texture features of UAS-based multispectral images for maize leaf area index estimation. *Remote Sens.* 14, 331. doi: 10.3390/rs14020331

Zhou, C., Gong, Y., Fang, S., Yang, K., Peng, Y., Wu, X., et al. (2022). Combining spectral and wavelet texture features for unmanned aerial vehicles remote estimation of rice leaf area index. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.957870

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 December 2024
DOI 10.3389/fpls.2024.1504119

Check for updates

OPEN ACCESS

EDITED BY
Liujun Xiao,
Nanjing Agricultural University, China

REVIEWED BY
Jingying Fu,
Chinese Academy of Sciences (CAS), China
Rui Jiang,
South China Agricultural University, China

*CORRESPONDENCE
Wanjie Liang
wanjie.liang@163.com

RECEIVED 30 September 2024
ACCEPTED 29 October 2024
PUBLISHED 06 December 2024

CITATION
Sun C, Zhang W, Zhao G, Wu Q, Liang W, Ren N, Cao H and Zou L (2024) Mapping rapeseed (*Brassica napus L.*) aboveground biomass in different periods using optical and phenotypic metrics derived from UAV hyperspectral and RGB imagery.
Front. Plant Sci. 15:1504119.
doi: 10.3389/fpls.2024.1504119

COPYRIGHT
© 2024 Sun, Zhang, Zhao, Wu, Liang, Ren, Cao and Zou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Mapping rapeseed (*Brassica napus L.*) aboveground biomass in different periods using optical and phenotypic metrics derived from UAV hyperspectral and RGB imagery

Chuanliang Sun[1,2], Weixin Zhang[1], Genping Zhao[3], Qian Wu[1], Wanjie Liang[1]*, Ni Ren[1], Hongxin Cao[1] and Lidong Zou[4,5]

[1]Department of Digital Technology, Institute of Agriculture Information, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [2]School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand, [3]School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China, [4]Institute of Applied Artificial Intelligence of the Guangdong-Hongkong-Macao Greater Bay, Shenzhen, China, [5]School of Artificial Intelligence, Shenzhen Polytechnic University, Shenzhen, China

Aboveground biomass (AGB) is a key indicator of crop nutrition and growth status. Accurately and timely obtaining biomass information is essential for crop yield prediction in precision management systems. Remote sensing methods play a key role in monitoring crop biomass. However, the saturation effect makes it challenging for spectral indices to accurately reflect crop changes at higher biomass levels. It is well established that rapeseed biomass during different growth stages is closely related to phenotypic traits. This study aims to explore the potential of using optical and phenotypic metrics to estimate rapeseed AGB. Vegetation indices (VI), texture features (TF), and structural features (SF) were extracted from UAV hyperspectral and ultra-high-resolution RGB images to assess their correlation with rapeseed biomass at different growth stages. Deep neural network (DNN), random forest (RF), and support vector regression (SVR) were employed to estimate rapeseed AGB. We compared the accuracy of various feature combinations and evaluated model performance at different growth stages. The results indicated strong correlations between rapeseed AGB at the three growth stages and the corresponding indices. The estimation model incorporating VI, TF, and SF showed higher accuracy in estimating rapeseed AGB compared to models using individual feature sets. Furthermore, the DNN model (R^2 = 0.878, RMSE = 447.02 kg/ha) with the combined features outperformed both the RF (R^2 = 0.812, RMSE = 530.15 kg/ha) and SVR (R^2 = 0.781, RMSE = 563.24 kg/ha) models. Among the growth stages, the bolting stage yielded slightly higher estimation accuracy than the seedling and early

blossoming stages. The optimal model combined DNN with VI, TF, and SF features. These findings demonstrate that integrating hyperspectral and RGB data with advanced artificial intelligence models, particularly DNN, provides an effective approach for estimating rapeseed AGB.

KEYWORDS

rapeseed (*Brassica napus L.*), aboveground biomass (AGB), phenotypic metrics, hyperspectral images (HSI), machine learning approach

1 Introduction

Winter oilseed rape (*Brassica napus L.*) is one of the most important oil crops globally, with China accounting for about one-third of the world's cultivated area and one-fifth of total production (Liu et al., 2019). The Yangtze River basin is the primary growing region for this crop in China. In addition to providing essential oil products, cultivating oilseed rape offers benefits such as improving soil fertility and serving as a potential raw material for bioenergy. Therefore, efficiently and accurately monitoring rapeseed growth is crucial for enhancing both yield and quality. Timely estimation of AGB is particularly important for diagnosing nutrient deficiencies, guiding precise fertilization, and predicting yield outcomes.

AGB is a critical physiological indicator for monitoring crop growth and guiding agricultural management. It is closely linked to the crop's nutritional status and the ability of leaves and stems to absorb organic matter, making it a key variable in crop phenotyping (Araus and Cairns, 2014). Accurate AGB monitoring is essential for effective crop management, yield prediction, and ensuring food security through a stable supply (Li et al., 2020, Li Z. et al., 2022). However, traditional methods of estimating AGB—such as destructive field sampling followed by laboratory drying and weighing—are time-consuming and inefficient. These approaches do not meet the need for large-scale, high-throughput, timely, and quantitative monitoring, thereby limiting real-time AGB assessment at the field scale (Chang et al., 2017; Yue et al., 2018b; Zeng et al., 2018).

In the past decade, research on crop growth monitoring using UAV-mounted spectral platforms has emerged as a new direction in precision agriculture. Hyperspectral imaging and UAV technology have significantly improved the flexibility of predicting crop AGB (Tao et al., 2020; Yue et al., 2020, 2023), enabling the collection of crop phenotypic information at the field scale for growth monitoring (Clevers et al., 2017; Gitelson et al., 2003; Kooistra and Clevers, 2016). Previous studies have successfully utilized these technologies for biomass monitoring in crops such as rice, maize, barley, wheat, and grasslands (Bendig et al., 2014, 2015; Derraz et al., 2023; Shu et al., 2023; Sinde-González et al., 2021; Zhai et al., 2023). Compared to UAV-mounted RGB and multispectral sensors, hyperspectral sensors cover a broader range of spectral bands, allowing for a deeper investigation of crop physiological characteristics (Daughtry et al., 2000). This technology has shown promising results in yield prediction for soybean and disease monitoring for wheat (Banerjee et al., 2020; Guo et al., 2021; Herrero-Huerta et al., 2020. Therefore, the application of spectral imaging technology is essential for accurately detecting spatial variability in crop biochemical composition.

Previous studies on the estimation of rapeseed AGB were based on the methods of index and texture characteristics, however there were few studies on the estimation of rapeseed biomass in combination with structural parameters. Current research on estimating AGB using optical data primarily focuses on leveraging vegetation indices (VI) that are sensitive to dry matter content in crop canopies (Cheng et al., 2017; Hansen and Schjoerring, 2003; Itoh et al., 2006). VI capture changes in crop physiological activity and canopy structure, facilitating accurate AGB estimation (Wang et al., 2019). Several studies have validated effective VI for estimating AGB in crops such as winter wheat, maize, rice, and cotton using traditional regression methods. Commonly used indices include the Normalized Difference Vegetation Index (NDVI), Visible Atmospherically Resistant Index (VARI), Transformed Vegetation Index (TVI), and Red-edge Chlorophyll Index (Han et al., 2019; Ma et al., 2022; Pugh et al., 2018; Varela et al., 2017). However, due to differences in canopy structure and growth stages across species, the performance of VI can vary, making AGB estimation less reliable at different growth stages. For instance, before canopy closure, interactions with soil background and spectral saturation can reduce the accuracy of VI-based AGB estimates (Yao et al., 2018; Yue et al., 2017, 2023).

High-resolution imagery provides rich texture features at the plot level. Previous studies have successfully used texture features derived from satellite data to estimate AGB, particularly in forested areas (Zha et al., 2020; Zhang et al., 2015). Additionally, AGB prediction can be achieved using 3D data from UAV sensors, which combine height information from LiDAR or stereo imagery with spectral features from multispectral images (Muharam et al., 2014; Zhu et al., 2019). These studies have demonstrated that texture and structural parameters significantly improve crop monitoring accuracy. However, the high cost and weight of LiDAR sensors make routine crop growth monitoring challenging. In contrast, RGB photogrammetric sensors are lighter and more practical, making them a viable alternative. Therefore, exploring the potential of extracting crop structural parameters from RGB data is crucial for crop monitoring. In recent years, the integration of

artificial intelligence (AI) algorithms with optical data for AGB estimation has emerged as a promising new approach. AI methods are particularly advantageous for processing multi-dimensional data, effectively addressing the issue of data redundancy that is common in traditional regression models (Li R. et al., 2022; Verrelst et al., 2012; Volpato et al., 2021; Wang et al., 2016).

Currently, there is limited research on the accuracy of AGB estimation models for rapeseed using canopy spectral, texture, and structural information extracted from UAV-based hyperspectral and RGB images. To assess the potential of combining optical and phenotypic parameters for rapeseed AGB estimation, the objectives of this study are: (1) to extract optical and phenotypic metrics—VI, TF, and SF—from UAV hyperspectral and ultra-high-resolution RGB images to investigate the correlation between rapeseed biomass and these metrics at different growth stages. (2) to estimate rapeseed AGB using DNN, SVR, and RF. The study compares the accuracy of VI, TF, SF, and their various combinations, while also evaluating the performance of the three artificial intelligence approaches in estimating AGB across different growth stages.

2 Materials and methods

2.1 Experimental design

The study was conducted at the Smart Agriculture Research Base of the Jiangsu Academy of Agricultural Sciences, Nanjing, China, located at 32°02′34″N, 118°26′25″E (Figure 1A). Two varieties of rapeseed, Zheza 903 (C1) and Ningyou 26 (C2), were used in the study. During the 2022–2023 growing season, nitrogen fertilizer treatments were applied at rates of 0, 90, 180, 270, and 360 kg/ha, labeled as N0, N1, N2, N3, and N4, respectively. Each treatment was replicated three times in a randomized block design on a 600 m^2 test plot. Phosphorus and potassium fertilizers were applied at rates of 120 kg/ha P_2O_5, 180 kg/ha K_2O, and 15 kg/ha boron. Sowing took place on October 10, 2022, with transplanting on November 5, 2022, at a planting density of 1.125×10^5 plants/hm^2. Fertilizer was distributed as base fertilizer: wax fertilizer: moss fertilizer in a 5:3:2 ratio, while other cultivation practices followed high-yield field management (Figure 1B). During the crop growing season, data on average daily precipitation and minimum temperature were collected in the field (Figure 1C). The highest daily average precipitation (53 mm) was recorded in June 2023, while the highest daily temperature occurred in May. The lowest temperature was observed in January.

2.2 Data acquisition

2.2.1 Field data acquisition

AGB and plant height (PH) measurement datasets were collected for rapeseed at the seedling stage (December 28, 2022), bolting stage (February 20, 2023), and early blossoming stage (March 2, 2023). Collection of AGB data: To ensure representative sampling, three plants reflecting the overall growth condition were randomly selected from each plot and placed in sealed plastic bags for transportation to the laboratory. Once separated, the stems and leaves were washed with running water and placed in an oven at 105°C for 1 hour, followed by drying at 80° C for more than 48 hours until a stable weight was reached. The stems and leaves were then weighed using a high-precision balance (accuracy 0.001 g), and the total weight of the samples was calculated. AGB was determined based on the population density and the total weight of the samples. Collection of PH data: Four representative rapeseed plants were selected from each plot, and the distance from the base to the tip of the leaf was measured using a ruler. The average of the four measurements was calculated as the PH for the rapeseed in the plot.

2.2.2 Hyperspectral and RGB data collection

A six-rotor UAV (Matrice 600 Pro, DJI, China) equipped with a Resonon PIKA CX imaging spectrometer (Resonon, USA) was used to capture hyperspectral data on sampling days. The hyperspectral sensor covers a spectral range of 400–1000 nm with 2.2 nm spectral resolution across 150 channels. To ensure data consistency, all flights were conducted from the same takeoff point between 12:00 and 13:30 under clear, windless conditions. Each flight followed a consistent route for all growth stages, flying at an altitude of 50 m (with a 6 m transect width) and a speed of 2 m/s. An 80% overlap was maintained in both forward and side directions. Radiometric calibration was performed before each flight using black and white reference panels. Hyperspectral images were acquired at four key growth stages: bare-soil (October 20), seedling, bolting, and early blossoming, with a spatial resolution of 2.5 cm.

RGB images were also captured on the same days and at the same altitude using a quad-rotor UAV (Dajiang Yu2, DJI, China). These images, along with digital surface models (DSM), were processed using Pix4D software. Ground control points (GCPs) were used to align the RGB photos for spatial consistency across growth stages. The images were processed into high-density point clouds, grids, and textures using structure-from-motion (SfM) algorithms, and mosaicked into digital orthophotos for each growth stage.

2.2.3 Hyperspectral data processing

The UAV hyperspectral images were preprocessed in three steps: (1) Correction and mosaicking: The data were corrected and mosaicked using Spectronon software (USA) for both radiometric and geographic corrections. (2) Image stitching: ENVI software (Harris Exelis, USA) was used to stitch the images within the specified navigation area, incorporating position data from GCPs to minimize correction error. This resulted in a new hyperspectral image. (3) Extracting rapeseed canopy reflectivity: Rapeseed canopy reflectance curves were extracted from the images. Using ArcGIS software, different maximum area vectors were delineated, and vector data were assigned numbers based on samples. The average spectral reflectance of each region of interest was then extracted using the interactive data language (IDL) in ENVI software. These values were considered the spectral reflectance of the rapeseed canopy in different plots. The images and the average values for each plot were used as the rapeseed canopy spectrum, as illustrated in Figures 2A–F.

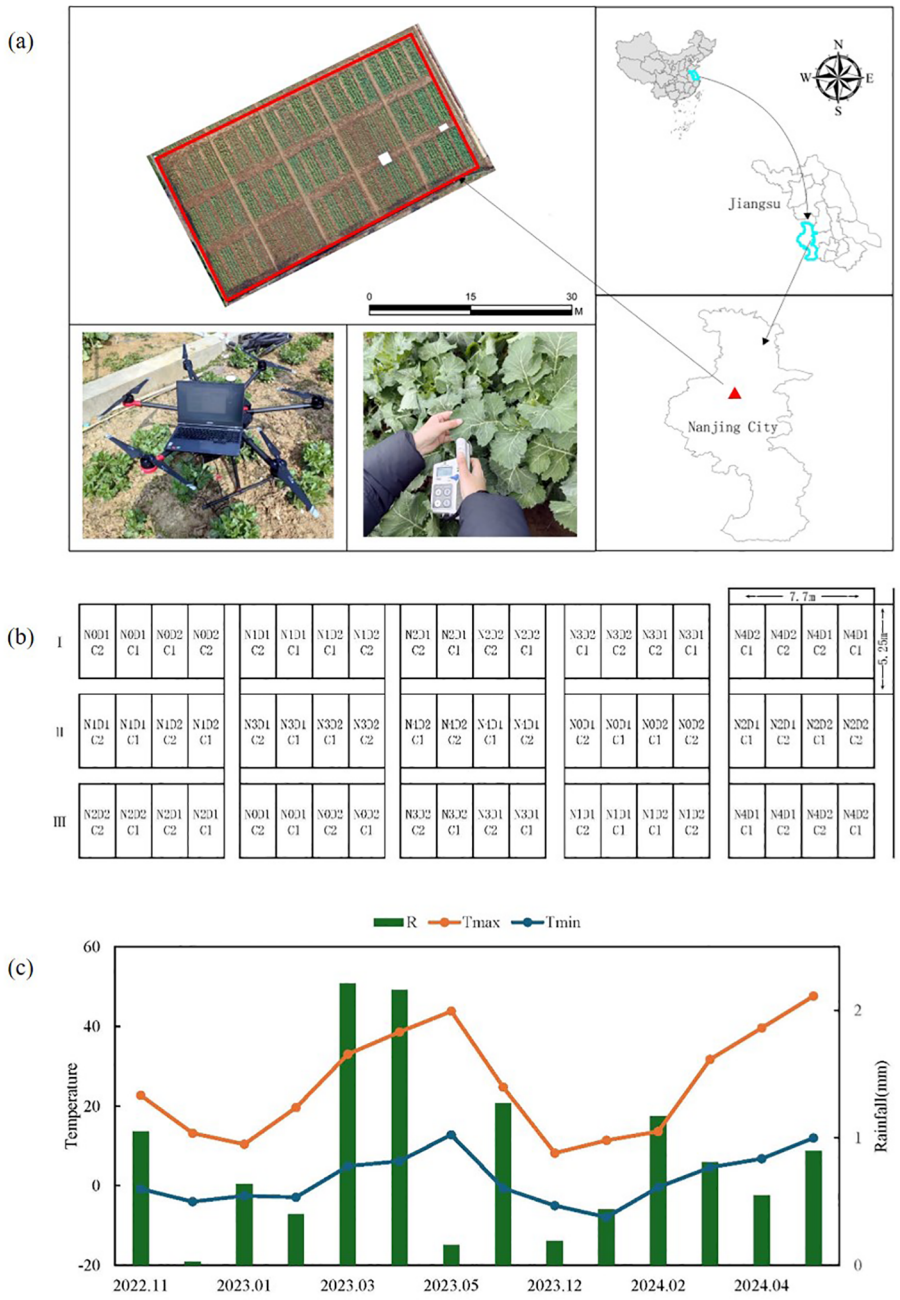


FIGURE 1
Geographical location of the study area and planting region for rapeseed. **(A)** Sampling area for Measurement data and spectral data collection, **(B)** Field distribution of processing condition. **(C)** Temperature of the crop growing season.

2.3 Optical and phenotypic metrics selection

2.3.1 VI metrics extraction

Canopy spectral information, obtained through optical sensors, is a crucial parameter for monitoring crop growth. The VI is closely linked to the physiological and biochemical characteristics of crops, making it an essential tool for assessing crop development. This index captures the interaction between spectral bands and enhances the response to specific crop properties. Based on previous research, 15 spectral vegetation indices were selected to estimate the AGB of rapeseed. The band calculation tool in ENVI 5.3 software was used to compute these VIs, as detailed in Table 1.

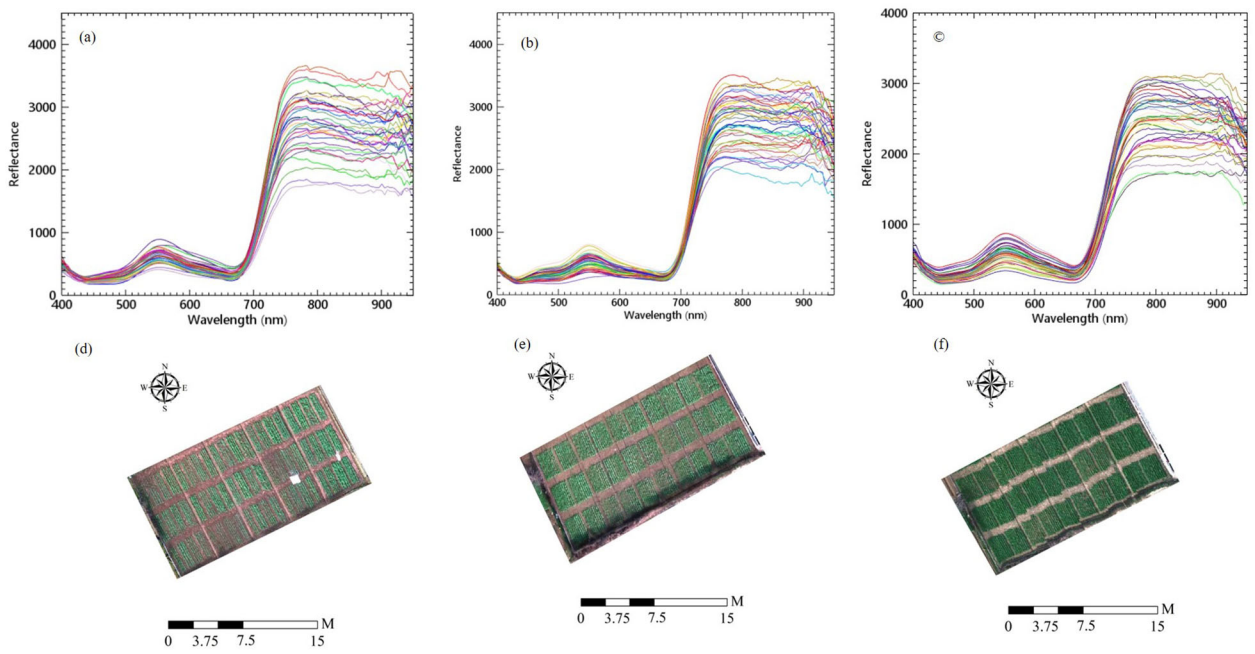


FIGURE 2
The hyperspectral images and canopy spectral reflectance of rapeseed at **(A, D)** seedling stage, **(B, E)** bolting stage and **(C, F)** early blossoming stage.

2.3.2 TF metrics extraction

The Gray Level Co-occurrence Matrix (GLCM) is one of the most widely used methods for texture extraction, originally proposed by Haralick (1973). GLCMs became popular due to their ability to maintain rotational invariance, capture multi-scale features, and allow for low-complexity calculations (Haralick et al., 1973). In this study, three texture metrics—Data Range (DR), Variation (VAR), and Entropy (ENT)—were extracted from UAV RGB bands. The selected window size effectively captures variations in spatial information among the rapeseed plants within the experimental plot. A window that is too small can increase computational complexity and the volume of calculations, while a window that is too large may result in the loss of detailed texture information (Bai et al., 2021). To address this, an averaging technique that combines the functionality of different window sizes was employed. Through trial and error, texture features were computed using the average values of two window sizes (3 pixels × 3 pixels and 5 pixels × 5 pixels) and four directional orientations (0°, 45°, 90°, and 135°) rotated clockwise along the x-axis. These features have been shown to be effective in quantifying changes in crop canopy structure and estimating AGB (Yue et al., 2018a).

2.3.3 SF metrics extraction

In this study, UAV RGB images of rapeseed were captured to create a base map for the DEM before sowing. Canopy point cloud images were then acquired to construct the DSM at various growth stages of rapeseed. The height model for each growth period was derived by subtracting the DEM from the DSM. Using the statistical toolbox in ArcGIS and the Kriging interpolation algorithm, the average plant height (PH) was extracted from the region of interest for each image (Xu et al., 2022). Plant roughness (PR), a metric that characterizes the irregularities of the canopy surface, was measured using 3D point clouds from UAV RGB images. PR has been shown to have a significant correlation with crop AGB (Herrero-Huerta et al., 2020). The fraction of plant cover (PC) was determined through image classification, where ground objects in the RGB images were categorized as either crops or soil (Maimaitijiang et al., 2019). PC for each image was calculated by dividing the number of cropped pixels by the total number of pixels in the image. Building on previous research, the volume method was applied to estimate crop biomass within a defined spatial range. The plant volume metric (PVM) of rapeseed was calculated as the product of PC and PH, along with the canopy elevation fluctuation rate (CEFR) to describe the relative shape of the canopy, as commonly used in forestry studies (Han et al., 2019). The canopy structure metrics extracted from UAV RGB images—PH, PR, PC, PVM, and CEFR—are defined in Table 2.

2.4 Model construction method

2.4.1 Model construction

The input layer of the DNN model used in this study consists of four hidden layers with 256, 128, 64, and 32 neurons, respectively (Hu et al., 2024). A ReLU activation function was applied after each hidden layer. To address overfitting, a dropout layer with a 0.2 ratio was added after the first hidden layer. The network was trained using the Adaptive Moment Estimation (ADAM) optimizer, with a maximum of 600 training iterations and a batch size of 256. The initial learning rate was set at 0.001, decreasing by 10% every 100 rounds. For the Random Forest (RF) model, bootstrap sampling was used to create a training dataset, and random decision trees were generated based on the integrated classifier (Niu et al., 2019). The RF model was configured

TABLE 1 The VI metrics extracted from hyperspectral images.

Vegetation indices	Equation	Reference
NDVI (normalized-difference vegetation index)	(R800 − R680)/(R800 + R680)	(Rouse et al., 1974)
RVI (ratio vegetation index)	R810/R660	(Rouse et al., 1974)
EXG (Excess green vegetation index)	2G-R-B	(Bendig et al., 2013)
EXB (Excess blue vegetation index)	1.4B-G	(Bendig et al., 2013)
NGBVI (Red green blue vegetation index)	(G^2-BR)/(G^2+BR)	(Bendig et al., 2013)
NGBDI (Normalized green blue difference index)	(G-B)/(G+B)	(Bendig et al., 2013)
EVI(enhanced vegetation index)	2.5 × (R800 − R670)/(R800 + 6 × R670 − 7.5 × R450 + 1)	(Tao et al., 2020)
SPVI (spectral-polygon vegetation index)	0.4 × [3.7 × (R800 − R670) − 1.2 × |R550 − R670|]	(Tao et al., 2020)
MCARI (modified chlorophyll-absorption ratio index)	((R700 − R670) − 0.2 × (R700 − R550))×(R700/R670)	(Tao et al., 2020)
RNDVI (renormalized-difference vegetation index)	(R800 − R670)/(R800 + R670)1/2	(Tao et al., 2020)
$CI_{red\ edge}$ (Red edge chlorophyll index)	R810/R690-1	(Gitelson et al., 2003)
VARI (Visible atmospherically resistance index)	(R555 − R680)/(R555 + R680 − R480)	(Gitelson et al., 2002)
SAVI (Soil-adjusted vegetation index)	(1 + 0.5) × (R800 − R670)/(R800 + R670 + 0.5)	(Gnyp et al., 2014)
GNDVI (Green normalized-difference vegetation index)	(R750 − R550)/(R750 + R550)	(Zheng et al., 2019)
SIPI (Structure-insensitive pigment index)	(R800 − R450)/(R800 + R680)	(Li B. et al., 2020)

with 80 decision trees (n tree = 80) and a maximum number of variables considered at each split (m try = 4). The final prediction was determined through a majority voting process among the decision trees. Support Vector Regression (SVR) was applied for linear and nonlinear regression tasks (Liu et al., 2023). The training dataset was binary-classified using a kernel function to minimize the distance of all samples from the hyperplane. The sample data were then fitted to generate predictions.

2.4.2 Model evaluation

A total of 60 datasets were collected for each period during the 2022-2023 season. Repeats 1 and 2 were selected as the calibration dataset, while plots from Repeat 3 were used as the validation dataset. The statistical results were presented in Table 3. To construct an AGB estimation model for rapeseed across various growth stages, a ten-fold cross-validation approach was employed. Pearson correlation analysis was performed to examine the relationship between features and AGB. The model's performance and stability were assessed using the coefficient of determination (R^2), prediction root mean square error (RMSE), and relative root mean square error (rRMSE). The study workflow was shown in Figure 3.

$$R^2 = 1 - \frac{\sum_i (y - y')^2}{\sum_i (y - \bar{y})^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y - y')^2}{n}} \quad (2)$$

$$rRMSE = \frac{RMSE}{\bar{y}} \quad (3)$$

Where y is the observed value (manual measurement), and y' is the predicted value (model extracted value), $\bar{y}$ is the average value, and n is the sample size.

TABLE 2 The SF metrics extracted from UAV-RGB images.

SF metrics	Equation	Reference
PH (Plant Height)	DSM-DEM	(Xu et al., 2022)
PR (Plant Roughness)	IQR^{med}	(Herrero-Huerta et al., 2020)
PC (Plant Cover)	Plant Pixel/total pixels	(Maimaitijiang et al., 2019)
PVM (Plant Volume Metric)	$\sum_i^N S_* PH_i$	(Han et al., 2019)
CEFR (Canopy Elevation Fluctuation Rate)	(PH_{mean}-$PH_{10\%min}$)/($PH_{10\%max}$-$PH_{10\%min}$)	(Han et al., 2019)

3 Results

3.1 Correlation of metrics and AGB

3.1.1 Statistical analysis of AGB measurements

For the AGB samples, the average value in the calibration dataset was 1785.6 kg/ha, with an overall coefficient of variation of 38.85%, while the validation dataset had a higher coefficient of variation at 42.38% (Table 3). The minimum AGB value observed was 147.5 kg/ha, and the maximum was 5573.39 kg/ha. For PH, the average was 43.74 cm in the calibration dataset and 42.76 cm in the validation dataset. The overall coefficient of variation for PH was 41.75%, while the validation dataset exhibited a larger variation of 53.18%. These results indicate that the validation dataset generally

TABLE 3 Descriptive statistics for AGB (kg/ha) and PH (cm) of calibration and validation datasets.

Dataset	Crop parameters	Min	Average	Max	Standard deviation	Coefficient of variation (%)
Calibration	AGB	147.50	1785.60	5573.39	688.56	38.85
	PH	8.90	43.74	110.50	18.26	41.75
Validation	AGB	223.40	1945.23	6061.60	824.47	42.38
	PH	10.10	42.76	105.20	22.74	53.18

showed larger coefficient of variation values compared to the calibration dataset.

3.1.2 Correlation of VI metrics and AGB

The correlation between VI metrics and AGB across different growth stages was illustrated in Figure 4. The AGB and VI values for the three different growth stages of rapeseed showed strong significance ($p< 0.01$). The strongest correlation during the seedling stage was observed with the RVI, which had a correlation coefficient of $r = 0.82$ ($p< 0.01$). Significant correlations were also noted between AGB and SAVI ($r = 0.75$, $p<0.01$) as well as NDVI ($r = 0.72$, $p< 0.01$). The results suggested a linear relationship between VI and AGB at all growth stages, although the strength of the correlation decreased as the crop developed. It was important to note that VI tends to saturate when AGB was high, meaning the accuracy of AGB estimation using a single VI may require validation through a more comprehensive estimation model.

3.1.3 Correlation of TF metrics and AGB

In this study, the correlation between nine TF metrics and rapeseed AGB across different growth stages was evaluated, as shown in Figure 5. The strongest correlation was observed between the GVAR and AGB across all three growth stages, with the average correlation exceeding 0.5 ($p< 0.01$). The correlations between GDR and BENT were also close to 0.5 ($p< 0.05$). The results indicate that DR, VAR and ENT metrics exhibit significant variability in relation to AGB, suggesting they fluctuate more throughout the growth stages in response to AGB.

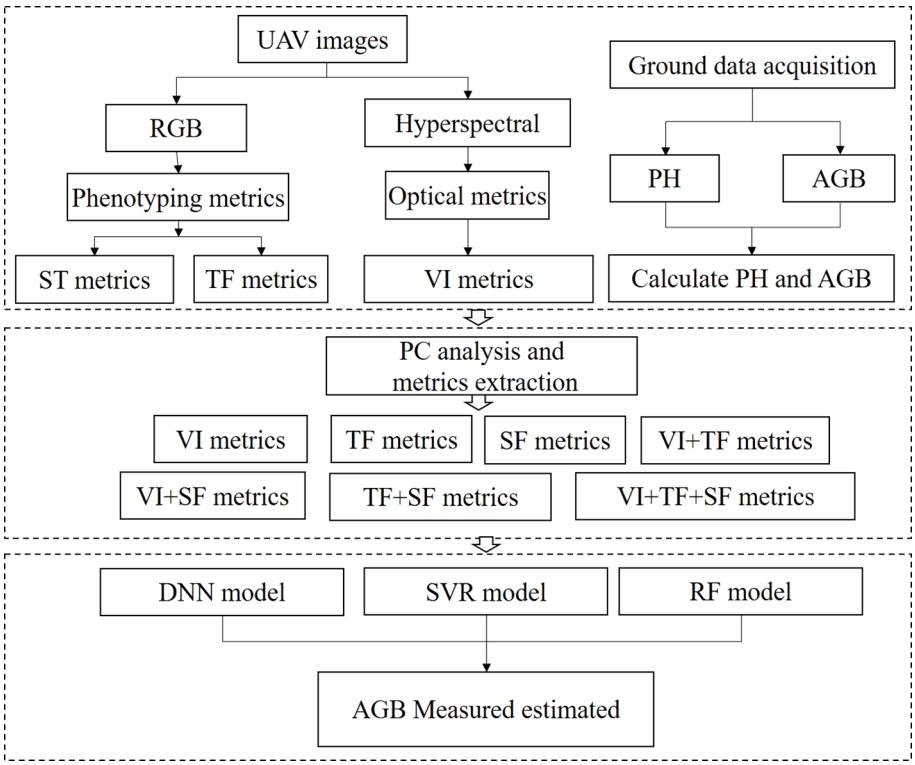


FIGURE 3
Research workflow.

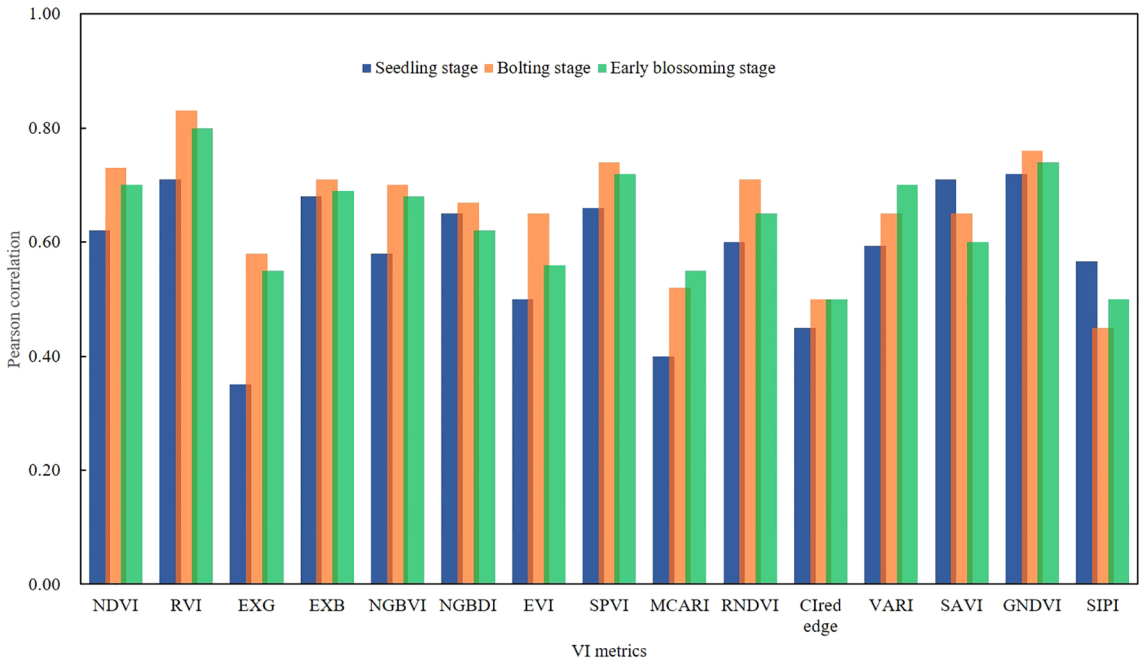


FIGURE 4
Pearson correlation analysis between VI metrics and AGB for three growth stages.

3.1.4 Correlation of SF metrics and AGB

The correlation between SF metrics and AGB at the seedling, bolting, and Early blossoming stages of rapeseed was shown in Figure 6. SF metrics, such as PH, PR, PC, and CEFR, showed significant correlations with AGB across all three growth stages (pp< 0.01). The correlation between these metrics and AGB increased progressively through the growth stages. Notably, the PVM showed a rise in correlation with AGB at first, followed by a decrease in the later growth stages. PH displayed the strongest correlation across all stages, with r values of 0.56 (p< 0.05), 0.67 (p< 0.01), and 0.75 (p< 0.01) at the

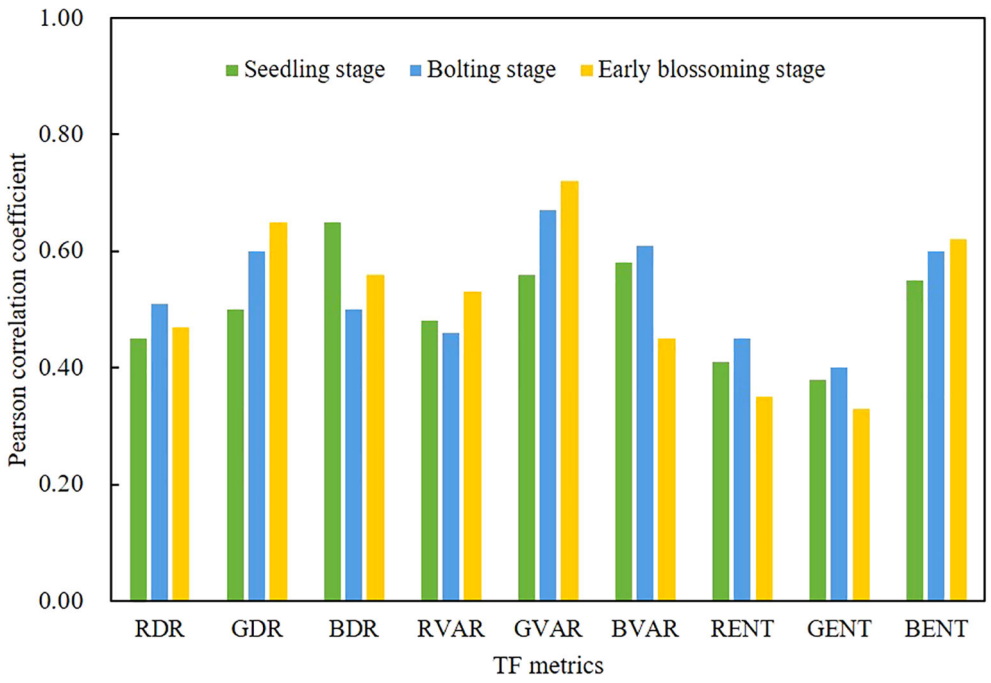


FIGURE 5
Pearson correlation analysis between TF metrics and AGB for three growth stages.

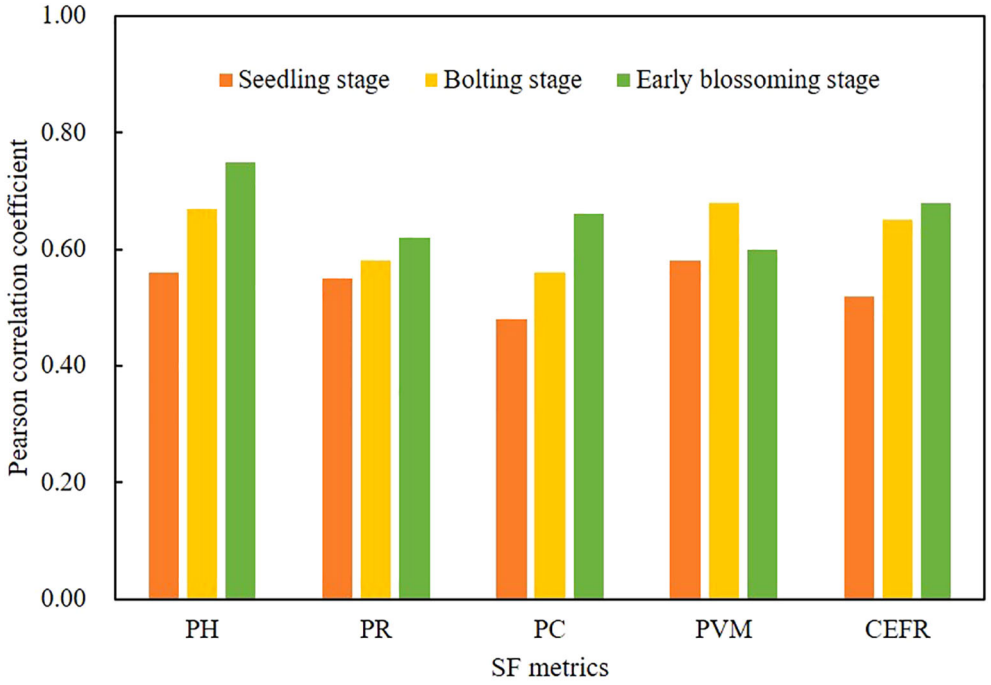


FIGURE 6
Pearson correlation analysis between SF metrics and AGB for three growth stages.

seedling, bolting, and Early blossoming stages, respectively. These results demonstrate that SF metrics have a linear relationship with AGB at each growth stage.

3.2 Construction of AGB estimation models

Based on the results of the correlation analysis, we selected the most significant correlation metrics ($r > 0.6$) as input variables for the AGB estimation models (Table 4). Utilizing both the calibration and validation datasets, we developed rapeseed AGB estimation models through three machine learning algorithms: ANN, SVR, and RF. The regression results, including R^2, RMSE, and rRMSE, were summarized in Table 5 and illustrated in Figure 6. The findings indicate that, among the estimation models constructed using individual features and the three algorithms, the ANN model combined with VI yielded the best performance. Specifically, the ANN model that employed the combination of VI, TF, and SF achieved the highest R^2 values of 0.878 and 0.864 for the training and test datasets, respectively. This model also resulted in the lowest RMSE and rRMSE values of 447.02 kg/ha and 0.171, respectively. In contrast, the SVR model using TF exhibited the lowest R^2 along with the highest RMSE and rRMSE values. The performance ranking of the three algorithms in constructing AGB estimation models was as follows: ANN > RF > SVR.

Figure 7 presents the R^2 values for models utilizing seven different feature combinations. In four combinations—VI, VI+TF, VI+SF, and VI+TF+SF—the ANN model consistently outperformed both SVR and RF models in terms of R^2. However, for the TF, SF, and TF+SF combinations, the RF models achieved the highest R^2 and the lowest RMSE and rRMSE across both training and test datasets. In the VI+TF model, the SVR model recorded the lowest R^2 and the highest RMSE and rRMSE, while the ANN model ranked second behind RF. In the VI+TF+SF model, the RF achieved the highest R^2 on the training set, whereas the SVR model displayed the lowest R^2. However, on the test set, the ANN model produced the highest R^2. These results confirm that the ANN model, when combined with VI, TF, and SF features, provides the best performance for both training and validation datasets, highlighting the superior capability of the ANN model in estimating rapeseed AGB.

3.3 Optimal estimation model of rapeseed at different growth stages

AGB estimation models for rapeseed were constructed for three key growth stages: seedling, bolting, and early blossoming, using three machine learning algorithms: ANN, RF, and SVR. The optimal feature combinations were selected, and the results were presented in Table 6 and Figure 8. Across all growth stages, the ANN model consistently demonstrated the highest accuracy, followed by the RF model, while the SVR model exhibited the lowest accuracy. The high consistency between the training and validation set results further confirms the superior performance of the ANN algorithm. In terms of growth stage comparisons, the ANN model achieved an AGB estimation accuracy of 0.783 during the seedling stage. As the rapeseed developed, the estimation accuracy improved, reaching 0.896 during the bolting stage. However, a slight decrease in accuracy was observed during the

TABLE 4 Input metrics were selected for algorithms.

Algorithm	Feature	Metrics
DNN, SVR, RF	VI	NDVI, RVI, EXB, NGBDI, RGBVI, EVI, SPVI, RNDVI, VARI, SAVI, GNDVI
	TF	GDR, GVAR, BDR, BVAR, BENT
	SF	PH, PR, PC, PVM, CEFR

early blossoming stage, with a value of 0.878. Similar trends were observed in the RF and SVR models, where estimation accuracy peaked during the bolting stage, outperforming both the seedling and Early blossoming stages. As a result, the AGB maps produced of different periods by optimal estimation model (Figure 10).

3.4 Evaluation of variable importance

To assess the contribution of different input metrics to the estimation models, we applied the RF importance evaluation method. Figure 9 illustrates the variable importance scores of three rapeseed growth stages. During the seedling stage, the RVI metric exhibited the highest importance, with vegetation indices contributing approximately 50% of the overall importance. However, as the crop progressed to the bolting stage, structural metrics gained prominence, with PH becoming a key factor in AGB estimation. The variable importance in the early blossoming stage closely mirrored the results of the bolting stage, highlighting the continued relevance of structural metrics at later stages of development. Overall, these findings were consistent with the performance of the AGB estimation models, reflecting the shift in the relative importance of vegetation and structural parameters as rapeseed matures.

4 Discussion

4.1 The correlation of metrics and rapeseed AGB

This study assessed the effectiveness of combining spectral, textural, and structural features derived from UAV-based

TABLE 5 Rapeseed AGB estimates based on different features combination with algorithms.

Features	Algorithm	Calibration			Validation		
		R^2	RMSE (kg/ha)	rRMSE	R^2	RMSE (kg/ha)	rRMSE
VI	DNN	**0.765**	**597.27**	**0.235**	**0.705**	**810.51**	**0.316**
	SVR	0.681	683.65	0.272	0.620	951.54	0.351
	RF	0.726	634.51	0.266	0. 661	927.16	0.348
TF	DNN	0.554	823.14	0.347	0.510	1135.84	0.434
	SVR	0.615	767.83	0.312	0.583	1015.21	0.395
	RF	**0.629**	**755.92**	**0.295**	**0.561**	**1054.33**	**0.406**
SF	DNN	0.647	722.16	0.272	0.614	968.78	0.364
	SVR	0.605	774.41	0.311	0.564	1054.65	0.401
	RF	**0.720**	**686.02**	**0.264**	**0.676**	**889.82**	**0.332**
VI+TF	DNN	**0.798**	**550.78**	**0.215**	**0.733**	**803.14**	**0.318**
	SVR	0.732	650.24	0.268	0.703	835.24	0.314
	RF	0.765	540.15	0.235	0.695	848.17	0.326
VI+SF	DNN	**0.823**	**524.95**	**0.201**	**0.785**	**687.82**	**0.264**
	SVR	0.741	645.18	0.244	0.712	788.15	0.292
	RF	0.804	556.47	0.206	0.768	754.52	0.277
TF+SF	DNN	0.755	630.13	0.23	0.712	788.15	0.295
	SVR	0.651	712.54	0.295	0.581	1020.45	0.397
	RF	**0.763**	**638.43**	**0.244**	**0.745**	**762.53**	**0.286**
VI+TF+SF	DNN	**0.878**	**447.02**	**0.171**	**0.864**	**583.85**	**0.224**
	SVR	0.781	563.24	0.227	0.733	775.16	0.295
	RF	0.812	530.15	0.205	0.761	761.58	0.283

The best model under each feature combination is shown in bold.

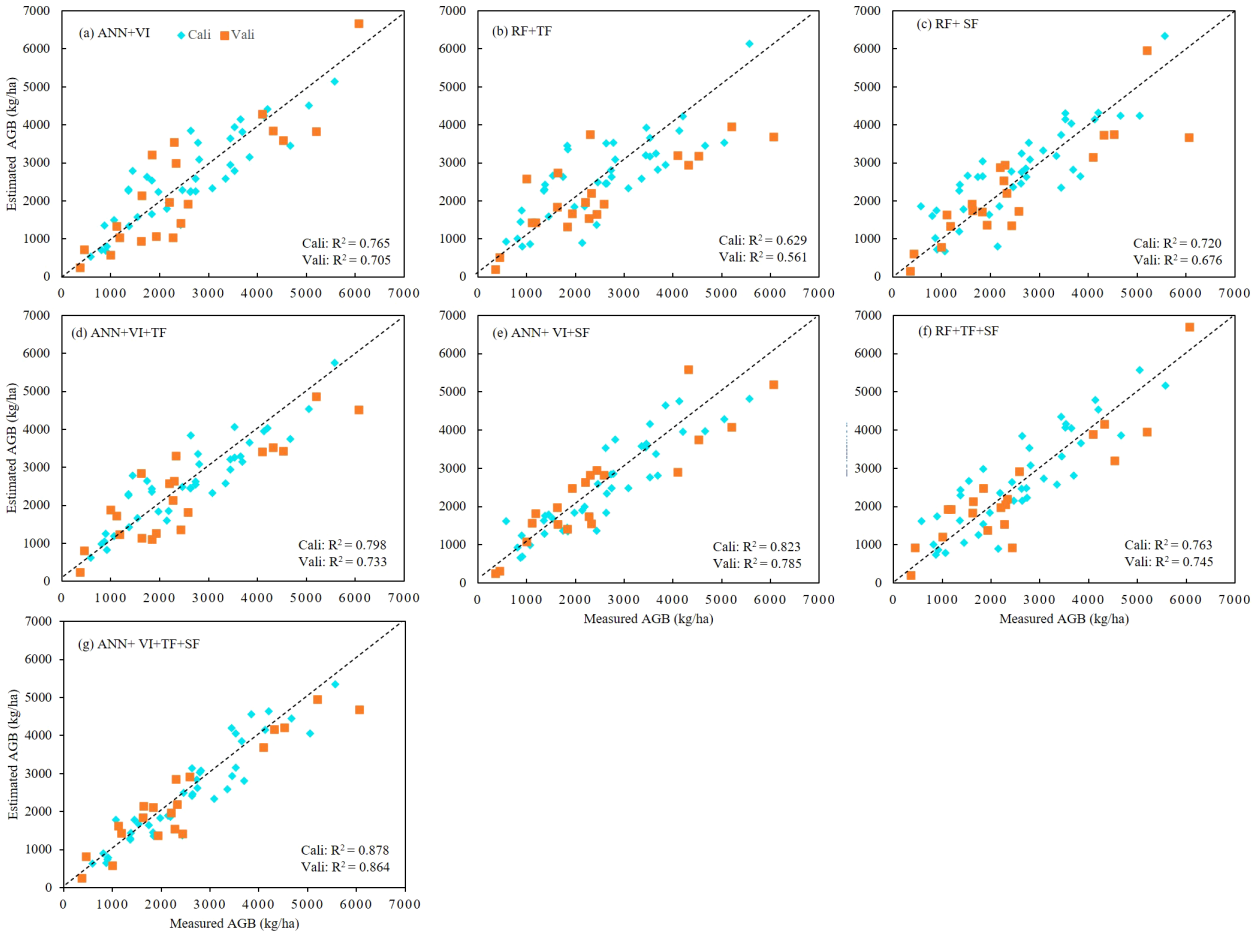


FIGURE 7
Scatter plots between measured and estimated AGB for calibration and validation datasets of different combination model. **(A)** Combine DNN and VI, **(B)** combine RF and TF, **(C)** combine RF and SF, **(D)** combine DNN and VI+TF, **(E)** combine DNN and VI+SF, **(F)** combine RF and TF+SF, **(G)** combine DNN and VI+TF+SF.

TABLE 6 Estimated AGB for calibration and validation datasets of three growth stages.

Growth stages	Algorithm	Calibration			Validation		
		R^2	RMSE (kg/ha)	rRMSE	R^2	RMSE (kg/ha)	rRMSE
Seedling stage	DNN	**0.783**	**101.67**	**0.195**	**0.762**	**115.64**	**0.213**
	SVR	0.655	129.15	0.251	0.584	155.32	0.284
	RF	0.751	106.93	0.227	0.702	128.53	0.236
Bolting stage	DNN	**0.896**	**193.18**	**0.184**	**0.866**	**169.51**	**0.175**
	SVR	0.774	264.65	0.242	0.722	244.36	0.254
	RF	0.830	229.6	0.214	0.797	209.18	0.216
Early blossoming stage	DNN	**0.878**	**447.01**	**0.177**	**0.831**	**598.77**	**0.238**
	SVR	0.745	609.17	0.236	0.695	804.08	0.319
	RF	0.824	513.32	0.194	0.787	660.14	0.262

The best model under each stage is shown in bold.

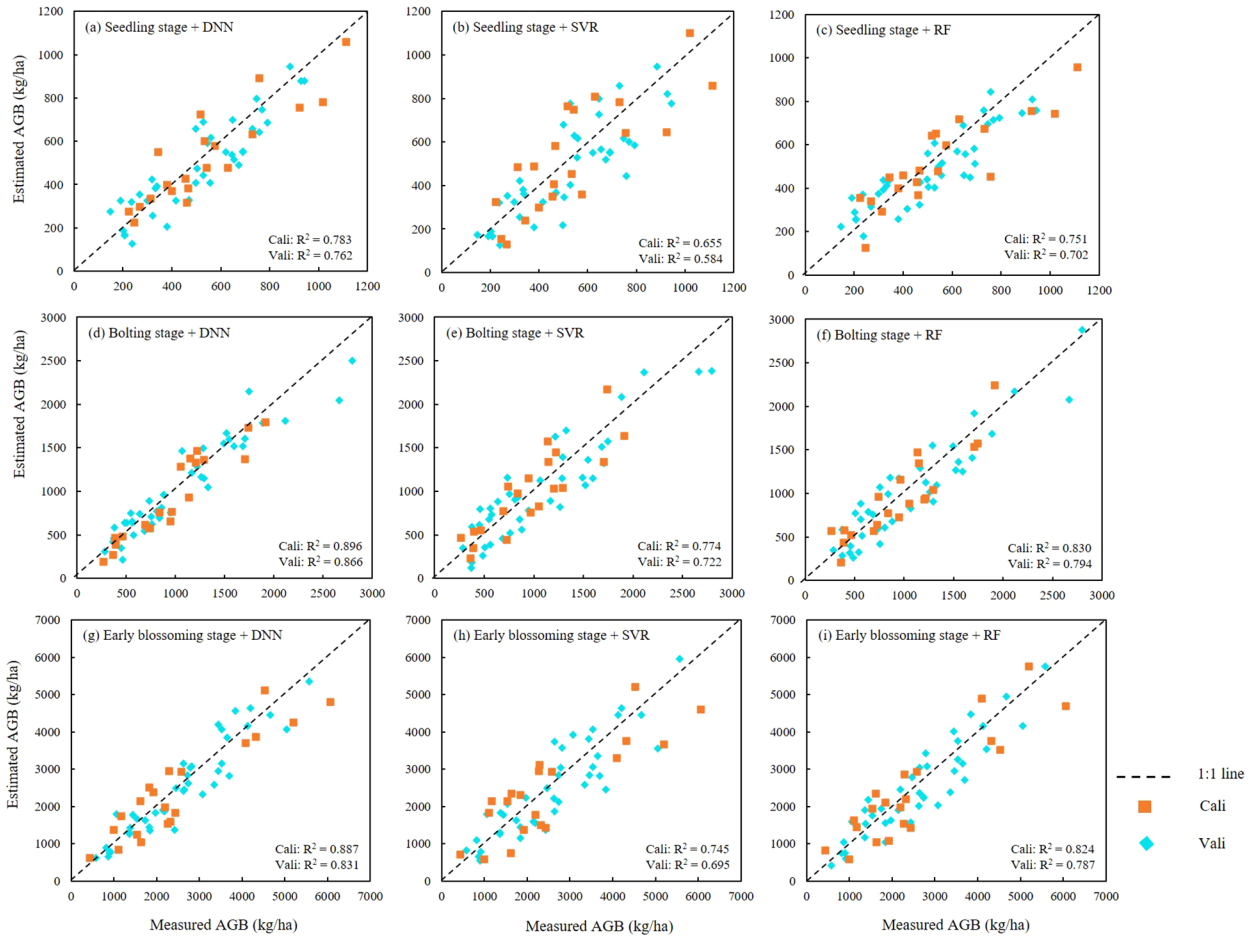


FIGURE 8
Scatter plots between measured and estimated AGB obtained by DNN, SVR and RF models. **(A–C)** were seedling stage with three algorithms, **(D–F)** were bolting stage with three algorithms, **(G–I)** were Early blossoming stage with three algorithms.

hyperspectral and RGB imagery to enhance the accuracy of AGB estimation in winter rapeseed. Previous research has demonstrated the high accuracy of UAV-based methods for estimating plant biomass (Liu et al., 2022, 2023; Niu et al., 2019). Consistent with these studies, our analysis found significant correlations between rapeseed AGB and the spectral, textural, and structural parameters extracted from UAV imagery. However, we observed that the stability of these correlations varied across different growth stages. While individual features, such as vegetation indices, texture metrics, and structural characteristics, were strongly correlated with AGB, their predictive power fluctuated as the crop developed. Notably, a combined approach leveraging the strengths of spectral, textural, and structural features offers considerable potential for achieving more accurate and consistent AGB estimates throughout rapeseed's growth cycle.

The correlation analysis between VI, TF, and SF with rapeseed AGB shows distinct trends across different growth stages. VI tend to saturate as spectral parameters stabilize, resulting in peak correlations during the bolting stage, following the seedling stage. These correlations gradually weaken as the crop enters the early blossoming stage (Liu et al., 2023). In contrast, TF and SF demonstrate increasing complexity with crop growth, reflecting the canopy's development. This complexity aligns with the rise in aboveground biomass, suggesting these features may provide complementary insights into rapeseed AGB estimation.

4.2 Estimation performance of various metrics

Previous studies have indicated that texture metrics can outperform VI in predicting above-ground biomass (AGB), though much of this research has focused on forests, with relatively limited applications in crop biomass estimation (Basyuni et al., 2023; Nichol and Sarker, 2011; Xu et al., 2024). In our study, we found that among individual feature types, SF produced more accurate AGB estimates than TF. Moreover, the integration of VI+TF+SF led to significant improvements in AGB estimation for winter rapeseed by reducing RMSE compared to models that relied solely on VI, TF, or SF. Contrary to earlier findings, our results did not show that texture metrics alone outperformed VI in estimating AGB. Instead, our study highlights that the combination of VI and SF yielded better AGB estimates than the combination of VI and TF (Schumacher et al., 2016; Xu et al., 2022, 2024).

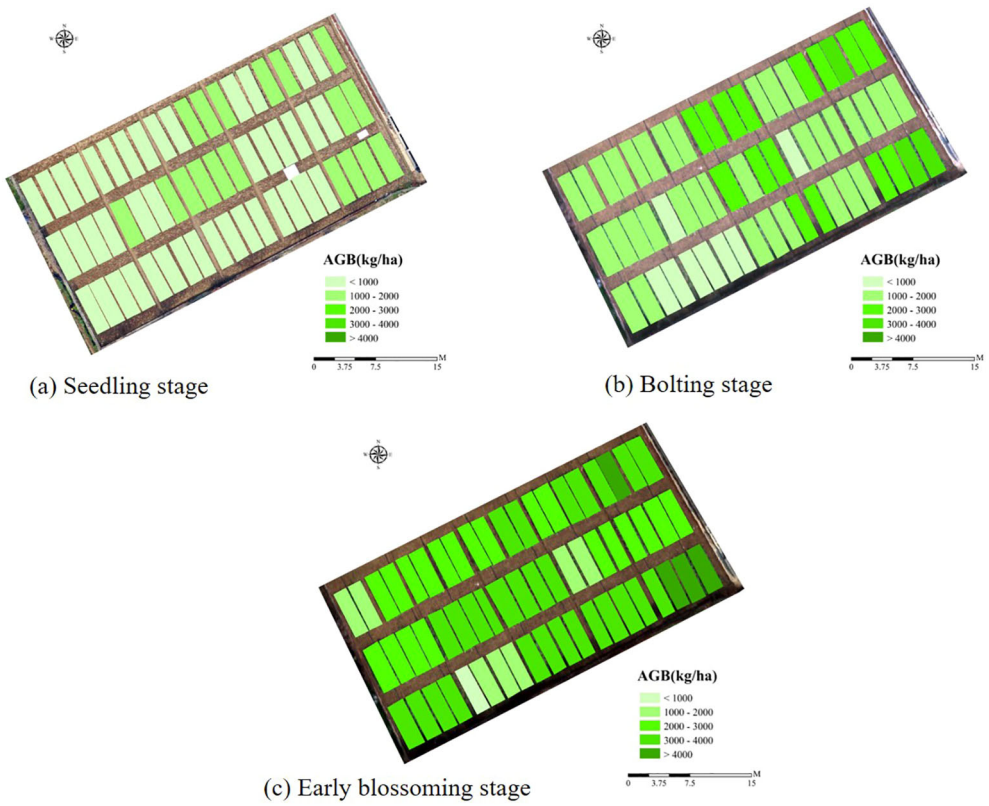


FIGURE 9
Mapping rapeseed AGB using three metrics and DNN model. **(A)** Seedling stage, **(B)** Bolting stage, **(C)** Early blossoming stage.

Comparing the estimation capabilities of different features at various growth stages of rapeseed, we found that structural parameters yield higher accuracy in biomass estimation as the crop matures. This aligns with the importance evaluation, which shows that SF like PH, PR, and PC become increasingly significant as the crop develops. Integrating multiple feature types—VI, TF, and SF—offers a more comprehensive approach to estimating AGB, as each capture unique and complementary information about crop growth. The highest estimation accuracy was observed during the bolting stage, providing crucial insights for guiding fertilization decisions during this key phase of rapeseed development.

4.3 Advantages of model estimation

Our findings demonstrate the effectiveness of machine learning models—specifically RF, SVR, and DNN—in estimating rapeseed AGB. Across the three growth stages, the DNN model consistently achieved the highest R^2 (0.896) and the lowest RMSE (193.18kg/ha) and rRMSE (0.184) (Table 6). Although the overall performance of the RF model was slightly lower than that of the DNN model, it demonstrated superior accuracy when TF and SF were combined as input variables. In this case, the RF model outperformed the DNN model in terms of R^2 and rRMSE on the test set, indicating that RF may be more adept at processing texture and structural data. The strong performance of the DNN model is likely due to its deep iterative layers, which allow it to capture complex patterns in the data, highlighting its potential for biomass estimation. However, the RF model's ability to handle TF and SF features effectively suggests that it is particularly well-suited for integrating these types of data. The evaluation of feature importance also revealed that the relative contribution of different input parameters significantly impacts model performance, further explaining the variations in accuracy among the algorithms.

The estimation results indicate that the inclusion of structural parameters significantly improved the performance of the estimation models, suggesting that the enhancement in estimation accuracy is related to the addition of key estimation factors (such as PH, PR, and PC). Among the three algorithms used to construct the estimation models, the DNN model demonstrated a clear advantage across all three growth stages, while the RF model performed better during the bolting and early flowering stages. In contrast, the SVR model showed weaker estimation performance in all stages. This suggests that the DNN algorithm exhibits good practical applicability for estimating rapeseed AGB at different growth stages.

4.4 Research outlook

The process from data collection to model development requires meticulous attention to detail and thorough analysis. Improving the quality of data obtained from UAV imagery,

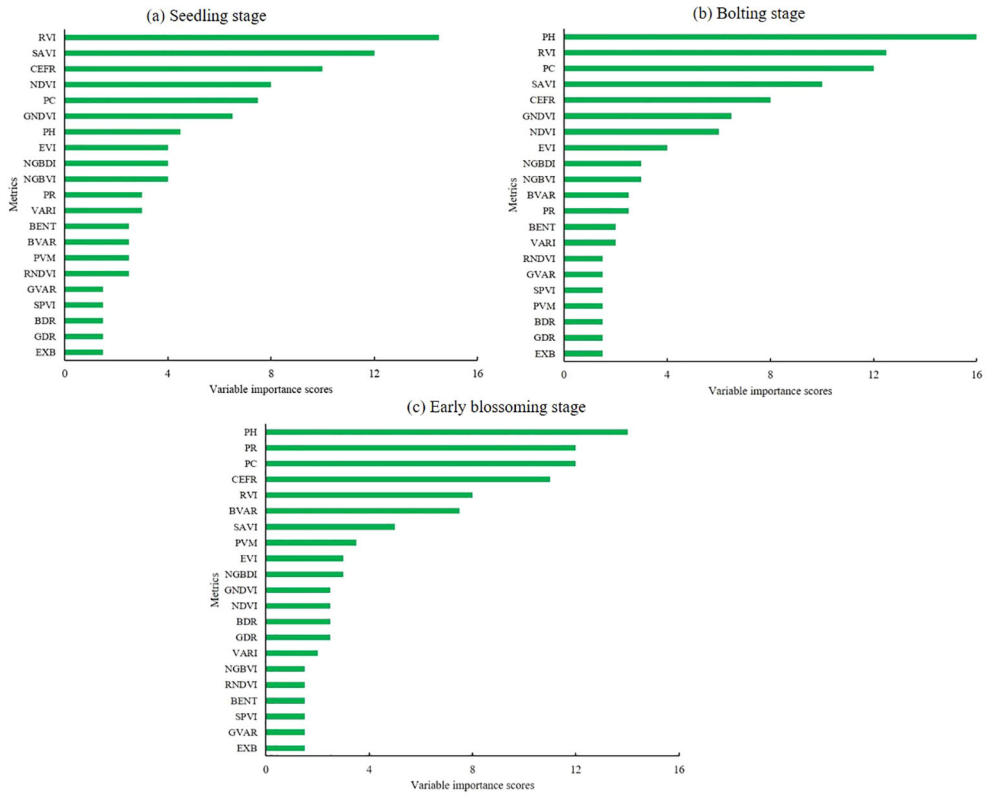


FIGURE 10
The variable importance scores of three growth stages. **(A)** Seedling stage, **(B)** Bolting stage, **(C)** Early blossoming stage.

ground observations, and modeling techniques is crucial, as these datasets can be prone to errors. Standardizing these procedures is essential for ensuring consistency and accuracy.

While this study presents an effective approach by integrating three feature types for estimating rapeseed AGB, several challenges remain. Data accuracy in hyperspectral acquisition and preprocessing must be carefully managed, and the mixed pixel problem due to resolution constraints may impact estimation performance. Additionally, the comparison of biomass under different nitrogen treatments is a valuable area of research that warrants further investigation. Future studies could extend the scope by exploring rapeseed biomass estimation over multiple years to evaluate the applicability and transferability of the developed models, thus enhancing their generalizability. Furthermore, delving deeper into the biophysical properties of plants and identifying potential error sources will be crucial for further refining estimation accuracy.

5 Conclusions

This study explored the potential of UAV hyperspectral and RGB imagery for estimating crop biomass by developing a multi-feature estimation model that incorporates VI, TF, and SF. The performance of these features in estimating rapeseed AGB across different growth stages was thoroughly evaluated. The RF importance evaluation method was used to assess the contribution of different input parameters to the estimation model. The results indicated that both DNN and RF outperformed SVR when using individual features for AGB estimation. Additionally, the DNN model surpassed the RF model in accuracy when feature combinations (VI, TF, and SF) were applied, achieving the best estimation performance across all growth stages. Furthermore, the DNN model (R^2 = 0.878, RMSE = 447.02 kg/ha) with the combined features outperformed both the RF (R^2 = 0.812, RMSE = 530.15 kg/ha) and SVR (R^2 = 0.781, RMSE = 563.24 kg/ha) models. Based on the variable importance analysis, the RVI index emerged as the most significant, while PH was identified as a key phenotypic metric in AGB estimation. These findings demonstrate that integrating hyperspectral and RGB data with advanced artificial intelligence models, particularly DNN, provides an effective approach for estimating rapeseed AGB. The estimation model incorporating VI, TF, and SF showed higher accuracy in estimating rapeseed AGB compared to models using individual feature sets. Among the growth stages, the bolting stage yielded slightly higher estimation accuracy than the seedling and early blossoming stages. The combination of VI, TF, and SF metrics offers significant improvements in biomass estimation accuracy, highlighting the potential of UAV-based multi-feature modeling in precision agriculture.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

CS: Data curation, Formal analysis, Writing – original draft, Writing – review & editing. WZ: Investigation, Writing – review & editing. GZ: Methodology, Writing – review & editing. QW: Project administration, Writing – review & editing. WL: Funding acquisition, Supervision, Writing – review & editing. NR: Resources, Writing – review & editing. HC: Writing – review & editing. LZ: Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Program for National Key R&D Program (Technology and equipment for waterlogging mitigation in grain-producing areas of southern China (2023YFD2300300)), Jiangsu Agricultural Science and Technology Innovation Fund (CX(22)3108), the Key Research and Development Program (Modern Agriculture) of Jiangsu Province (BE2023302), the Key Research and Development Program of Hainan Province (ZDYF2024XDNY169), the Postdoctoral Later-stage Foundation Project of Shenzhen Polytechnic University (6023271029K), and was also carried out with the aid of the Guangdong Basic and Applied Basic Research Foundation under Grant (2021A1515110769), the Guangdong Basic and Applied Basic Research Foundation (2024A1515010036) and the funded by China Scholarship Council (CSC).

Acknowledgments

In addition, we are especially grateful to the reviewers and editors for appraising our manuscript and for offering instructive comments.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/J.TPLANTS.2013.09.008

Bai, X., Chen, Y., Chen, J., Cui, W., Tai, X., Zhang, Z., et al. (2021). Optimal window size selection for spectral information extraction of sampling points from UAV multispectral images for soil moisture content inversion. *Comput. Electron. Agric.* 190, 106456. doi: 10.1016/J.COMPAG.2021.106456

Banerjee, B. P., Spangenberg, G., and Kant, S. (2020). Fusion of spectral and structural information from aerial images for improved biomass estimation. *Remote Sens.* 12, 3164. doi: 10.3390/RS12193164

Basyuni, M., Wirasatriya, A., Iryanthony, S. B., Amelia, R., Slamet, B., Sulistiyono, N., et al. (2023). Aboveground biomass and carbon stock estimation using UAV photogrammetry in Indonesian mangroves and other competing land uses. *Ecol. Inf.* 77, 102227. doi: 10.1016/J.ECOINF.2023.102227

Bendig, J., Bolten, A., and Bareth, G. (2013). UAV-based imaging for multi-temporal, very high-resolution crop surface models to monitor crop growth variability. Photogramm Fernerkund. *Geoinf.* 6, 551–562. doi: 10.1127/1432-8364/2013/0200

Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., and Bareth, G. (2014). Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging. *Remote Sens.* 6, 10395–10412. doi: 10.3390/RS61110395

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., et al. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Observation Geoinformation* 39, 79–87. doi: 10.1016/J.JAG.2015.02.012

Chang, A., Jung, J., Maeda, M. M., and Landivar, J. (2017). Crop height monitoring with digital imagery from Unmanned Aerial System (UAS). *Comput. Electron. Agric.* 141, 232–237. doi: 10.1016/J.COMPAG.2017.07.008

Cheng, T., Song, R., Li, D., Zhou, K., Zheng, H., Yao, X., et al. (2017). Spectroscopic estimation of biomass in canopy components of paddy rice using dry matter and chlorophyll indices. *Remote Sens.* 9, 319. doi: 10.3390/RS9040319

Clevers, J. G. P. W., Kooistra, L., and van den Brande, M. M. M. (2017). Using sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sens.* 9, 405. doi: 10.3390/RS9050405

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., De Colstoun, E. B., and McMurtrey, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens. Environ.* 74, 229–239. doi: 10.1016/S0034-4257(00)00113-9

Derraz, R., Melissa Muharam, F., Nurulhuda, K., Ahmad Jaafar, N., and Keng Yap, N. (2023). Ensemble and single algorithm models to handle multicollinearity of UAV vegetation indices for predicting rice biomass. *Comput. Electron. Agric.* 205, 107621. doi: 10.1016/J.COMPAG.2023.107621

Gitelson, A. A., Vina, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., and Leavitt, B. (2003). Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical Res. Lett.* 30, 1248. doi: 10.1029/2002GL016450

Gitelson, A. A., Zur, Y., Chivkunova, O. B., and Merzlyak, M. N. (2002). Assessing carotenoid content in plant leaves with reflectance spectroscopyll. *Photochem. Photobiol.* 75, 272–281. doi: 10.1562/0031-8655(2002)0750272ACCIPL2.0.CO2

Gnyp, M. L., Miao, Y. X., Yuan, F., Ustin, S. L., Yu, K., Yao, Y. K, et al. (2014). Hyperspectral canopy sensing of paddy rice aboveground biomass at different growth stages. *Field Crop. Res.* 155, 42–55. doi: 10.1016/J.FCR.2013.09.023

Guo, A., Huang, W., Dong, Y., Ye, H., Ma, H., Liu, B., et al. (2021). Wheat yellow rust detection using UAV-based hyperspectral technology. *Remote Sens.* 13, 123. doi: 10.3390/RS13010123

Han, L., Yang, G., Dai, H., Xu, B., Yang, H., Feng, H., et al. (2019). Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods* 15, 1–19. doi: 10.1186/S13007-019-0394-Z/FIGURES/13

Hansen, P. M., and Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens. Environ.* 86, 542–553. doi: 10.1016/S0034-4257(03)00131-7

Haralick, R. M. (1973). Glossary and index to remotely sensed image pattern recognition concepts. *Pattern Recognition* 5, 391–403. doi: 10.1016/0031-3203(73)90029-0

Haralick, R. M., Dinstein, I., and Shanmugam, K. (1973). Textural features for image classification. *IEEE Trans. Systems Man Cybernetics* SMC-3, 610–621. doi: 10.1109/TSMC.1973.4309314

Herrero-Huerta, M., Bucksch, A., Puttonen, E., and Rainey, K. M. (2020a). Canopy roughness: A new phenotypic trait to estimate aboveground biomass from unmanned aerial system. *Plant Phenomics* 2020. doi: 10.34133/2020/6735967

Herrero-Huerta, M., Rodriguez-Gonzalvez, P., and Rainey, K. M. (2020b). Yield prediction by machine learning from UAS-based multi-sensor data fusion in soybean. *Plant Methods* 16, 78. doi: 10.1186/s13007-020-00620-6

Hu, H., Zhou, H., Cao, K., Lou, W., Zhang, G., Gu, Q., et al. (2024). Biomass estimation of milk vetch using UAV hyperspectral imagery and machine learning. *Remote Sens.* 16, 2183. doi: 10.3390/RS16122183

Itoh, J. I., Sato, Y., Nagato, Y., and Matsuoka, M. (2006). Formation, maintenance and function of the shoot apical meristem in rice. *Plant Mol. Biol.* 60, 827–842. doi: 10.1007/S11103-005-5579-3/METRICS

Kooistra, L., and Clevers, J. G. P. W. (2016). Estimating potato leaf chlorophyll content using ratio vegetation indices. *Remote Sens. Lett.* 7, 611–620. doi: 10.1080/2150704X.2016.1171925

Li, B., Xu, X., Zhang, L., Han, J., and Bian, C. (2020). Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *ISPRS J. Photogramm* 162, 161–172. doi: 10.1016/J.ISPRSJPRS.2020.02.013

Li, P., Zhang, X., Wang, W., Zheng, H., Yao, X., Tian, Y., et al. (2020). Estimating aboveground and organ biomass of plant canopies across the entire season of rice growth with terrestrial laser scanning. *Int. J. Appl. Earth Observation Geoinformation* 91, 102132. doi: 10.1016/J.JAG.2020.102132

Li, R., Wang, D., Zhu, B., Liu, T., Sun, C., and Zhang, Z. (2022). Estimation of nitrogen content in wheat using indices derived from RGB and thermal infrared imaging. *Field Crops Res.* 289, 108735. doi: 10.1016/J.FCR.2022.108735

Li, Z., Zhao, Y., Taylor, J., Gaulton, R., Jin, X., Song, X., et al. (2022). Comparison and transferability of thermal, temporal and phenological-based in-season predictions of above-ground biomass in wheat crops from proximal crop reflectance data. *Remote Sens. Environ.* 273, 112967. doi: 10.1016/J.RSE.2022.112967

Liu, Y., Feng, H., Yue, J., Fan, Y., Bian, M., Ma, Y., et al. (2023). Estimating potato above-ground biomass by using integrated unmanned aerial system-based optical, structural, and textural canopy measurements. *Comput. Electron. Agric.* 213, 108229. doi: 10.1016/J.COMPAG.2023.108229

Liu, Y., Feng, H., Yue, J., Fan, Y., Jin, X., Zhao, Y., et al. (2022). Estimation of potato above-ground biomass using UAV-based hyperspectral images and machine-learning regression. *Remote Sens.* 14, 5449. doi: 10.3390/rs14215449

Liu, Y., Liu, S., Li, J., Guo, X., Wang, S., and Lu, J. (2019). Estimating biomass of winter oilseed rape using vegetation indices and texture metrics derived from UAV multispectral images. *Comput. Electron. Agric.* 166, 105026. doi: 10.1016/j.compag.2019.105026

Ma, L., Chen, X., Zhang, Q., Lin, J., Yin, C., Ma, Y., et al. (2022). Estimation of nitrogen content based on the hyperspectral vegetation indexes of interannual and multi-temporal in cotton. *Agronomy* 12, 1319. doi: 10.3390/AGRONOMY12061319

Maimaitijiang, M., Sagan, V., Sidike, P., Maimaitiyiming, M., Hartling, S., Peterson, K. T., et al. (2019). Vegetation Index Weighted Canopy Volume Model (CVMVI) for soybean biomass estimation from Unmanned Aerial System-based RGB imagery. *ISPRS J. Photogrammetry Remote Sens.* 151, 27–41. doi: 10.1016/J.ISPRSJPRS.2019.03.003

Muharam, F. M., Bronson, K. F., Maas, S. J., and Ritchie, G. L. (2014). Inter-relationships of cotton plant height, canopy width, ground cover and plant nitrogen status indicators. *Field Crops Res.* 169, 58–69. doi: 10.1016/J.FCR.2014.09.008

Nichol, J. E., and Sarker, M. L. R. (2011). Improved biomass estimation using the texture parameters of two high-resolution optical sensors. *IEEE Trans. Geosci. Remote Sens.* 49, 930–948. doi: 10.1109/TGRS.2010.2068574

Niu, Y., Zhang, L., Zhang, H., Han, W., and Peng, X. (2019). Estimating above-ground biomass of maize using features derived from UAV-based RGB imagery. *Remote Sens.* 11, 1261. doi: 10.3390/RS11111261

Pugh, N. A., Horne, D. W., Murray, S. C., Carvalho, G., Malambo, L., Jung, J., et al. (2018). Temporal estimates of crop growth in sorghum and maize breeding enabled by unmanned aerial systems. *Plant Phenome J.* 1, 1–10. doi: 10.2135/TPPJ2017.08.0006

Rouse, J. W. Jr., Haas, R. H., Schell, J., and Deering, D. (1974). Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation.

Schumacher, P., Mislimshoeva, B., Brenning, A., Zandler, H., Brandt, M., Samimi, C., et al. (2016). Do red edge and texture attributes from high-resolution satellite data improve wood volume estimation in a semi-arid mountainous region? *Remote Sens.* 8, 540. doi: 10.3390/RS8070540

Shu, M., Li, Q., Ghafoor, A., Zhu, J., Li, B., and Ma, Y. (2023). Using the plant height and canopy coverage to estimation maize aboveground biomass with UAV digital images. *Eur. J. Agron.* 151, 126957. doi: 10.1016/J.EJA.2023.126957

Sinde-González, I., Gil-Docampo, M., Arza-García, M., Grefa-Sánchez, J., Yánez-Simba, D., Pérez-Guerrero, P., et al. (2021). Biomass estimation of pasture plots with multitemporal UAV-based photogrammetric surveys. *Int. J. Appl. Earth Observation Geoinformation* 101, 102355. doi: 10.1016/J.JAG.2021.102355

Tao, H., Feng, H., Xu, L., Miao, M., Long, H., Yue, J., et al. (2020). Estimation of crop growth parameters using UAV-based hyperspectral remote sensing data. *Sensors* 20, 1296. doi: 10.3390/S20051296

Varela, S., Assefa, Y., Vara Prasad, P. V., Peralta, N. R., Griffin, T. W., Sharda, A., et al. (2017). Spatio-temporal evaluation of plant height in corn via unmanned aerial systems. *J. Appl. Remote Sens.* 11, 1. doi: 10.1117/1.jrs.11.036013

Verrelst, J., Alonso, L., Camps-Valls, G., Delegido, J., and Moreno, J. (2012). Retrieval of vegetation biophysical parameters using Gaussian process techniques. *IEEE Trans. Geosci. Remote Sens.* 50, 1832–1843. doi: 10.1109/TGRS.2011.2168962

Volpato, L., Pinto, F., González-Pérez, L., Thompson, I. G., Borém, A., Reynolds, M., et al. (2021). High throughput field phenotyping for plant height using UAV-based RGB imagery in wheat breeding lines: feasibility and validation. *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.591587/BIBTEX

Wang, L., Zhou, X., Zhu, X., Dong, Z., and Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Crop J.* 4, 212–219. doi: 10.1016/J.CJ.2016.01.008

Wang, Y., Zhang, K., Tang, C., Cao, Q., Tian, Y., Zhu, Y., et al. (2019). Estimation of rice growth parameters based on linear mixed-effect model using multispectral images from fixed-wing unmanned aerial vehicles. *Remote Sens.* 11, 1371. doi: 10.3390/RS11111371

Xu, T., Wang, F., Shi, Z., and Miao, Y. (2024). Multi-scale monitoring of rice aboveground biomass by combining spectral and textural information from UAV hyperspectral images. *Int. J. Appl. Earth Observation Geoinformation* 127, 103655. doi: 10.1016/J.JAG.2024.103655

Xu, T., Wang, F., Xie, L., Yao, X., Zheng, J., Li, J., et al. (2022). Integrating the textural and spectral information of UAV hyperspectral images for the improved estimation of rice aboveground biomass. *Remote Sens.* 14, 2534. doi: 10.3390/RS14112534

Yao, X., Si, H., Cheng, T., Jia, M., Chen, Q., Tian, Y. C., et al. (2018). Hyperspectral estimation of canopy leaf biomass phenotype per ground area using a continuous wavelet analysis in wheat. *Front. Plant Sci.* 9. doi: 10.3389/FPLS.2018.01360/BIBTEX

Yue, J., Feng, H., Jin, X., Yuan, H., Li, Z., Zhou, C., et al. (2018a). A comparison of crop parameters estimation using images from UAV-mounted snapshot hyperspectral sensor and high-definition digital camera. *Remote Sens.* 10, 1138. doi: 10.3390/RS10071138

Yue, J., Feng, H., Tian, Q., and Zhou, C. (2020). A robust spectral angle index for remotely assessing soybean canopy chlorophyll content in different growing stages. *Plant Methods* 16, 1–18. doi: 10.1186/S13007-020-00643-Z/FIGURES/11

Yue, J., Feng, H., Yang, G., and Li, Z. (2018b). A comparison of regression techniques for estimation of above-ground winter wheat biomass using near-surface spectroscopy. *Remote Sens.* 10, 66. doi: 10.3390/RS10010066

Yue, J., Yang, G., Li, C., Li, Z., Wang, Y., Feng, H., et al. (2017). Estimation of winter wheat above-ground biomass using unmanned aerial vehicle-based snapshot hyperspectral sensor and crop height improved models. *Remote Sens.* 9, 708. doi: 10.3390/RS9070708

Yue, J., Yang, H., Yang, G., Fu, Y., Wang, H., and Zhou, C. (2023). Estimating vertically growing crop above-ground biomass based on UAV remote sensing. *Comput. Electron. Agric.* 205, 107627. doi: 10.1016/J.COMPAG.2023.107627

Zeng, Y., Xu, B., Yin, G., Wu, S., Hu, G., Yan, K., et al. (2018). Spectral invariant provides a practical modeling approach for future biophysical variable estimations. *Remote Sens.* 10, 1508. doi: 10.3390/RS10101508

Zha, H., Miao, Y., Wang, T., Li, Y., Zhang, J., Sun, W., et al. (2020). Improving unmanned aerial vehicle remote sensing-based rice nitrogen nutrition index prediction with machine learning. *Remote Sens.* 12, 215. doi: 10.3390/RS12020215

Zhai, W., Li, C., Fei, S., Liu, Y., Ding, F., Cheng, Q., et al. (2023). CatBoost algorithm for estimating maize above-ground biomass using unmanned aerial vehicle-based multi-source sensor data and SPAD values. *Comput. Electron. Agric.* 214, 108306. doi: 10.1016/J.COMPAG.2023.108306

Zhang, B., Zhang, L., Xie, D., Yin, X., Liu, C., and Liu, G. (2015). Application of synthetic NDVI time series blended from landsat and MODIS data for grassland biomass estimation. *Remote Sens.* 8, 10. doi: 10.3390/RS8010010

Zheng, H., Cheng, T., Zhou, M., Li, D., and Yao, X. (2019). Improved estimation of rice aboveground biomass combining textural and spectral analysis of UAV imagery. *Precis. Agric.* 20, 611–629. doi: 10.1007/S11119-018-9600-7/FIGURES/7

Zhu, Y., Zhao, C., Yang, H., Yang, G., Han, L., Li, Z., et al. (2019). Estimation of maize above-ground biomass based on stem-leaf separation strategy integrated with LiDAR and optical remote sensing data. *PeerJ* 2019, e7593. doi: 10.7717/PEERJ.7593/SUPP-10

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 10 December 2024
DOI 10.3389/fpls.2024.1492528

Check for updates

OPEN ACCESS

EDITED BY
Wenyu Zhang,
Jiangsu Academy of Agricultural Sciences Wuxi Branch, China

REVIEWED BY
Zhenwang Li,
Yangzhou University, China
Qing Gu,
Zhejiang Academy of Agricultural Sciences, China

*CORRESPONDENCE
Haitao Xiang
htxiang@issas.ac.cn

RECEIVED 07 September 2024
ACCEPTED 20 November 2024
PUBLISHED 10 December 2024

CITATION
Wang Y, Shi P, Qian Y, Chen G, Xie J, Guan X, Shi W and Xiang H (2024) Enhancing Nitrogen Nutrition Index estimation in rice using multi-leaf SPAD values and machine learning approaches.
Front. Plant Sci. 15:1492528.
doi: 10.3389/fpls.2024.1492528

COPYRIGHT
© 2024 Wang, Shi, Qian, Chen, Xie, Guan, Shi and Xiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Enhancing Nitrogen Nutrition Index estimation in rice using multi-leaf SPAD values and machine learning approaches

Yuan Wang[1], Peihua Shi[2], Yinfei Qian[3], Gui Chen[4], Jiang Xie[3], Xianjiao Guan[3], Weiming Shi[1] and Haitao Xiang[1]*

[1]State Key Laboratory of Soil and Sustainable Agriculture, Changshu National Agro-Ecosystem Observation and Research Station, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, China, [2]Department of Agronomy and Horticulture, Jiangsu Vocational College of Agriculture and Forestry, Jurong, China, [3]Soil and Fertilizer & Resources and Environmental Institute, Jiangxi Academy of Agricultural Sciences, Nanchang, China, [4]Institute of Biotechnology, Jiaxing Academy of Agricultural Science, Jiaxing, China

Accurate nitrogen diagnosis is essential for optimizing rice yield and sustainability. This study investigates the potential of using multi-leaf SPAD measurements combined with machine learning models to improve nitrogen nutrition diagnostics in rice. Conducted across five locations with 15 rice cultivars, SPAD values from the first to fifth fully expanded leaves were collected at key growth stages. The study demonstrates that integrating multi-leaf SPAD data with advanced machine learning models, particularly Random Forest and Extreme Gradient Boosting, significantly improves the accuracy of Leaf Nitrogen Concentration (LNC) and Nitrogen Nutrition Index (NNI) estimation. The second fully expanded Leaf From the Top (2LFT) emerged as the most critical variable for predicting LNC, while the 3LFT was pivotal for NNI estimation. The inclusion of statistical metrics, such as maximum and median SPAD values, further enhanced model performance, underscoring the importance of considering both original SPAD measurements and derived indices. This approach provides a more precise method for nitrogen assessment, facilitating improved nitrogen use efficiency and contributing to sustainable agricultural practices through targeted and effective nitrogen management strategies in rice cultivation.

KEYWORDS

rice nitrogen diagnosis, multi-leaf SPAD values, machine learning, leaf nitrogen concentration, nitrogen nutrition index, statistical metrics

1 Introduction

Nitrogen is one of the most essential nutrients in crop growth, particularly in rice cultivation, where it plays a key role in photosynthesis, biomass accumulation, and overall yield (Zhang et al., 2021). As a staple food for over half of the world's population, rice production is vital for global food security (Bandumula, 2018). Effective nitrogen management is crucial not only for maximizing rice yields but also for reducing environmental impacts associated with nitrogen over-application (Anas et al., 2020), such as water pollution and greenhouse gas emissions (Yu et al., 2019). Therefore, accurately understanding and monitoring the nitrogen status of rice plants is essential for optimizing nitrogen use efficiency and promoting sustainable agricultural practices (Colaço and Bramley, 2018; Berger et al., 2020).

Leaf Nitrogen Concentration (LNC) and Nitrogen Nutrition Index (NNI) are key indicators of a plant's nitrogen status (Makowski et al., 2020; Li et al., 2022a). LNC represents the nitrogen content in the most important photosynthetic and assimilation organs, while NNI provides a relative measure of the nitrogen supply in relation to the plant's needs. Accurate and timely estimation of these indices is crucial for effective nitrogen management, enabling farmers to apply the right amount of nitrogen at the right time to optimize crop performance (Yao et al., 2023).

The SPAD meter is a widely used tool for diagnosing crop nitrogen nutrition (Li et al., 2022b; Lu et al., 2022; Karaca et al., 2023; Wang et al., 2023; Wu et al., 2024). It measures chlorophyll content in leaves by comparing light absorption at specific wavelengths, which correlates with nitrogen levels. SPAD meters have achieved considerable success in diagnosing nitrogen nutrition in rice (Yuan et al., 2016b, 2016a; Ravier et al., 2017; Yang et al., 2018; Wu et al., 2024). However, these studies have often reported that diagnostic accuracy can be influenced by factors such as growth stage, cultivar, and environmental conditions (Ravier et al., 2017; Li et al., 2022b; Fu et al., 2023; Wang et al., 2023; Yao et al., 2023), leading to variability in results. Additionally, chlorophyll is only one form of nitrogen within the leaf, and a saturation effect on chlorophyll content can occur in specific leaves (Jiang et al., 2021). As a result, when chlorophyll content is high, the SPAD value is prone to being affected by this saturation effect (Gabriel et al., 2017; Souza et al., 2019).

To address these challenges, researchers have explored various methods. One approach involves using Relative SPAD (RSPAD) values to estimate nitrogen nutrition status, thereby minimizing the effects of cultivar and growth stage differences (Yue et al., 2020; Lu et al., 2022). Studies have shown that nitrogen-split application based on RSPAD can save nitrogen fertilizer and greatly improve nitrogen use efficiency (Wu et al., 2024). However, this method requires a well-fertilized reference area as a control, complicating practical applications and conflicting with the goal of simplifying nitrogen diagnostics (Ravier et al., 2017; Yue et al., 2020; Lu et al., 2022). Another approach uses SPAD indices, such as differences and ratios between SPAD values of different leaf positions, to enhance diagnostic accuracy (Zhang et al., 2019; Karaca et al., 2023). While this method shows promise, it lacks validation on large-scale datasets, limiting its applicability.

Machine learning methods offer significant advantages in overcoming these limitations by effectively handling complex nonlinear relationships in agricultural data (Chlingaryan et al., 2018). By integrating multi-leaf SPAD measurements with machine learning algorithms, it is possible to enhance the accuracy of nitrogen nutrition diagnosis in rice. This integration allows for better interpretation of SPAD data affected by factors like growth stage and cultivar, thus improving nitrogen management strategies (Yang et al., 2023; Mandal et al., 2024).

Rice leaves differ significantly in their physiological characteristics depending on their position on the plant (Sun et al., 2018). These differences can affect the accuracy of nitrogen nutrition diagnosis, as leaves at various positions respond differently to nitrogen availability (Yuan et al., 2016b; Zhao et al., 2018). For instance, younger leaves at the top of the plant are in dynamic growth phases with higher nitrogen content, while older leaves are more stable or in senescence, affecting their SPAD readings (Zhang et al., 2019; Li et al., 2022b; Wang et al., 2023). By leveraging information from multiple leaf positions and exploring the relationships between SPAD values at these positions and the plant's nitrogen status, the accuracy of nitrogen nutrition diagnosis in rice can be improved (Yuan et al., 2016a).

In this study, we conducted multi-location, multi-cultivar field trials in rice, collecting data across various growth stages to build a comprehensive dataset. We applied several machine learning methods and performed feature importance analyses to evaluate the role of multi-leaf SPAD information in enhancing nitrogen nutrition diagnosis. This research not only contributes to improving the accuracy of nitrogen diagnostics in rice but also offers insights into the future prospects of integrating machine learning with agronomic practices.

2 Materials and methods

2.1 Study area and experimental design

This study was conducted across five locations within the Jiangsu, Zhejiang, and Jiangxi provinces, situated in the middle and lower reaches of the Yangtze River in China: Suzhou, Wuxi, Zhenjiang, Jiaxing, and Yichun prefecture-level cities (Figure 1). Seventeen field experiments were conducted, involving fifteen rice cultivars—five indica and ten japonica cultivars. The field experiments were arranged in a randomized block design, with each experiment comprising two to seven nitrogen fertilizer treatments, including at least a control with no nitrogen and a conventional nitrogen application to induce varied nitrogen statuses in the rice plants. Each treatment was replicated across three to four plots, with plot sizes ranging from 30 to 100 m^2.

Abbreviations: LNC, Leaf Nitrogen Concentration; $N_{critical}$, Critical Nitrogen concentration; NNI, Nitrogen Nutrition Index; CNDC, Critical Nitrogen Dilution Curve; LFT, fully expanded Leaves From the Top; LR, Linear Regression; PLS, Partial Least Squares; SVR, Support Vector Regression; RF, Random Forest; XGB, Extreme Gradient Boosting; ATPA, Average Testing Prediction Accuracy; SHAP, SHapley Additive exPlanations.

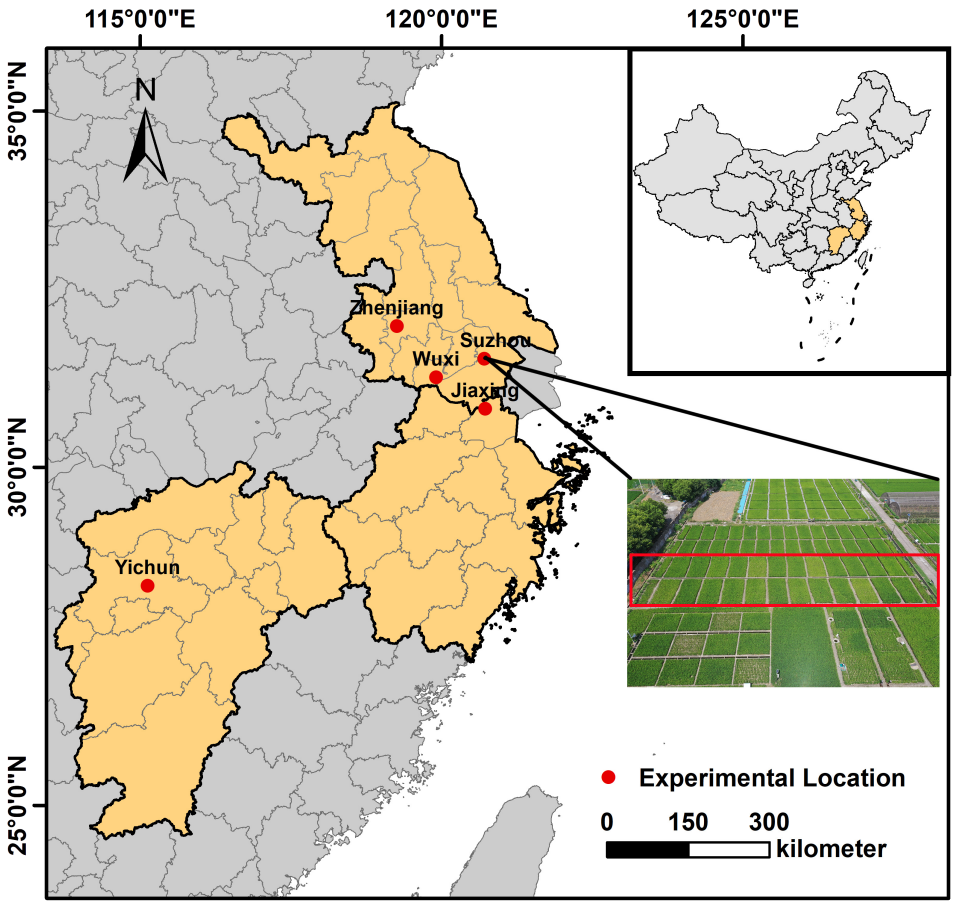


FIGURE 1
Geographic locations of the five experimental sites in Jiangsu, Zhejiang, and Jiangxi provinces, China, along with representative image of field experiments.

Detailed information on each experiment and experimental site, including soil types, is provided in Tables 1 and 2. All field management practices, except nitrogen levels, conformed to local conventional cultivation methods.

2.2 Data collection and plant sample testing

Samples were collected at regular intervals of 10-15 days throughout the rice growing season, covering the main growth stages: Tillering (TI), Stem Elongation (SE), Panicle Initiation (PI), Heading (HD), and Grain Filling (GF) (the GF stage sampling was only for the varieties NG46_1, WYG35, and TY398). Due to varietal differences and planting locations, different rice cultivars may be at different growth stages on the same sampling date. To account for this, we scheduled destructive sampling at fixed intervals and, during data analysis, grouped the samples according to the actual growth stages of each cultivar.

At each stage, three to six uniformly growing plants were destructively sampled from the experimental plots. The plants were divided into leaves, sheaths, and panicles, which were oven-dried at 105°C for 30 minutes, followed by further drying at 75°C to a constant weight. The dried samples from each organ were weighed, ground, and the nitrogen concentration was determined using the Kjeldahl method. The overall plant nitrogen concentration was calculated as the ratio of total nitrogen content to the total dry mass.

The NNI was calculated as the ratio of plant nitrogen concentration to the critical nitrogen concentration ($N_{critical}$). An NNI close to 1 indicates optimal nitrogen supply, while an NNI greater than 1 suggests excess nitrogen (where additional nitrogen does not increase biomass). An NNI less than 1 indicates nitrogen deficiency, with the severity of deficiency inversely related to the NNI value. The method for determining the $N_{critical}$ in this study was based on the approach of Justes et al. (1994). For each experiment and sampling date, variance analysis of dry matter (DM) was conducted across treatments, followed by LSD testing to identify the treatment with the highest biomass but the lowest plant nitrogen concentration, designated as the $N_{critical}$. To reduce the likelihood of Type II errors in the LSD test, the significance level was set at 0.10. For experiments with only two nitrogen levels or where the $N_{critical}$ could not be determined using this method, a Critical Nitrogen Dilution Curve (CNDC) was plotted using other

TABLE 1 Summary of field experiment information.

No.	Cultivars	Subspecies	Rice Type	Number of Nitrogen Treatments	Experimental Site
1	NG46_1	japonica	Conventional	6	Suzhou
2	NG46_2	japonica	Conventional	6	Suzhou
3	NG46_3	japonica	Conventional	6	Wuxi
4	WYG35	japonica	Conventional	6	Wuxi
5	NG5055	japonica	Conventional	7	Wuxi
6	J67	japonica	Conventional	3	Jiaxing
7	XS14	japonica	Conventional	6	Jiaxing
8	JH218	japonica	Conventional	5	Jiaxing
9	YG13	japonica	Conventional	5	Zhenjiang
10	CY5	japonica	hybrid	4	Yichun
11	CY6	japonica	hybrid	5	Suzhou
12	JYZK6	japonica	hybrid	2	Jiaxing
13	HHZ	indica	Conventional	4	Yichun
14	MXXZ	indica	Conventional	4	Yichun
15	TY398	indica	hybrid	4	Yichun
16	TYXZ	indica	hybrid	4	Yichun
17	TYHZ	indica	hybrid	4	Yichun

Each treatment has 3-4 replicates.

confirmed $N_{critical}$, and the NNI was calculated from this curve. The CNDC established in this study was $N_{critical} = 3.44DM^{-0.44}$.

Simultaneously with the destructive sampling, 15 to 20 rice plants were selected from each experimental plot to measure the SPAD values of the first to fifth fully expanded Leaves From the Top (1-5 LFT). Each leaf was measured at three positions (middle, upper third, and lower third) on one side of the main vein, and the SPAD values were averaged. SPAD measurements were performed using a SPAD-502 PLUS (Konica Minolta) chlorophyll meter.

2.3 Estimation of rice LNC and NNI

This study employed several modeling techniques to predict LNC and the NNI in rice, including univariate Linear Regression (LR), Partial Least Squares (PLS) regression, Support Vector Regression (SVR), Random Forest (RF), and Extreme Gradient Boosting (XGB).

PLS regression addresses multicollinearity by projecting predictors and responses onto latent variables that maximize their covariance. The key hyperparameter in PLS is the number of latent components, which strikes a balance between dimensionality reduction and predictive accuracy. SVR aims to find the optimal hyperplane for data fitting, with critical hyperparameters including the regularization parameter C, the kernel function [e.g., linear, polynomial, or Radial Basis Function (RBF)], the kernel coefficient γ for RBF, and the epsilon (ε) margin. RF builds a forest of decision trees, with hyperparameters such as the number of trees, maximum tree depth, minimum samples required for node splitting, and the maximum number of features considered at each split, controlling model complexity and diversity. XGB constructs sequential trees to

TABLE 2 Climatic and soil conditions for each experimental site.

Experimental Site	Average annual temperature (°C)	Annual precipitation (mm)	sunshine hours (h)	Soil Type
Suzhou	16.9	1615	1711	Gleyi-Stagnic Anthrosols
Wuxi	17.1	1553	1898	Hapli-Stagnic Anthrosols
Jiaxing	15.9	1168	2017	Hapli-Stagnic Anthrosols
Zhenjiang	15.1	1019	2116	Hapli-Stagnic Anthrosols
Yichun	17.5	1560	1780	Argi-Udic Ferrosols

Soil types are classified according to the Chinese Soil Taxonomy (CST).

correct errors, with essential hyperparameters including learning rate, number of boosting iterations, tree depth, subsample ratios for training instances and features, and L1/L2 regularization to mitigate overfitting. These models were chosen for their ability to capture complex, nonlinear relationships between SPAD measurements and nitrogen status indicators. Machine learning methods like RF and XGB are particularly effective in handling high-dimensional data and interactions among variables, which are common in agronomic datasets.

Hyperparameter tuning was conducted using Bayesian optimization through the 'Optuna' framework (Akiba et al., 2019), combined with five-fold cross-validation in Python 3.12.0. The hyperparameters tested are detailed in Table 3.

In the LR model, the SPAD value from a single leaf position was used as the input to estimate rice LNC and NNI. For the other models, which accommodate multiple input variables, the variables were grouped based on original SPAD measurements, SPAD indices between different leaf positions, and SPAD statistical metrics. Detailed information on the input variables for each group is provided in Table 4.

TABLE 3 Hyperparameters and tested range for each modeling method.

Modeling Method	Hyperparameter	Tested Range
PLS	Number of components	[1, 5], Integer
SVR	Regularization parameter (C)	[10^{-3}, 10^{3}], Log-uniform (float)
	Kernel coefficient (γ)	[10^{-3}, 10], Log-uniform (float)
	Epsilon (ε)	[10^{-4}, 1], Log-uniform (float)
RF	Number of trees	[50, 500], Integer
	Tree depth	[1, 20], Integer
	Minimum samples to split	[2, 11], Integer
	Minimum samples per leaf	[1, 11], Integer
	Maximum features for split	sqrt, log2, None
XGB	Learning rate	[10^{-3}, 10^{-1}], Log-uniform (float)
	Boosting iterations	[50, 500], Integer
	Tree depth	[3, 10], Integer
	Subsample ratio (instances)	[0.5, 1.0], Uniform (float)
	Subsample ratio (features)	[0.5, 1.0], Uniform (float)
	L1 regularization term (α)	[0.1, 1.0], Uniform (float)
	L2 regularization term (λ)	[0.1, 1.0], Uniform (float)

Values within square brackets indicate the hyperparameter search range. "Integer" denotes that the hyperparameter values are integers sampled within the specified range. "Uniform (float)" indicates that the hyperparameter values are floating-point numbers sampled uniformly from the specified range, while "Log-uniform (float)" indicates that the values are sampled from a logarithmic distribution within the specified range.

2.4 Model evaluation and feature importance analysis

The performance of each model was evaluated using the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Average Testing Prediction Accuracy (ATPA). These indicators were calculated for both the training and validation datasets, which were split in an 8:2 ratio, to assess the models' ability to generalize to unseen data. The ATPA was calculated using the following formula:

$$\mathrm{ATPA} = \left(1 - \frac{1}{n}\sum_{i=1}^{n}\frac{|TA_i - TP_i|}{TA_i}\right) \times 100$$

where TA_i represents the observed value, TP_i represents the predicted value, and n is the number of observations.

Feature importance was analyzed using SHapley Additive exPlanations (SHAP) to determine the contribution of each input variable to the model's predictions (Lundberg and Lee, 2017). SHAP values provided insights into the influence of each feature on prediction outcomes, enabling a detailed understanding of the models' decision-making processes. This analysis offered both global interpretation—by averaging SHAP values across the entire dataset to assess overall feature importance—and local interpretation, by examining individual data points to understand specific feature contributions. In this study, SHAP analysis was conducted for the RF and XGB models using the four input variable combinations (Table 4).

3 Results

3.1 Distribution of SPAD values across different leaf positions in rice

A statistical analysis of SPAD values across different leaf positions in all rice cultivars revealed significant differences in the distribution of SPAD values from the 1LFT to 5LFT (Figure 2). The SPAD values from the 1LFT to 5LFT followed a trend of initially increasing, and then gradually decreasing. The 1LFT recorded the lowest average SPAD value of 38.9, whereas the highest average SPAD value of 42.2 was observed in the 2LFT. After the peak, the average SPAD values diminished progressively, with an accompanying increase in the range of SPAD value distributions, indicating greater variability in older leaves.

Variance analysis revealed no significant differences in average SPAD values between 2LFT and 3LFT or between 4LFT and 5LFT, while significant differences were detected between the other leaf positions ($p < 0.05$). As the leaves aged from 1LFT to 5LFT, the difference between the median and mean SPAD values increased, with values of -0.09, 0.36, 0.55, 0.84, and 1.26, respectively. This suggests a growing skew in SPAD values with leaf age, where older leaves tend to have lower SPAD values, causing the mean to decline relative to the median. The statistical analysis of SPAD values at different leaf positions indicates that specific leaf positions exhibit

TABLE 4 Input variable combinations for the rice nitrogen estimation model.

Variable Combination	Description	Included Variables
comb_1	SPAD values of the five fully expanded leaves from the top of rice plant	1LFT, 2LFT, 3LFT, 4LFT, 5LFT
comb_2	Normalized difference index, ratios, and difference of SPAD values between specific leaf positions	ND2_1, ND3_1, ND3_2, ND4_1, ND4_2, ND4_3, R2_1, R3_1, R3_2, R4_1, R4_2, R4_3, D2_1, D3_1, D3_2, D4_1, D4_2, D4_3
comb_3	Combination of comb_2 variables with additional statistical metrics (std, min, max, median) of SPAD values	ND2_1, ND3_1, ND3_2, ND4_1, ND4_2, ND4_3, R2_1, R3_1, R3_2, R4_1, R4_2, R4_3, D2_1, D3_1, D3_2, D4_1, D4_2, D4_3, 1LFT_std, 1LFT_min, 1LFT_max, 1LFT_median, 2LFT_std, 2LFT_min, 2LFT_max, 2LFT_median, 3LFT_std, 3LFT_min, 3LFT_max, 3LFT_median
comb_4	Combination of all the above variables including original SPAD measurements, SPAD indices, and statistical metrics	All the above variables

1-5LFT refers to the 1st, 2nd, 3rd, 4th and 5th fully expanded leaf form the top of the rice plant, where SPAD measurements are taken. The variables NDi_j, Ri_j, and Di_j represent the Normalized Difference Index, Ratio Index, and Difference between SPAD values at iLFT and jLFT, respectively, calculated as $(SPAD_{iLFT}-SPAD_{jLFT})/(SPAD_{iLFT}+SPAD_{jLFT})$, $SPAD_{iLFT}/SPAD_{jLFT}$, and $SPAD_{iLFT}-SPAD_{jLFT}$. The variables std, min, max, and median represent the standard deviation, minimum, maximum, and median of the SPAD values from the corresponding leaves.

distinct responses to changes in nitrogen concentration and growth stages. This finding suggests that selecting appropriate leaf positions for SPAD measurements is crucial for accurate nitrogen monitoring in rice.

3.2 Simple Linear Regression for LNC and NNI estimation

A simple Linear Regression (LR) analysis was conducted using SPAD values from individual leaf positions (1LFT to 5LFT) and their average to estimate rice LNC and NNI. The overall mean LNC across all samples was 2.96%, with lower and upper quartiles of 2.42% and 3.43%. The mean NNI was 0.94, with quartiles of 0.74 and 1.12.

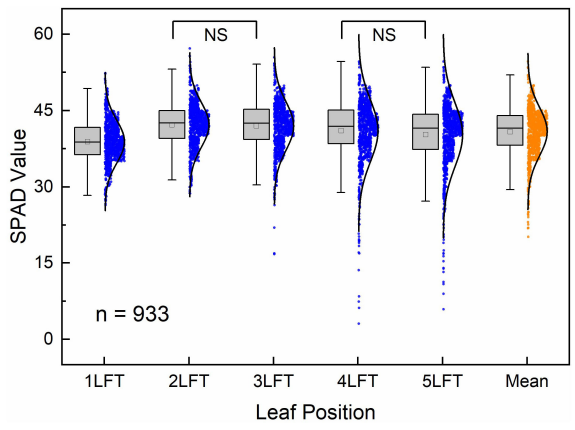


FIGURE 2
Distribution of SPAD values across different leaf positions and their means in all rice cultivars. Each box plot displays the range of SPAD values, with whiskers extending to 1.5 times the interquartile range, a horizontal line representing the median, and a point within the box indicating the mean. Accompanying density distribution curves and scatter plots provide additional insights into the data spread and distribution shape. Significant differences in SPAD values are present among all leaf positions, except those marked as "NS".

Across 17 independent experiments, SPAD values effectively estimated LNC and NNI for most cultivars, with R^2 values ranging from 0.14 to 0.92 (Tables 5, 6). Cultivars NG5055, J67, and MXXZ exhibited the highest estimation performance, while WYG35, XS14, and TY398 showed lower R^2 values for LNC estimation. Similarly, for NNI estimation, WYG35, XS14, and JYZK6 had lower R^2 values. No significant difference in estimation performance was found between japonica and indica subspecies.

The estimation accuracy varied with leaf position. The average R^2 values increased from 1LFT, peaked at 3LFT, and then decreased, with 3LFT providing the most accurate estimates for both LNC and NNI (average R^2 of 0.70). In contrast, 1LFT had the lowest estimation accuracy.

When LR was applied to the pooled dataset including all cultivars and growth stages, R^2 values decreased to a maximum of 0.49 for LNC and 0.42 for NNI, indicating that variability among cultivars and growth stages significantly impacted regression accuracy (Figure 3). For LNC estimation, 2LFT provided the highest accuracy, while for NNI estimation, 3LFT was most accurate. Using the average SPAD values from 1-5LFT offered estimation accuracy close to that of the best individual leaf positions.

Scatter plots of LNC versus SPAD values (Figure 3) revealed that early growth stages had higher nitrogen concentrations, which declined as rice developed. However, SPAD values remained similar across growth stages, causing data grouping by growth stage and reducing LR model accuracy. For NNI versus SPAD plots, NNI values clustered between 0.5 and 1.5, mitigating the grouping effect, but LR accuracy did not improve compared to LNC estimation.

Regarding rice subspecies, indica clustered in the lower SPAD value range, while japonica was concentrated in the higher SPAD value range. This difference likely reflects variations in nitrogen uptake and utilization between the two subspecies.

3.3 PLS and machine learning for LNC and NNI estimation

Using the pooled dataset, four groups of feature variables—comprising leaf SPAD values, SPAD indices, and SPAD statistical

10.3389/fpls.2024.1492528

TABLE 5 Linear regression analysis between SPAD values and LNC (%) at different leaf positions.

Cultivars	1LFT		2LFT		3LFT		4LFT		5LFT		Average SPAD	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
NG46_1	0.48(1)	0.27	0.67	0.22	0.76	0.19	0.77	0.19	0.79	0.18	0.79	0.18
NG46_2	0.63	0.24	0.73	0.19	0.73	0.20	0.64	0.21	0.71	0.20	0.74	0.19
NG46_3	0.61	0.21	0.73	0.19	0.74	0.18	0.65(1)	0.20	0.74	0.18	0.76	0.17
WYG35	0.19(2a)	0.28	0.25(1a)	0.27	0.39(1a)	0.23	0.50	0.21	0.45(a)	0.22	0.45(a)	0.22
NG5055	0.66	0.24	0.82	0.19	0.84	0.18	0.73	0.23	0.80	0.20	0.85	0.18
J67	0.67	0.20	0.82	0.14	0.79	0.16	0.80	0.16	0.71	0.18	0.82	0.14
XS14	0.38(1)	0.22	0.57(1)	0.18	0.58	0.18	0.42(1)	0.22	0.30(1)	0.24	0.51(1)	0.19
JH218	0.68	0.23	0.77	0.21	0.60	0.29	0.45(1)	0.34	0.55	0.30	0.68	0.26
YG13	0.56(a)	0.16	0.64(a)	0.14	0.65	0.15	0.43(1)	0.19	0.56	0.17	0.62	0.15
CY5	0.60(1)	0.18	0.62(1)	0.17	0.55(1)	0.19	0.51(1a)	0.20	0.63(1)	0.19	0.61(1)	0.18
CY6	0.64	0.22	0.86	0.14	0.82	0.16	0.59(1)	0.24	0.77	0.18	0.81	0.17
JYZK6	0.82	0.22	0.88	0.17	0.92	0.14	0.86	0.19	0.90	0.16	0.90	0.16
HHZ	0.69	0.26	0.53	0.31	0.61	0.28	0.57	0.30	0.49(1)	0.33	0.60	0.29
MXXZ	0.85	0.18	0.82	0.22	0.86	0.19	0.58(1)	0.30	0.72	0.27	0.82	0.20
TY398	0.44(2)	0.32	0.48(1a)	0.31	0.47(1a)	0.31	0.46(2)	0.31	0.21(2a)	0.39	0.48(1a)	0.30
TYXZ	0.43(1)	0.34	0.67	0.25	0.78	0.21	0.70	0.23	0.79	0.20	0.75	0.22
TYHZ	0.63	0.24	0.70	0.22	0.77	0.19	0.71	0.21	0.63	0.24	0.76	0.19
Mean	0.59	0.24	0.68	0.21	0.70	0.20	0.61	0.23	0.63	0.23	0.70	0.20

The R^2 and RMSE values for each cultivar represent the averages of regression analyses conducted separately across different growth stages. Numbers in parentheses indicate the number of growth stages where the regression coefficients for that cultivar are not statistically significant. The letter 'a' indicates that the regression coefficients for one growth stage are significant at the 0.1 level. All other unmarked regression models have coefficients that are significant at the 0.05 level.

metrics—were used as inputs for PLS regression, SVR, RF, and XGB models to estimate rice LNC and NNI. Compared to the results of LR using SPAD values from single leaf positions, models incorporating multiple feature variables significantly improved estimation accuracy, with R^2 values increasing from 0.3-0.4 to 0.5-0.7 (Table 7; Figure 4). Among the four modeling methods, SVR, RF, and XGB showed similar performance, with the RF model achieving the highest accuracy in estimating LNC. Notably, when comb_3 and comb_4 were used as input variables, the R^2 values reached 0.73 and 0.74, respectively, and the ATPA also achieved the highest values among all models, at 88.84 and 88.89, respectively. For NNI estimation, the RF and XGB models performed consistently, with average ATPA values of 83.44 and 83.61, respectively. While the SVR model showed slightly lower R^2 values compared to RF and XGB, its ATPA was comparable to the other two models. The PLS model, although effective, performed slightly worse than the machine learning models.

Overall, the models provided better estimation accuracy for LNC than for NNI, with average ATPA values of 86.60 and 83.07, respectively. Among the four input variable combinations, using the second combination (comb_2)—which includes normalized difference indices, ratios, and differences of SPAD values—resulted in significantly lower estimation accuracy for both LNC and NNI compared to the other three combinations, particularly for NNI estimation. In contrast, combinations that included original SPAD values from specific leaf positions (comb_1, comb_4) and SPAD statistical metrics (comb_3, comb_4) provided higher estimation accuracy for both LNC and NNI. The accuracy of the models followed the trend comb_4 > comb_1 > comb_3, indicating that the inclusion of multi-leaf position SPAD measurements and more comprehensive variable combinations improved prediction accuracy.

Using the RF model as an example, a 1:1 plot of predicted versus observed values on the validation dataset was generated (Figure 4). The distribution of data points demonstrates that the combination of multi-leaf SPAD features with machine learning methods effectively eliminated the grouping effect observed in LR models when using data from different growth stages, significantly improving the model's accuracy in estimating LNC and NNI. However, across all input variables and target variables, the models tended to underestimate in the high-value regions, particularly when using comb_2 as input. This underestimation may be attributed to the saturation effect of SPAD values in estimating nitrogen concentration and the nitrogen dilution effect observed as the plant grows.

TABLE 6 Linear regression analysis between SPAD values and NNI at different leaf positions.

Cultivars	1LFT		2LFT		3LFT		4LFT		5LFT		Average SPAD	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
NG46_1	0.45(1)	0.18	0.66	0.14	0.77	0.12	0.74	0.13	0.77	0.12	0.77	0.12
NG46_2	0.68	0.11	0.79	0.08	0.79	0.09	0.74	0.09	0.80	0.09	0.83	0.08
NG46_3	0.57	0.11	0.61(a)	0.10	0.65(a)	0.09	0.61(1)	0.10	0.67	0.09	0.68	0.09
WYG35	0.29(1)	0.13	0.29(a)	0.13	0.35(a)	0.13	0.38(a)	0.12	0.43	0.12	0.43	0.12
NG5055	0.68	0.11	0.80	0.09	0.81	0.08	0.74	0.09	0.78	0.09	0.84	0.08
J67	0.76	0.09	0.90	0.06	0.88	0.06	0.85	0.07	0.79	0.08	0.91	0.06
XS14	0.31(1)	0.13	0.56(a)	0.10	0.58	0.10	0.37(1)	0.12	0.27(1)	0.14	0.47(1)	0.11
JH218	0.46	0.12	0.64	0.10	0.81	0.07	0.68	0.09	0.81	0.07	0.81	0.07
YG13	0.47	0.06	0.69	0.05	0.67	0.05	0.53(1)	0.05	0.67	0.05	0.68	0.05
CY5	0.54(1)	0.06	0.57(1)	0.06	0.61(1)	0.06	0.60(1)	0.06	0.71(a)	0.06	0.64(1)	0.05
CY6	0.48	0.14	0.78	0.10	0.80	0.09	0.63	0.12	0.77	0.10	0.79	0.09
JYZK6	0.56(1a)	0.19	0.54(1)	0.18	0.52(1)	0.19	0.52(1a)	0.19	0.56(1)	0.18	0.54(1)	0.19
HHZ	0.62	0.13	0.60	0.13	0.60	0.13	0.62	0.13	0.50(a)	0.15	0.62	0.13
MXXZ	0.77	0.10	0.80	0.09	0.87	0.07	0.61(a)	0.13	0.63	0.13	0.79	0.10
TY398	0.59(b)	0.10	0.67(a)	0.09	0.66	0.09	0.68	0.09	0.39(1)	0.14	0.68	0.09
TYXZ	0.60	0.13	0.78	0.10	0.91	0.06	0.83	0.09	0.90	0.07	0.90	0.07
TYHZ	0.59	0.10	0.59(a)	0.10	0.64(a)	0.09	0.65	0.09	0.37(1)	0.13	0.61	0.10
Mean	0.55	0.12	0.66	0.10	0.70	0.09	0.63	0.10	0.64	0.11	0.70	0.09

The R^2 and RMSE values for each cultivar represent the averages of regression analyses conducted separately across different growth stages. Numbers in parentheses indicate the number of growth stages where the regression coefficients for that cultivar are not statistically significant. The letters 'a' or 'b' indicate that the regression coefficients for one or two growth stages, respectively, are significant at the 0.1 level. All other unmarked regression models have coefficients that are significant at the 0.05 level.

3.4 Feature importance of the random forest model

SHAP analysis was performed on the input features of the RF model to assess their impact on predicting LNC and NNI in rice. When only SPAD measurements from the 1-5LFT were used as input variables (Figure 5A), the 2LFT emerged as the most influential feature for LNC estimation, with an average SHAP value of approximately 0.35, significantly higher than those of the other leaf positions. The second most important feature was the 5LFT, with an average SHAP value slightly above 0.1. Interestingly, 3LFT, which performed well in linear regression, was ranked only fourth in importance. Generally, higher SPAD values contributed positively to LNC predictions, while lower SPAD values had a negative impact, except for 3LFT, which exhibited the opposite pattern.

When SPAD indices were used as input variables (Figure 5B), the difference SPAD indices (D2_1, D3_2, D4_2, D4_1) were the most influential for LNC prediction, followed by ratio and normalized difference indices. Adding statistical metrics as input variables (comb_3) significantly enhanced the model's predictive power, increasing the R^2 from 0.44 to 0.73. In this combination, the top five most important features were statistical metrics (Figure 5C), with the median, maximum, and minimum values of 2LFT ranking highest. When using all features (Figure 5D), the median, original measurement, and maximum values of 2LFT remained the most influential, with average SHAP values far exceeding those of other features. Across all four feature combinations, the SPAD values of 2LFT and their derived indices consistently emerged as the most important variables, underscoring the critical role of 2LFT in estimating rice LNC.

In the SHAP analyses of the comb_2, comb_3, and comb_4 combinations, variables such as D3_2, R3_2, and ND3_2 consistently demonstrated that smaller values contributed more positively to LNC, indicating that a larger SPAD difference between 2LFT and 3LFT had a more significant positive impact. However, these variables had relatively lower influence on LNC predictions, ranking 6th, 9th, and 10th in importance in comb_3, and 4th, 6th, and 7th in comb_4. Although D3_2 ranked 2nd in comb_2, this combination had lower overall explanatory power for LNC.

For NNI estimation, SHAP analysis revealed that when using only SPAD values as inputs (comb_1), 3LFT, 4LFT, and 1LFT were the most influential features, with 3LFT having the highest average SHAP value (Figure 6). The SHAP value distributions indicated that higher SPAD values at these leaf positions positively contributed to NNI predictions, suggesting better nitrogen status in the plant. When using SPAD indices as input variables (comb_2), R4_2,

10.3389/fpls.2024.1492528

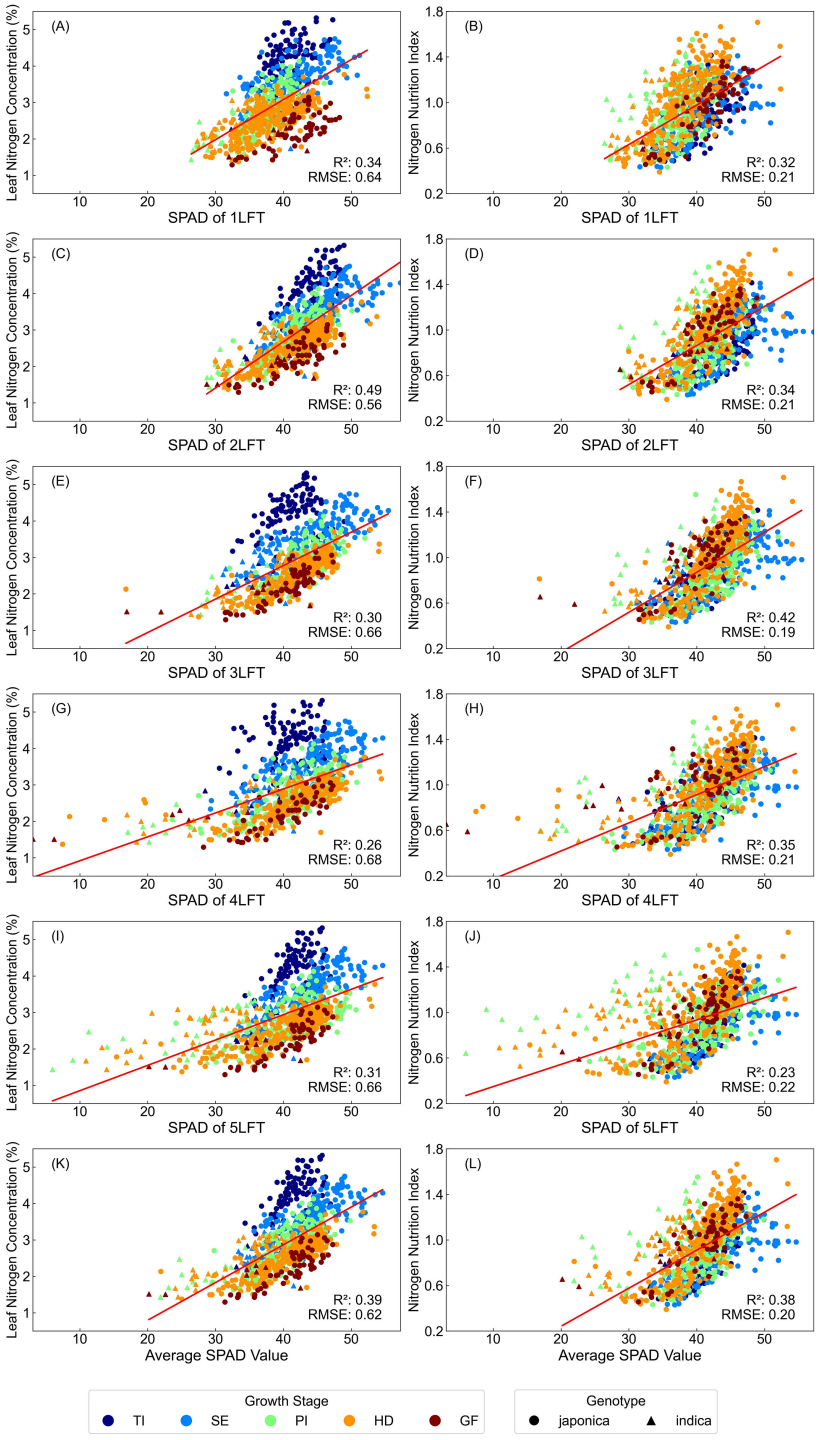


FIGURE 3
Linear regression analysis between SPAD readings from different leaf positions and LNC and NNI across different growth stages and cultivars (n=933). **(A–J)** represent scatter plots for SPAD readings from the 1st to 5th leaf from the top (LFT) and their corresponding LNC **(A, C, E, G, I)** or NNI **(B, D, F, H, J)** values. **(K, L)** show the scatter plots for the average SPAD readings across the 1st to 5th LFT and their corresponding LNC **(K)** or NNI **(L)** values.

ND4_2, and D4_1 emerged as the most critical features, with larger values leading to higher NNI predictions. This pattern suggests that older leaves with higher SPAD values reflect a plant in a well-supplied nitrogen state.

When statistical metrics (such as median, minimum, maximum, and standard deviation) were added to the input variables (comb_3), the most important features were the minimum, median, and maximum SPAD values of 3LFT. Higher values of these metrics corresponded to higher NNI predictions, emphasizing that the distribution characteristics of SPAD measurements within a plot are key indicators of plant nitrogen status. When using all variables (comb_4), the model's

TABLE 7 Validation results of different modeling methods based on the pooled dataset.

Modeling Method	Input Variable	LNC Estimation			NNI Estimation		
		R^2	RMSE(%)	ATPA	R^2	RMSE	ATPA
PLS	comb_1	0.59	0.53	85.55	0.55	0.18	82.62
	comb_2	0.53	0.56	83.85	0.32	0.22	77.68
	comb_3	0.62	0.51	86.22	0.52	0.19	82.37
	comb_4	0.62	0.51	86.17	0.58	0.18	83.83
SVR	comb_1	0.71	0.44	88.58	0.68	0.15	86.18
	comb_2	0.51	0.57	82.66	0.33	0.22	78.29
	comb_3	0.69	0.46	88.16	0.58	0.18	84.09
	comb_4	0.69	0.46	88.24	0.66	0.16	85.90
RF	comb_1	0.68	0.46	87.60	0.69	0.15	86.23
	comb_2	0.44	0.61	82.13	0.34	0.22	78.03
	comb_3	0.73	0.42	88.84	0.62	0.17	84.21
	comb_4	0.74	0.42	88.89	0.67	0.16	85.30
XGB	comb_1	0.68	0.47	87.34	0.68	0.15	85.93
	comb_2	0.58	0.53	84.73	0.36	0.22	78.61
	comb_3	0.69	0.45	88.34	0.63	0.17	84.12
	comb_4	0.69	0.46	88.31	0.69	0.15	85.77

Model training and validation, using an 8:2 split, were conducted on a pooled dataset comprising data from all varieties and growth stages. The results shown in the table are from the model's performance on the validation dataset.

reliance on 3LFT SPAD values (3LFT, 3LFT_min, and 3LFT_median) remained strong, followed by 4LFT, consistent with LR results.

Additional validation results for LNC and NNI using the XGB model, along with SHAP analyses of feature importance, are provided in the Supplementary Materials (Supplementary Figures S1-S3). These results offer further insights into the model's performance and the key variables influencing nitrogen estimation in rice.

4 Discussion

4.1 The relationship between rice leaf characteristics and plant nitrogen nutrition

The physiological state of rice leaves varies significantly depending on their position on the plant and their developmental stage (Yuan et al., 2016a; Sun et al., 2018). These variations critically influence leaf SPAD values and their relationship with LNC and the NNI. Younger leaves, particularly those at the top of the plant, are the primary growth centers and therefore receive prioritized resource allocation. These leaves typically undergo rapid growth, with quick expansion of leaf area and substantial increases in nitrogen and chlorophyll content. However, this dynamic growth phase results in greater measurement variability for the SPAD values of the uppermost leaves (Yuan et al., 2016b). Consequently, the correlation between SPAD values of the 1LFT and the plant's overall nitrogen status tends to be lower compared to other leaf positions (Tables 5, 6), consistent with findings from previous studies (Hu et al., 2014; Wang et al., 2014).

In contrast, the 2LFT and 3LFT are typically in a more stable phase of growth and physiological activity. These leaves are crucial for photosynthesis and nutrient assimilation, as they maintain relatively high levels of chlorophyll and nitrogen concentration. Research on different leaf layers within the rice canopy has shown that the nitrogen concentration in the second layer from the top (corresponding to 2LFT and 3LFT) is more stable compared to the lower and uppermost layers (He et al., 2022). This stability closely reflects the canopy-level nitrogen status, making it a crucial indicator for assessing the plant's nitrogen status.

Meanwhile, the 4LFT and 5LFT typically enter or are already in the senescence phase, characterized by a decline in chlorophyll content, nitrogen concentration, and leaf area compared to 2LFT and 3LFT. This process aligns with the physiological redistribution of nutrients, wherein nutrients, including nitrogen, are translocated from older leaves to younger, actively growing parts of the plant. Under nitrogen-deficient conditions, this nutrient translocation occurs earlier and more prominently, leading to a marked decrease in SPAD values and skewed distribution patterns. Although SPAD values for 4LFT and 5LFT are more variable than those for 2LFT and 3LFT, the chlorosis and nutrient

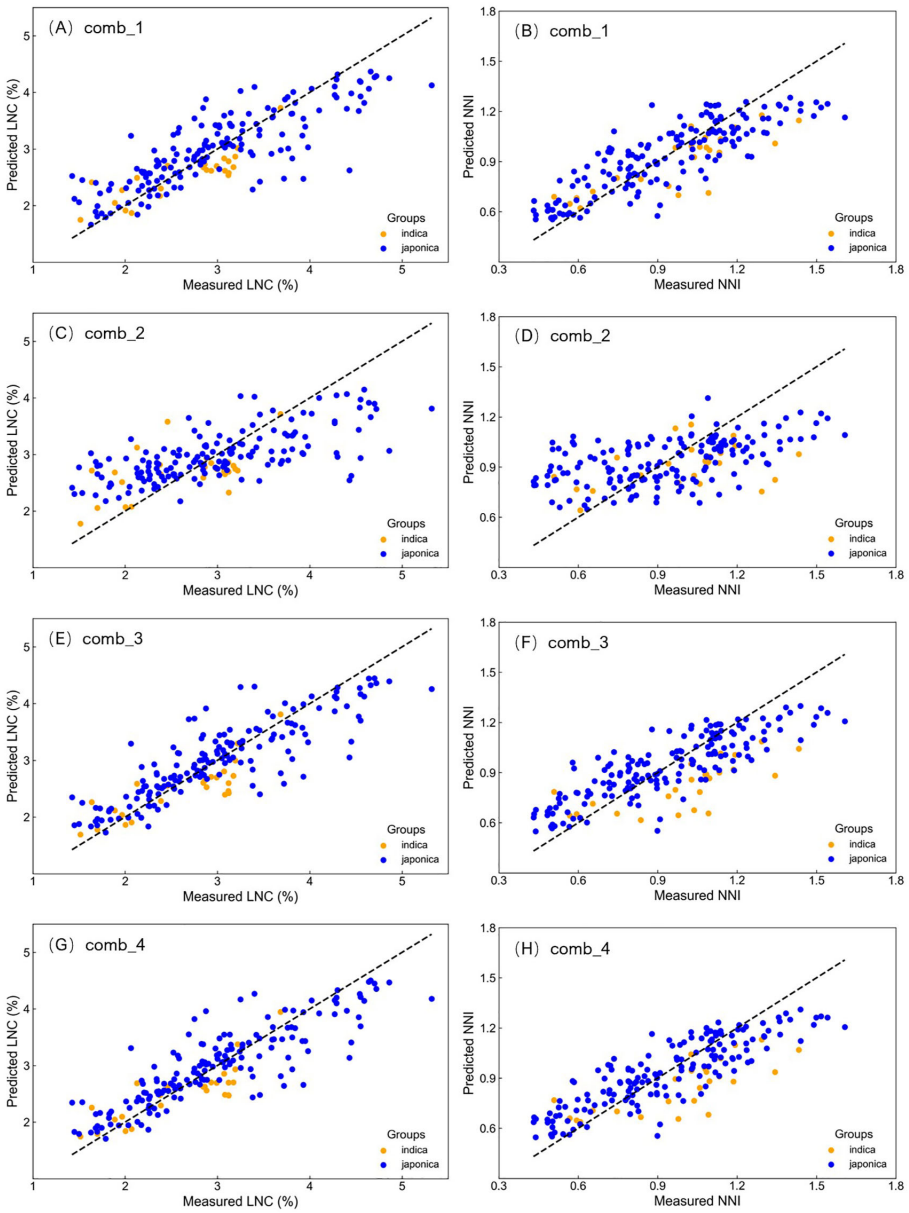


FIGURE 4
Validation results for predicted vs. measured LNC **(A, C, E, G)** and NNI **(B, D, F, H)** using random forest with input variable combinations comb_1, comb_2, comb_3, and comb_4, respectively. Each plot shows 1:1 scatter relationships.

translocation observed in these leaves provide valuable insights into the plant's overall nutritional status, making them useful indicators for diagnosing nitrogen nutrition in rice (Li et al., 2022b).

4.2 The role of multi-leaf variables and machine learning models in estimating plant nitrogen

Growth stages, rice cultivars, and environmental factors can significantly influence the accuracy of nitrogen nutrition diagnosis in crops (Huang et al., 2019; He et al., 2022; Lu et al., 2022). Delloye et al. (2018) reported that the relationship between canopy chlorophyll content retrieved from Sentinel-2 and actual nitrogen absorption in wheat was influenced by canopy structure complexity and saturation effects at different growth stages. One of the challenges in using SPAD measurements for nitrogen estimation is the saturation effect at high chlorophyll and nitrogen levels, where SPAD values plateau and become less sensitive to increases in nitrogen concentration (Gabriel et al., 2017; Souza et al., 2019). This saturation effect can reduce the sensitivity and accuracy of nitrogen diagnosis, particularly under high nitrogen availability.

In this study, the explanatory power of the linear regression between SPAD values and LNC and NNI significantly decreased after data consolidation, indicating substantial variability in these relationships across different growth stages and rice cultivars. This

10.3389/fpls.2024.1492528

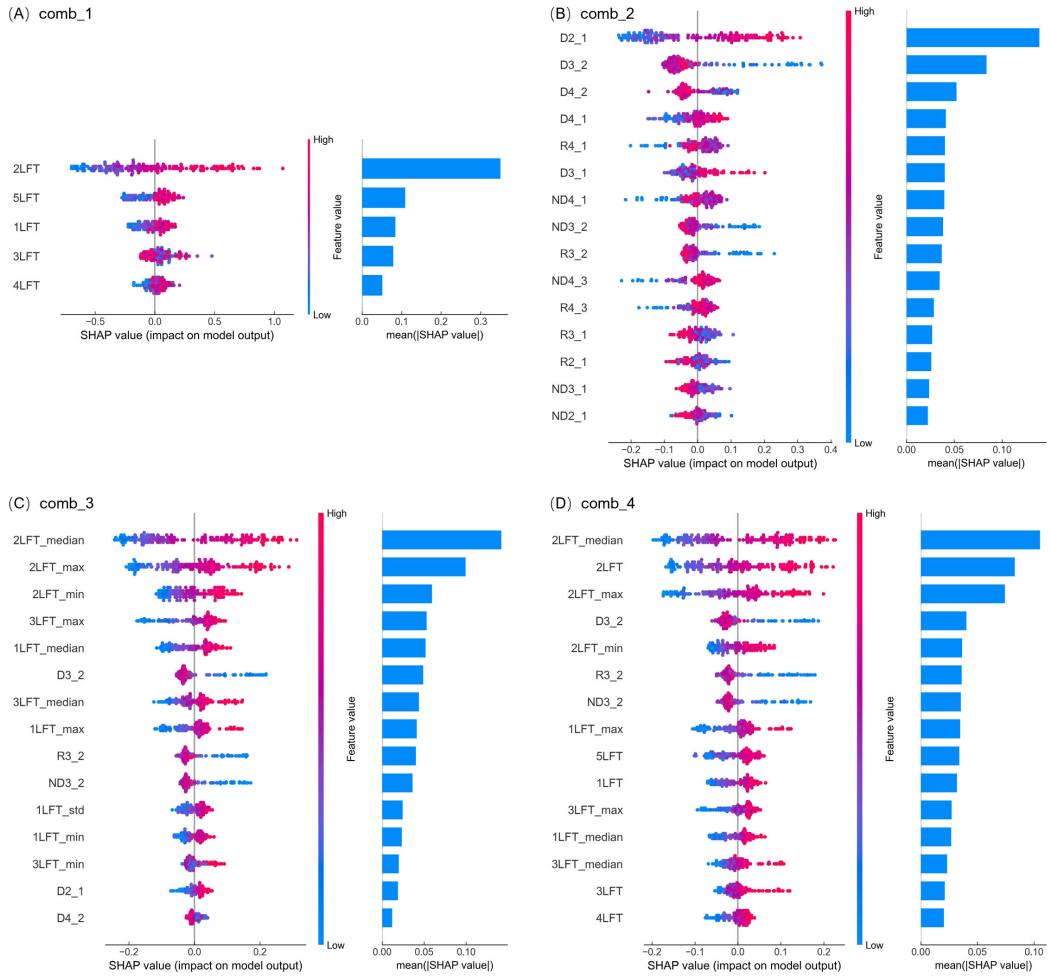


FIGURE 5
SHAP analysis of feature importance in predicting rice LNC with the random forest model across four input variable combinations. **(A–D)** correspond to the results for input variable combinations comb_1, comb_2, comb_3, and comb_4, respectively. The left side of each subplot shows SHAP value distributions, where each dot represents a data point from the validation dataset. The x-axis indicates the feature's impact on the model's prediction, and the dot color reflects the feature value, ranging from blue (low) to pink (high). Positive SHAP values indicate that the feature increases the prediction, while negative values indicate a decrease. The right side presents a bar chart (sharing the Y-axis with the left-side chart), with features ranked in descending order of their average absolute SHAP values, highlighting the most influential variables for the model's accuracy.

variability suggests that SPAD values from a single leaf position are insufficient for accurately predicting LNC or NNI. However, the SPAD values from 1-5LFT exhibit certain patterns of variation across different growth stages and nitrogen levels. By leveraging information from multiple leaf positions, especially those less prone to saturation, and incorporating features that reflect variability (e.g., minimum and median SPAD values), the models mitigate the impact of saturation on prediction accuracy.

Several studies have also demonstrated the contribution of multi-leaf position data to enhancing the accuracy of crop nitrogen nutrition diagnosis models (Lin et al., 2010; Zhao et al., 2018; Zhang et al., 2019). The use of Dualex (a leaf-clip meter that measures chlorophyll and flavonoid content) measurements from multiple leaf positions, combined with key environmental and management variables in multiple linear regression models, was critical in enhancing the precision of maize NNI estimation (Dong et al., 2021). By employing the Normalized SPAD Index (NSI), especially NSI4, Yuan et al. (2016a) achieved a notable improvement in the accuracy of nitrogen accumulation estimation, effectively reducing the influence of non-nitrogen-related factors.

Traditional statistical methods often rely on single or a limited number of variables, which can overlook the complex, nonlinear relationships that exist within biological data. In contrast, machine learning techniques offer significant advantages in managing multivariate inputs, allowing for the identification of intricate patterns and interactions (Chlingaryan et al., 2018). In our study, machine learning models like Random Forest (RF) and Extreme Gradient Boosting (XGB) demonstrated robustness in handling the nonlinear relationships caused by saturation effects. These models capture complex patterns by considering interactions among multiple features, including SPAD values from different leaf positions and statistical metrics. By leveraging data from leaves less affected by saturation and incorporating features that reflect variability, the models enhance the estimation of nitrogen status even when individual SPAD measurements reach their upper limits.

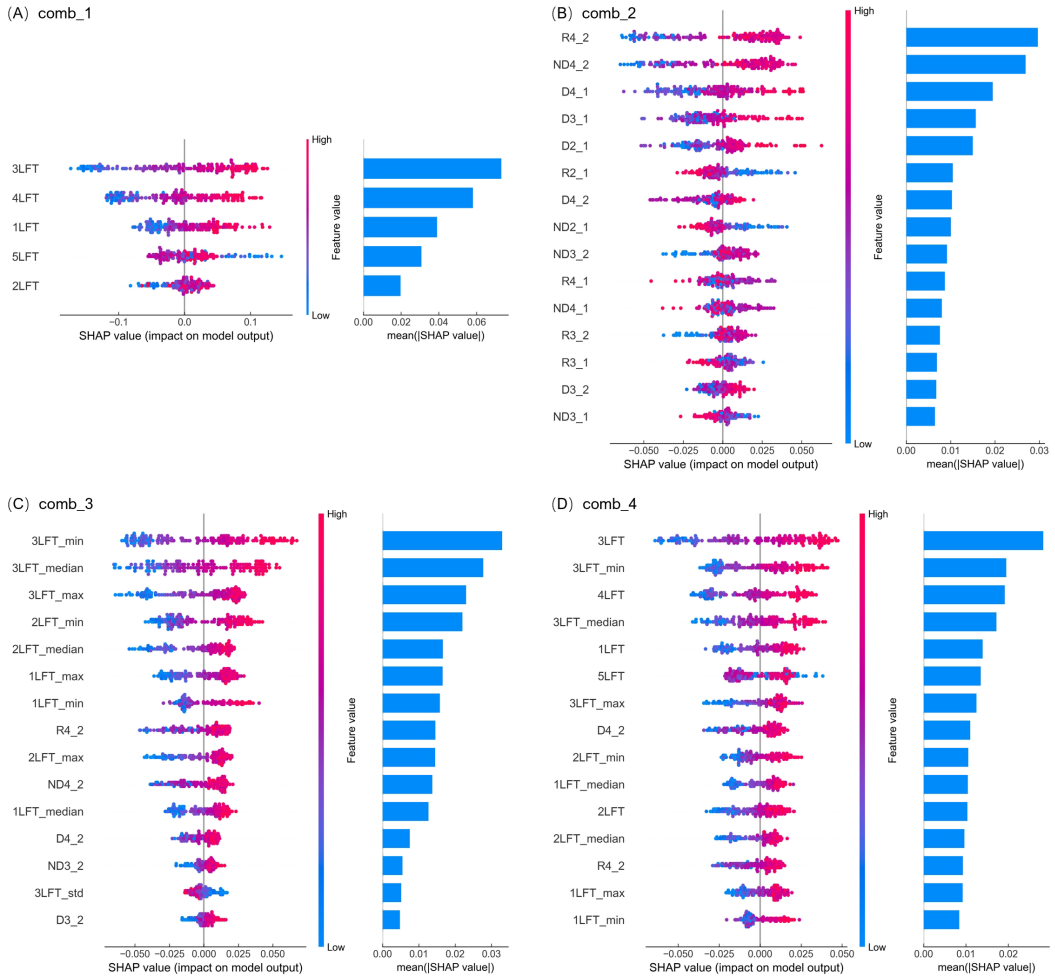


FIGURE 6
SHAP analysis of feature importance in predicting rice NNI with the random forest model across four input variable combinations. **(A–D)** correspond to the results for input variable combinations comb_1, comb_2, comb_3, and comb_4, respectively. The left side of each subplot shows SHAP value distributions, where each dot represents a data point from the validation dataset. The x-axis indicates the feature's impact on the model's prediction, and the dot color reflects the feature value, ranging from blue (low) to pink (high). Positive SHAP values indicate that the feature increases the prediction, while negative values indicate a decrease. The right side presents a bar chart (which shares the Y-axis with the left-side chart), with features ranked from top to bottom in descending order of their average absolute SHAP values, highlighting the most influential variables for the model's accuracy.

Among the four variable combinations tested, comb_4 consistently demonstrated superior performance in both LNC and NNI predictions. The superior performance of comb_4 can be attributed to its ability to leverage both raw SPAD values and derived indices. The inclusion of statistical metrics enhances the model's robustness by accounting for variability and extremities in the data, which are critical for accurately assessing nitrogen status. Previous studies have highlighted the importance of using comprehensive feature sets in machine learning models, particularly in agricultural applications where complex biological processes often underlie the observed data (Lobell et al., 2015; Shi et al., 2021; Yang et al., 2023; Mandal et al., 2024). The success of comb_4 in this study aligns with these findings, underscoring the importance of detailed feature engineering in developing accurate predictive models for nitrogen estimation.

Previous research on diagnosing nitrogen nutrition in rice using leaf SPAD values has predominantly focused on normalized, ratio, or difference indices of SPAD, often achieving satisfactory estimation results (Lin et al., 2010; Zhao et al., 2018; Dong et al., 2021). These methods have been effective when applied to models built for single cultivars and specific growth stages, where SPAD values typically exhibit consistent patterns in relation to nitrogen supply. However, this study's findings suggest that when using variable combination comb_2, which includes normalized difference indices, ratios, and differences between major leaf positions, the prediction accuracy was notably the lowest among the four tested combinations. This unexpected result may be attributed to the broader scope of this study, which encompassed multiple cultivars and growth stages, introducing additional complexity and variability that was not accounted for in previous studies. Under multi-cultivar and multi-growth-stage conditions, the generally narrow range of SPAD indices may fail to adequately capture the distinct nitrogen characteristics associated with different phenological and environmental factors. The variability introduced by these factors likely disrupts the otherwise consistent relationship between SPAD indices and nitrogen levels (Yuan et al., 2016a). As a

result, the model using comb_2, which relies heavily on SPAD indices, may struggle to generalize across diverse datasets, leading to reduced prediction accuracy.

In contrast, variable combination comb_3 demonstrated significantly improved prediction accuracy. Unlike comb_2, comb_3 incorporated a broader range of features, including minimum, maximum, and median SPAD values. These features not only expand the range of input variables but also introduce statistical information that captures the variability and distribution of SPAD measurements in certain leaf positions. By doing so, comb_3 provides a more comprehensive representation of nitrogen status, allowing the model to more effectively distinguish between different growth stages and cultivars. This approach enhances the model's robustness and generalization capability, resulting in better performance across a wide range of conditions.

4.3 Feature importance analysis in random forest models

The feature importance analysis in this study highlights the pivotal roles of the 2LFT and the 3LFT in estimating LNC and NNI, respectively. The findings indicate that higher SPAD values in 2LFT and 3LFT contribute significantly to increases in LNC and NNI, underscoring these leaf positions as key indicators of the plant's nitrogen status.

This conclusion is supported by data from 17 independent experiments involving 15 rice cultivars, which consistently showed that 2LFT and 3LFT had the highest average SPAD values, with 2LFT slightly higher than 3LFT. The prominence of 2LFT can be attributed to its critical role in nitrogen allocation during the vegetative growth stage. As one of the younger, actively growing leaves, 2LFT receives a larger proportion of the plant's nitrogen resources, which are essential for photosynthesis and biomass accumulation (Li et al., 2017; Wang et al., 2023). The relatively higher nitrogen concentration and larger leaf area of 2LFT increase its overall nitrogen content, making it a more accurate representative of the plant's total leaf nitrogen status.

In contrast, 3LFT, which transitions from maturity to senescence, is identified as the most critical leaf position for estimating NNI. Previous research has shown that leaf development and senescence are closely linked to nitrogen availability, with low soil nitrogen supply accelerating leaf chlorosis and senescence (Heyneke et al., 2019; Li et al., 2022b). As 3LFT moves from maturity to senescence, its SPAD value becomes a crucial marker of the plant's nitrogen status, especially in the early stages of vegetative growth when the number of leaves is limited. This makes 3LFT a valuable indicator for assessing plant NNI, particularly under conditions of nitrogen deficiency, where early onset of senescence in this leaf position can signal broader nutritional challenges within the plant (Sun et al., 2018).

While multi-leaf SPAD measurements enhance estimation accuracy, we recognize that collecting data from multiple leaves may not be feasible for all farmers, particularly those with limited resources or on smaller farms. To balance complexity and practicality, focusing on the most critical leaf positions identified by our feature importance analysis offers a viable solution. By concentrating on SPAD measurements from 2LFT for LNC prediction and 3LFT for NNI estimation, data collection can be simplified without substantially compromising accuracy. Implementing this targeted approach can make the method more accessible and practical for widespread adoption, facilitating efficient nitrogen management in rice cultivation. Future studies should explore the development of user-friendly tools or protocols that assist farmers in easily collecting SPAD data from these specific leaf positions.

5 Conclusion

This study demonstrates the effectiveness of using multi-leaf SPAD values and advanced machine learning models—particularly RF and XGB—to accurately estimate LNC and the NNI in rice. The 2LFT consistently emerged as the most critical variable for LNC prediction, while the 3LFT was pivotal for NNI estimation. Incorporating statistical metrics, such as maximum and median SPAD values, significantly enhanced model performance, underscoring the importance of considering both original SPAD measurements and derived indices.

While multi-leaf SPAD measurements improve estimation accuracy, focusing on the key leaves identified by our analysis can simplify data collection, making the method more practical for farmers. Targeted monitoring of SPAD values in these specific leaf positions can improve the precision of nitrogen assessments, thereby enhancing crop management practices and optimizing nitrogen use efficiency for sustainable agriculture. Future research should focus on refining these models under varying environmental conditions, exploring their applicability to other crops, and addressing challenges related to model generalizability and integration with other data sources.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

YW: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Project administration, Methodology, Investigation, Conceptualization. PS: Writing – review & editing, Resources, Project administration, Methodology. YQ: Writing – review & editing, Resources, Project administration, Methodology. GC: Writing – review & editing, Resources, Project administration, Methodology. JX: Writing – review & editing, Resources, Methodology. XG: Writing – review & editing, Resources, Methodology. WS: Writing – review & editing, Resources, Methodology, Investigation. HX:

Writing – review & editing, Supervision, Resources, Methodology, Investigation, Conceptualization.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by the National Natural Science Foundation of China (No. 32072681, 32360530, 32101616), Frontier Project from the Institute of Soil Science, Chinese Academy of Sciences (ISSASIP2406), Yafu Technology Innovation and Service Major Project of Jiangsu Vocational College of Agriculture and Forestry, China (No.2024kj01), the Fund Support Project of Jiangsu Vocational College of Agriculture and Forestry, China (No. 2021kj15), and the Innovative and Entrepreneurial Talents Project of Jiangxi Province (No. S2021DQKJ0001).

Acknowledgments

We would like to express our sincere gratitude to all individuals who contributed to the field management and plant sampling efforts throughout the course of this study. Their dedication and hard work were crucial to the success of the large-scale field trials, ensuring the accurate collection of data across multiple locations and growth stages. We also acknowledge the use of ChatGPT (version GPT-4o) for language refinement and grammar correction of the manuscript. The AI tool was employed to enhance the clarity and coherence of the text under the guidance and supervision of the authors, ensuring that the scientific content and interpretations were solely determined by the authors.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1492528/full#supplementary-material

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). "Optuna: a next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining* (Association for Computing Machinery, New York, NY, USA), 2623–2631. doi: 10.1145/3292500.3330701

Anas, M., Liao, F., Verma, K. K., Sarwar, M. A., Mahmood, A., Chen, Z. L., et al. (2020). Fate of nitrogen in agriculture and environment: agronomic, eco-physiological and molecular approaches to improve nitrogen use efficiency. *Biol. Res.* 53, 47. doi: 10.1186/s40659-020-00312-4

Bandumula, N. (2018). Rice production in Asia: key to global food security. *Proc. Natl. Acad. Sci. India Sect. B Biol. Sci.* 88, 1323–1328. doi: 10.1007/s40011-017-0867-7

Berger, K., Verrelst, J., Féret, J. B., Wang, Z., Wocher, M., Strathmann, M., et al. (2020). Crop nitrogen monitoring: recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sens. Environ.* 242, 111758. doi: 10.1016/j.rse.2020.111758

Chlingaryan, A., Sukkarieh, S., and Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput. Electron. Agric.* 151, 61–69. doi: 10.1016/j.compag.2018.05.012

Colaço, A. F., and Bramley, R. G. V. (2018). Do crop sensors promote improved nitrogen management in grain crops? *Field Crops Res.* 218, 126–140. doi: 10.1016/j.fcr.2018.01.007

Delloye, C., Weiss, M., and Defourny, P. (2018). Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. *Remote Sens. Environ.* 216, 245–261. doi: 10.1016/j.rse.2018.06.037

Dong, R., Miao, Y., Wang, X., Chen, Z., and Yuan, F. (2021). Improving maize nitrogen nutrition index prediction using leaf fluorescence sensor combined with environmental and management variables. *Field Crops Res.* 269, 108180. doi: 10.1016/j.fcr.2021.108180

Fu, Z., Zhang, R., Zhang, J., Zhang, K., Cao, Q., Tian, Y., et al. (2023). Potential of establishing the universal critical nitrogen dilution curve for japonica rice. *Plant Phenomics* 5, 36. doi: 10.34133/plantphenomics.0036

Gabriel, J. L., Zarco-Tejada, P. J., López-Herrera, P. J., Pérez-Martín, E., Alonso-Ayuso, M., and Quemada, M. (2017). Airborne and ground level sensors for monitoring nitrogen status in a maize crop. *Biosyst. Eng.* 160, 124–133. doi: 10.1016/j.biosystemseng.2017.06.003

He, J., Ma, J., Cao, Q., Wang, X., Yao, X., Cheng, T., et al. (2022). Development of critical nitrogen dilution curves for different leaf layers within the rice canopy. *Eur. J. Agron.* 132, 126414. doi: 10.1016/j.eja.2021.126414

Heyneke, E., Watanabe, M., Erban, A., Duan, G., Buchner, P., Walther, D., et al. (2019). Effect of senescence phenotypes and nitrate availability on wheat leaf metabolome during grain filling. *Agronomy* 9, 305. doi: 10.3390/agronomy9060305

Hu, Y., Yang, J., Lv, Y., and He, J. (2014). SPAD values and nitrogen nutrition index for the evaluation of rice nitrogen status. *Plant Prod. Sci.* 17, 81–92. doi: 10.1626/pps.17.81

Huang, S., Miao, Y., Yuan, F., Cao, Q., Ye, H., Lenz-Wiedemann, V. I. S., et al. (2019). In-season diagnosis of rice nitrogen status using proximal fluorescence canopy sensor at different growth stages. *Remote Sens.* 11, 1847. doi: 10.3390/rs11161847

Jiang, J., Wang, C., Wang, H., Fu, Z., Cao, Q., Tian, Y., et al. (2021). Evaluation of three portable optical sensors for non-destructive diagnosis of nitrogen status in winter wheat. *Sensors* 21, 5579. doi: 10.3390/s21165579

Justes, E., Mary, B., Meynard, J. M., Machet, J. M., and Thelier-Huche, L. (1994). Determination of a critical nitrogen dilution curve for winter wheat crops. *Ann. Bot.* 74, 397–407. doi: 10.1006/anbo.1994.1133

Karaca, C., Thompson, R. B., Peña-Fleitas, M. T., Gallardo, M., and Padilla, F. M. (2023). Evaluation of absolute measurements and normalized indices of proximal optical sensors as estimators of yield in muskmelon and sweet pepper. *Remote Sens.* 15, 2174. doi: 10.3390/rs15082174

Li, J., Feng, Y., Mou, G., Xu, G., Luo, Q., Luo, K., et al. (2017). Construction and application effect of the leaf value model based on SPAD value in rice. *Sci. Agric. Sin.* 50, 4714–4724. doi: 10.3864/j.issn.0578-1752.2017.24.005

Li, X., Ata-UI-Karim, S. T., Li, Y., Yuan, F., Miao, Y., Kato, Y., et al. (2022a). Advances in the estimations and applications of critical nitrogen dilution curve and

nitrogen nutrition index of major cereal crops. A review. *Comput. Electron. Agric.* 197, 106998. doi: 10.1016/j.compag.2022.106998

Li, Y. Y., Ming, B., Fan, P. P., Liu, Y., Wang, K. R., Hou, P., et al. (2022b). Effects of nitrogen application rates on the spatio-temporal variation of leaf SPAD readings on the maize canopy. *J. Agric. Sci.* 160, 32–44. doi: 10.1017/S0021859621001052

Lin, F. F., Qiu, L. F., Deng, J. S., Shi, Y. Y., Chen, L. S., and Wang, K. (2010). Investigation of SPAD meter-based indices for estimating rice nitrogen status. *Comput. Electron. Agric.* 71S, S60–S65. doi: 10.1016/j.compag.2009.09.006

Lobell, D. B., Thau, D., Seifert, C., Engle, E., and Little, B. (2015). A scalable satellite-based crop yield mapper. *Remote Sens. Environ.* 164, 324–333. doi: 10.1016/j.rse.2015.04.021

Lu, J., Lu, J., Nie, W., Song, J., Zhan, Q., Wang, M., et al. (2022). Estimation of nitrogen nutrition index in chrysanthemum using chlorophyll meter readings. *Ind. Crops Prod.* 187, 115459. doi: 10.1016/j.indcrop.2022.115459

Lundberg, S. M., and Lee, S. I. (2017). "A unified approach to interpreting model predictions," in *Advances in neural information processing systems*. Eds. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et al (Curran Associates, Inc, Long Beach, CA, USA). Available at: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

Makowski, D., Zhao, B., Ata-Ul-Karim, S. T., and Lemaire, G. (2020). Analyzing uncertainty in critical nitrogen dilution curves. *Eur. J. Agron.* 118, 126076. doi: 10.1016/j.eja.2020.126076

Mandal, D., Siqueira, R., Longchamps, L., and Khosla, R. (2024). Machine learning and fluorosensing for estimation of maize nitrogen status at early growth-stages. *Comput. Electron. Agric.* 225, 109341. doi: 10.1016/j.compag.2024.109341

Ravier, C., Quemada, M., and Jeuffroy, M. (2017). Use of a chlorophyll meter to assess nitrogen nutrition index during the growth cycle in winter wheat. *Field Crops Res.* 214, 73–82. doi: 10.1016/j.fcr.2017.08.023

Shi, P., Wang, Y., Xu, J., Zhao, Y., Yang, B., Yuan, Z., et al. (2021). Rice nitrogen nutrition estimation with RGB images and machine learning methods. *Comput. Electron. Agric.* 180, 105860. doi: 10.1016/j.compag.2020.105860

Souza, R., Peña-Fleitas, M. T., Thompson, R. B., Gallardo, M., Grasso, R., and Padilla, F. M. (2019). The use of chlorophyll meters to assess crop N status and derivation of sufficiency values for sweet pepper. *Sensors* 19, 2949. doi: 10.3390/s19132949

Sun, Y., Zhu, S., Yang, X., Weston, M. V., Wang, K., Shen, Z., et al. (2018). Nitrogen diagnosis based on dynamic characteristics of rice leaf image. *PloS One* 13, e0196298. doi: 10.1371/journal.pone.0196298

Wang, G., Bronson, K. F., Thorp, K. R., Mon, J., and Badaruddin, M. (2014). Multiple leaf measurements improve effectiveness of chlorophyll meter for durum wheat nitrogen management. *Crop Sci.* 54, 817–826. doi: 10.2135/cropsci2013.03.0160

Wang, Z., Li, H., Xia, Y., Fan, M., Zhao, H., Lu, J., et al. (2023). Relationship between blade SPAD value and protein content of rice in cold region. *Agric. Res. Arid. Areas* 41, 150–161. doi: 10.7606/j.issn.1000-7601.2023.05.16

Wu, G., Chen, X., Zang, Y., Ye, Y., Qian, X., Zhang, W., et al. (2024). An optimized strategy of nitrogen-split application based on the leaf positional differences in chlorophyll meter readings. *J. Integr. Agric.* 23, 2605–2617. doi: 10.1016/j.jia.2023.07.014

Yang, H., Yang, J. P., Li, F. H., and Liu, N. (2018). Replacing the nitrogen nutrition index by SPAD values and analysis of effect factors for estimating rice nitrogen status. *Agron. J.* 110, 545–554. doi: 10.2134/agronj2017.09.0532

Yang, H., Yin, H., Li, F., Hu, Y., and Yu, K. (2023). Machine learning models fed with optimized spectral indices to advance crop nitrogen monitoring. *Field Crops Res.* 293, 108844. doi: 10.1016/j.fcr.2023.108844

Yao, B., Ata-Ul-Karim, S. T., Li, Y., Ye, T., Zhu, Y., Cao, W., et al. (2023). Plant nitrogen status at phenological stages can well estimate wheat yield and its components. *Field Crops Res.* 297, 108950. doi: 10.1016/j.fcr.2023.108950

Yu, C., Huang, X., Chen, H., Godfray, H. C. J., Wright, J. S., Hall, J. W., et al. (2019). Managing nitrogen to restore water quality in China. *Nature* 567, 516–520. doi: 10.1038/s41586-019-1001-1

Yuan, Z., Ata-Ul-Karim, S. T., Cao, Q., Lu, Z., Cao, W., Zhu, Y., et al. (2016a). Indicators for diagnosing nitrogen status of rice based on chlorophyll meter readings. *Field Crops Res.* 185, 12–20. doi: 10.1016/j.fcr.2015.10.003

Yuan, Z., Cao, Q., Zhang, K., Ata-Ul-Karim, S. T., Tian, Y., Zhu, Y., et al. (2016b). Optimal leaf positions for SPAD meter measurement in rice. *Front. Plant Sci.* 7, 719. doi: 10.3389/fpls.2016.00719

Yue, X., Hu, Y., Zhang, H., and Schmidhalter, U. (2020). Evaluation of both SPAD reading and SPAD index on estimating the plant nitrogen status of winter wheat. *Int. J. Plant Prod.* 14, 67–75. doi: 10.1007/s42106-019-00068-2

Zhang, K., Liu, X., Tahir Ata-Ul-Karim, S., Lu, J., Krienke, B., Li, S., et al. (2019). Development of chlorophyll-meter-index-based dynamic models for evaluation of high-yield japonica rice production in Yangtze River Reaches. *Agronomy* 9, 106. doi: 10.3390/agronomy9020106

Zhang, X., Zou, T., Lassaletta, L., Mueller, N. D., Tubiello, F. N., Lisk, M. D., et al. (2021). Quantification of global and national nitrogen budgets for crop production. *Nat. Food* 2, 529–540. doi: 10.1038/s43016-021-00318-5

Zhao, B., Ata-Ul-Karim, S. T., Liu, Z., Zhang, J., Xiao, J., Liu, Z., et al. (2018). Simple assessment of nitrogen nutrition index in summer maize by using chlorophyll meter readings. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00011

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 16 December 2024
DOI 10.3389/fpls.2024.1518272

Check for updates

OPEN ACCESS

EDITED BY
Wenyu Zhang,
Jiangsu Academy of Agricultural Sciences Wuxi Branch, China

REVIEWED BY
Ben Zhao,
Chinese Academy of Agricultural Sciences, China
Kai Zhou,
Nanjing Forestry University, China

*CORRESPONDENCE
Yinfei Qian
qyftfs@163.com
Gui Chen
chenzhao2004@163.com

†These authors have contributed equally to this work

RECEIVED 28 October 2024
ACCEPTED 27 November 2024
PUBLISHED 16 December 2024

CITATION
Shi P, Wang Y, Yin C, Fan K, Qian Y and Chen G (2024) Mitigating saturation effects in rice nitrogen estimation using Dualex measurements and machine learning. *Front. Plant Sci.* 15:1518272. doi: 10.3389/fpls.2024.1518272

COPYRIGHT
© 2024 Shi, Wang, Yin, Fan, Qian and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Mitigating saturation effects in rice nitrogen estimation using Dualex measurements and machine learning

Peihua Shi[1], Yuan Wang[2], Congfei Yin[1], Kaiqing Fan[1], Yinfei Qian[3*†] and Gui Chen[4*†]

[1]Department of Agronomy and Horticulture, Jiangsu Vocational College of Agriculture and Forestry, Jurong, China, [2]State Key Laboratory of Soil and Sustainable Agriculture, Changshu National Agro-Ecosystem Observation and Research Station, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, China, [3]Soil and Fertilizer & Resources and Environmental Institute, Jiangxi Academy of Agricultural Sciences, Nanchang, China, [4]Institute of Biotechnology, Jiaxing Academy of Agricultural Science, Jiaxing, China

Nitrogen is essential for rice growth and yield formation, but traditional methods for assessing nitrogen status are often labor-intensive and unreliable at high nitrogen levels due to saturation effects. This study evaluates the effectiveness of flavonoid content (Flav) and the Nitrogen Balance Index (NBI), measured using a Dualex sensor and combined with machine learning models, for precise nitrogen status estimation in rice. Field experiments involving 15 rice varieties under varying nitrogen application levels collected Dualex measurements of chlorophyll (Chl), Flav, and NBI from the top five leaves at key growth stages. Incremental analysis was performed to quantify saturation effects, revealing that chlorophyll measurements saturated at high nitrogen levels, limiting their reliability. In contrast, Flav and NBI remained sensitive across all nitrogen levels, accurately reflecting nitrogen status. Machine learning models, particularly random forest and extreme gradient boosting, achieved high prediction accuracy for leaf and plant nitrogen concentrations (R^2 > 0.82), with SHAP analysis identifying NBI and Flav from the top two leaves as the most influential predictors. By combining Flav and NBI measurements with machine learning, this approach effectively overcomes chlorophyll-based saturation limitations, enabling precise nitrogen estimation across diverse conditions and offering practical solutions for improved nitrogen management in rice cultivation.

KEYWORDS

rice nitrogen estimation, Dualex measurements, saturation effect, incremental analysis, machine learning, nitrogen balance index, SHAP analysis

1 Introduction

Nitrogen is a fundamental nutrient for rice growth and yield formation, playing a crucial role in processes such as photosynthesis, protein synthesis, and enzyme activity (Liu et al., 2021; Luo et al., 2021). Effective nitrogen management not only boosts rice yield and quality but also reduces environmental pollution from excessive fertilizer use, thereby supporting sustainable agriculture (Guo et al., 2020; Cai et al., 2023). Therefore, accurately monitoring the nitrogen nutritional status of rice is vital for developing effective fertilization strategies.

Traditional nitrogen assessment methods largely depend on laboratory analyses, which are often cumbersome, time-consuming, and expensive, making them impractical for large-scale, real-time monitoring (Berger et al., 2020). To address the need for fast, non-destructive measurements, devices such as the SPAD chlorophyll meter have been widely used (Tan et al., 2021; Singh et al., 2022). These tools estimate chlorophyll content by analyzing the optical properties of leaves, thus providing an indirect measure of nitrogen levels. However, chlorophyll measurements often show saturation effects at high nitrogen levels, limiting their ability to accurately capture differences in nitrogen content (Jiang et al., 2021; Gao et al., 2024). This drawback affects their precision under high-nitrogen conditions, hindering accurate nitrogen management.

To overcome these limitations, multispectral and hyperspectral remote sensing technologies at the canopy level have been introduced for agricultural nitrogen monitoring (Fu et al., 2021; Qian et al., 2022). These methods utilize the reflectance characteristics of vegetation across various spectral bands to develop indices such as the Normalized Difference Vegetation Index (NDVI) and the Optimized Soil-Adjusted Vegetation Index (OSAVI). These indices help estimate crop nitrogen content and biomass (Qian et al., 2022; Yang et al., 2023; Li et al., 2024). Although remote sensing is beneficial for large-scale, non-contact monitoring, it is also subject to limitations. Data reliability can be affected by weather conditions, observation angles, and variations in light, leading to inconsistencies (Ma et al., 2020; Wu et al., 2023). Furthermore, at high biomass or nitrogen levels, these indices may also exhibit saturation effects, resulting in reduced accuracy in nitrogen estimation (Zhang et al., 2019b; Chang et al., 2022).

Saturation effects present a major challenge in nitrogen estimation. When measurements saturate at high nitrogen levels, distinguishing between different nitrogen statuses becomes difficult, regardless of whether chlorophyll content or canopy-level indices are used. This leads to reduced accuracy in high-nitrogen conditions, complicating the implementation of precise nitrogen management. Therefore, new approaches that can mitigate or eliminate saturation effects are urgently needed to enhance the accuracy and reliability of nitrogen estimation.

Abbreviations: Chl, Chlorophyll; Flav, Flavonoid; NBI, Nitrogen Balance Index; LNC, Leaf Nitrogen Concentration; PNC, Plant Nitrogen Concentration; NNI, Nitrogen Nutrition Index; PLS, Partial Least Squares Regression; SVR, Support Vector Regression; RF, Random Forest; XGB, Extreme Gradient Boosting; NN, Neural Network; SHAP, SHapley Additive exPlanations; RMSE, Root Mean Square Error; RRMSE, Relative Root Mean Square Error.

Flavonoids (Flav) are a diverse group of phenolic compounds that are abundantly present in plants and play key roles in their responses to biotic and abiotic stresses (Gitelson et al., 2017; Hamdane et al., 2023). Under stress conditions, such as nitrogen deficiency, Flav levels increase, making them reliable indicators of nitrogen stress (Li et al., 2021; Waqas et al., 2023). The portable Dualex sensor facilitates the simultaneous measurement of chlorophyll (Chl) and Flav in leaves, allowing the calculation of the Nitrogen Balance Index (NBI), which is derived as the ratio of Chl to Flav (Cerovic et al., 2012, 2015). Measurements of Chl and Flav have been effective in differentiating genotypic responses to nitrogen and water stress in wheat (Raya-Sereno et al., 2024). Studies have shown that the Dualex sensor maintains a stronger linear correlation with actual chlorophyll concentration than devices like SPAD-502 and CCM-200, especially at higher chlorophyll levels. It also demonstrates higher reliability in tests on crops such as maize, soybean, and wheat, being less susceptible to the effects of uneven chlorophyll distribution (Dong et al., 2019, 2022). Dong et al. (2021) successfully estimated the Nitrogen Nutrition Index (NNI) in maize across various growth stages by applying multiple linear regression models that integrated Dualex readings with environmental variables. Overall, Dualex measurements prove to be valuable in estimating plant chlorophyll and nitrogen content, as well as in distinguishing different levels of nitrogen and water stress, particularly maintaining high sensitivity even under high nitrogen conditions (Rubo and Zinkernagel, 2022). However, there is still a lack of comprehensive research on the role of Flav and NBI in rice nitrogen estimation and their performance under varying nitrogen conditions.

Nitrogen distribution within rice plants shows variability across different leaf positions (Chang et al., 2019; Ouyang et al., 2021). Karaca et al. (2023) highlighted that optical sensor data collected from various leaf positions can improve the estimation of crop nitrogen status. In crops like melon and sweet pepper, data from lower leaves demonstrated superior predictive performance at certain growth stages. Another study reported that Dualex sensor measurements across different leaf positions performed similarly when estimating the NNI. However, data from the third fully expanded Leaf From the Top (LFT) often showed higher predictive accuracy, especially during the later growth stages of maize (Dong et al., 2021). Some studies also suggest that variations in leaf layers or specific measurement positions do not significantly impact the assessment of nitrogen content in wheat when using spectral reflectance (Röll et al., 2019). In rice, SPAD readings taken from the upper-middle section of the fourth LFT have been found to provide more accurate and stable reflections of nitrogen status (Yuan et al., 2016). These findings suggest that relying on a single leaf position may not comprehensively capture the overall nitrogen status of the plant. Therefore, using multi-leaf, multi-indicator data could improve nitrogen diagnosis accuracy. However, integrating data from multiple leaf positions and indicators to build accurate nitrogen estimation models remains a challenge that requires further investigation.

Machine learning models are particularly effective at handling complex nonlinear relationships and multivariate data, making them well-suited for agricultural data analysis and prediction tasks (Shi et al., 2021; Huang et al., 2023; Mandal et al., 2024).

Combining machine learning with Dualex measurements offers the potential to enhance nitrogen estimation accuracy and mitigate saturation effects. Based on this framework, our study systematically investigates the relationships between Dualex measurements and rice nitrogen indicators through multi-location, multi-variety, and multi-leaf-position testing. We assess the effectiveness of Flav and NBI in reducing saturation effects at high nitrogen levels. We also develop rice nitrogen nutritional indicator estimation models by integrating machine learning, focusing on optimizing feature selection and model structure to enhance interpretability and practical application. By addressing the limitations of traditional methods under high nitrogen conditions and extending applicability across different growth stages, we aim to ultimately improve the efficiency and sustainability of agricultural production.

2 Materials and methods

2.1 Study area and experimental design

Field experiments were conducted across Jiangsu, Zhejiang, and Jiangxi provinces in China. These experiments followed a randomized block design, incorporating 3 to 6 different nitrogen application rates, each with 3 to 4 replicates. A total of 15 rice varieties were tested, including 10 japonica, 4 indica, and 1 indica-japonica hybrid, covering a range of regions and cultivation systems. Basic details of these varieties are shown in Table 1. Except for variations in nitrogen application, all other agronomic practices followed local conventional standards.

2.2 Data collection

Samples were obtained every 10-15 days from the Tillering (TI) stage to the Heading (HD) stage, corresponding to key growth stages: TI, Stem Elongation (SE), Panicle Initiation (PI), and HD. Due to differences in growth durations among the rice varieties, some early-maturing varieties had entered the Grain Filling (GF) stage when most varieties were at HD. Consequently, a small number of samples were collected during the GF stage for these varieties. At each sampling time, 3 to 6 hills of uniformly growing plants were destructively sampled from each plot. Samples were separated into leaves, sheaths, and panicles, then oven-dried at 105°C for 30 minutes, followed by 75°C until a constant weight was achieved. Nitrogen concentration was determined using the Kjeldahl method.

The NNI was derived by calculating the ratio of Plant Nitrogen Concentration (PNC) to the Critical Nitrogen concentration (N_c). NNI values greater than 1 indicate nitrogen surplus, 1 indicates sufficiency, and less than 1 indicates deficiency.

$$NNI = \frac{PNC}{N_c}$$

where N_c is calculated using the nitrogen dilution curve established from our dataset:

$$N_c = 3.44 \times W^{-0.43}$$

where W represents the above-ground dry matter ($t \cdot ha^{-1}$). The parameters (3.44 and -0.43) were derived from our experimental data using the method described by Justes et al. (1994). This method involves plotting plant nitrogen concentration against biomass and

TABLE 1 Summary of field experiments and nitrogen treatments.

No.	Varieties	Subspecies	Experimental Site	Growth Period	Nitrogen Rate
1	NG3908	japonica	Nanjing	May–Oct	0, 60, 120, 180, 240, 300
2	NG5055	japonica	Nanjing	May–Oct	0, 100, 150, 200, 250, 300
3	NG46	japonica	Zhenjiang	May–Oct	0, 60, 120, 180, 240, 300
4	YG13	japonica	Zhenjiang	May–Oct	0, 75, 150, 225,300
5	NG3	japonica	Zhenjiang	May–Oct	0, 75, 150, 225,300
6	JH218	japonica	Jiaxing	May–Oct	0, 80, 120, 140, 160
7	JYZK6	japonica	Jiaxing	May–Oct	0, 200, 300
8	XS14	japonica	Jiaxing	May–Oct	0, 100, 200
9	YY15	Hybrid[1]	Jiaxing	May–Oct	0, 189, 230, 270
10	J67	japonica	Jiaxing	May–Oct	0, 127, 169
11	HHZ	indica	Yichun	Jun–Oct	0, 135, 180, 225
12	MXXZ	indica	Yichun	Jun–Oct	0, 135, 180, 225
13	CY5	japonica	Yichun	Jun–Oct	0, 135, 180, 225
14	TYXZ	indica	Yichun	Jun–Oct	0, 135, 180, 225
15	TYHZ	indica	Yichun	Jun–Oct	0, 135, 180, 225

[1]This rice variety is an indica-japonica hybrid bred by the Ningbo Academy of Agricultural Sciences.

fitting a power function to the minimum nitrogen concentration observed at each biomass level, representing the N_c required for optimal growth.

On the same day as the destructive sampling, 10 to 20 rice plants were selected from each plot to measure the 1-5 LFT using a Dualex Scientific+ instrument (Force-A Inc., France). Each leaf was measured at three positions: middle, upper one-third, and lower one-third, and the average value was calculated for that leaf. The Dualex Scientific+ is a portable optical sensor designed for rapid, non-destructive detection of various physiological parameters in plant leaves. By measuring the transmission and reflection of specific wavelengths of light, it provides readings for Chl, Flav, Anthocyanin content (Anth), and the NBI. NBI is the ratio of Chl to Flav, serves as an indirect measure of the plant's nitrogen nutritional status. Although the Dualex Scientific+ instrument also measures Anth, preliminary analyses indicated that anthocyanin levels in rice leaves were consistently low and showed minimal variation across treatments and growth stages. Consequently, Anth was not included in the subsequent analyses.

During early growth stages, such as the TI stage, it was sometimes not possible to identify five fully expanded leaves in a shoot due to the limited number of leaves. In these cases, we measured at least the 1-4 LFT and calculated the average values. This average was then used as the value for the 5 LFT. This data preprocessing step ensured consistency across samples and prevented errors due to unequal data when applying machine learning models.

2.3 Modeling methods

The study employed several modeling approaches including Simple Linear Regression (LR), Quadratic Curve Regression (QCR), Partial Least Squares Regression (PLS), and four machine learning models—Support Vector Regression (SVR), Random Forest (RF), Extreme Gradient Boosting (XGB), and Neural Network (NN)—to estimate rice Leaf Nitrogen Concentration (LNC), PNC, and NNI. The hyperparameters for each machine learning model were selected based on preliminary analyses that involved testing various parameter combinations using grid search and cross-validation. This approach helped balance model complexity and performance, ensuring robust generalization. By comparing the performance of these models, their applicability in estimating rice nitrogen nutrition indices was evaluated.

2.3.1 Simple linear regression and quadratic curve regression

Initial analyses used LR and QCR to examine the relationships between individual Dualex measurements (Chl, Flav, NBI) and nitrogen indices (LNC, PNC, NNI) to gain a preliminary understanding of variable interactions and fitting effectiveness.

2.3.2 Partial least squares regression

PLS regression, a linear method suitable for multivariate, highly correlated data, was employed. It extracts latent variables that maximize the covariance between independent and dependent variables. In this study, the PLS model used two principal components to balance complexity and explanatory power.

2.3.3 Support vector regression

SVR, an application of support vector machines in regression, is capable of handling nonlinear and high-dimensional data. The SVR model utilized a radial basis function kernel, with the following key hyperparameters:

- Penalty parameter C = 1.0: Controls model complexity to prevent overfitting.
- Epsilon ε = 0.2: Defines the width of the support vector boundary, influencing the model's robustness.

2.3.4 Random forest

RF is an ensemble learning method that constructs multiple decision trees and averages their outcomes to reduce overfitting. Key hyperparameters of the RF model included:

- Number of trees $n_{\text{estimators}}$ = 30: Ensures sufficient model diversity.
- Maximum depth *max_depth* = 5: Limits tree depth to prevent overfitting.

2.3.5 Extreme gradient boosting

XGB, a tree-based ensemble model, employs gradient boosting for efficient and accurate predictions. Its key hyperparameters were set as follows:

- Number of trees $n_{\text{estimators}}$ = 30.
- Maximum depth *max_depth* = 4: Limits tree depth to prevent overfitting.
- Learning rate *learning_rate* = 0.1: Regulates each tree's contribution; a smaller rate enhances generalization.
- Regularization parameters *reg_alpha* = 0.5 and *reg_lambda* = 0.5: Apply L1 and L2 regularization to prevent overfitting.

2.3.6 Neural network

The NN model was constructed as a multilayer perceptron regressor with the following structure and parameters:

- Hidden layer size *hidden_layer_sizes* = (100),: One hidden layer with 100 neurons.
- Maximum iterations *max_iter* = 500: Restricts training time.
- Activation function: Rectified Linear Unit (ReLU) to enhance nonlinear representation.
- Learning rate: Adaptive learning rate to optimize training.

2.3.7 Model training and validation

For LR and QCR, the models used average values of Chl, Flav, or NBI from the 1-5 LFT as single input variables. For PLS and machine learning models, four different variable combinations served as inputs:

- Combination 1: Chl measurements from the 1-5 LFT and their average (six variables).
- Combination 2: Flav measurements from the 1-5 LFT and their average.
- Combination 3: NBI measurements from the 1-5 LFT and their average.
- Combination 4: All variables from the three combinations above.

A 10-fold cross-validation was employed to randomly divide the dataset into training and validation sets, ensuring reliable and stable model evaluation. This method reduces the model's dependence on data partitioning and enhances generalization.

2.4 Saturation effect analysis

To evaluate the saturation effects of different spectral measurements on rice nitrogen indices, we conducted an incremental change analysis that included the following steps:

2.4.1 Data normalization

Normalization was performed using Min-Max scaling on both independent (X) and dependent (Y) variables to make them comparable.

$$X_{scaled} = \frac{X - X_{\min}}{X_{\max} - X_{\min}},\ Y_{scaled} = \frac{Y - Y_{\min}}{Y_{\max} - X_{\min}}$$

2.4.2 Sorting

Data were sorted in ascending order based on X_{scaled} to facilitate the calculation of incremental changes, ensuring accurate reflection of *X's* impact on *Y*.

2.4.3 Increment calculation

We quantified the effect of X_{scaled} changes on Y_{scaled} by computing increments between adjacent data points.

$$\Delta X_i = X_{scaled,i+1} - X_{scaled,i},\ \Delta Y_i = Y_{scaled,i+1} - Y_{scaled,i}$$

2.4.4 Threshold for ΔX

A minimum $|\Delta X_i|$ threshold of 1×10^{-3} was applied to exclude data points with very small ΔX_i values. This step prevents excessively large rates of change in the subsequent calculations, ensuring stability and accuracy in the rate computations.

$$\Delta X_i = \begin{cases} \Delta X_i, & if\,|\Delta X_i| \geq 1 \times 10^{-3} \\ NaN, & if\,|\Delta X_i| < 1 \times 10^{-3} \end{cases}$$

2.4.5 Rate of change computation

The rate of change was determined to assess Y_{scaled} variations relative to X_{scaled}, with rates scaled to the range [-1, 1] using the maximum absolute rate.

$$Rate_i = \frac{\Delta Y_i}{\Delta X_i},\ Rate_Scaled_i = \frac{Rate_i}{max(|Rate|)}$$

2.4.6 Data smoothing

A moving average with a window size of 5 was used to smooth the rates, reducing random noise and emphasizing overall trends.

$$\text{Rate_Smoothed}_i = \frac{1}{5}\sum_{j=i-2}^{i+2} \text{Rate_Scaled}_j$$

2.4.7 Density curve analysis

To examine the behavior at high nitrogen levels, we focused on the top 20% of Y_{scaled} values and used Gaussian Kernel Density Estimation (KDE) on the corresponding X_{scaled} to obtain a smooth estimate of their probability density function. This non-parametric method does not assume any specific underlying distribution, making it suitable for real-world data. It helps in assessing whether the measurements continue to respond to increases in nitrogen levels or exhibit saturation.

2.5 Feature importance analysis and model evaluation

To gain deeper insights into the contributions of each input variable to the model's predictions, we employed the SHAP (SHapley Additive exPlanations) method (Lundberg and Lee, 2017), which is based on game theory, to analyze feature importance for the RF and XGB models. SHAP values quantify the impact of each feature on the prediction outcomes by calculating its marginal contribution across all possible feature combinations. Compared to traditional feature importance metrics, the SHAP method offers consistency and additivity, enhancing model interpretability and providing guidance for optimizing measurement strategies in practical applications.

To comprehensively evaluate the predictive performance of each model, we used the coefficient of determination (R^2), Root Mean Square Error (RMSE), and Relative Root Mean Square Error (RRMSE) as evaluation metrics. The formulas for these indicators are as follows:

- **Coefficient of determination**:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

- **Root mean square error**:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y})^2}$$

- **Relative root mean square error**:

$$RRMSE = \frac{RMSE}{\bar{Y}} \times 100\,\%$$

where Y_i is the actual value, $\hat{Y}_i$ is the predicted value, $\bar{Y}$ is the mean of the actual values, and n is the number of samples.

2.6 Statistical analysis tools

All statistical analyses and model construction in this study were performed using Python 3.12.4. Data processing and analysis utilized libraries such as Pandas, NumPy, and SciPy. Machine learning modeling was done using Scikit-learn, XGB, TensorFlow, and Keras, while visualization was carried out using Matplotlib. Statistical tests were conducted with Statsmodels and SciPy.stats, and feature importance analysis was performed using the SHAP library. To ensure reproducibility of results, all random processes were set with the same random seed during modeling and testing.

3 Results

3.1 Relationship between Dualex measurements and rice nitrogen indicators

Average Dualex values (Chl, Flav, NBI) from the first to fifth LFT were analyzed using LR and QCR models to explore their relationship with rice nitrogen nutritional indicators—LNC, PNC, and the NNI (Figure 1). The findings revealed notable differences in the effectiveness of these indicators for nitrogen estimation depending on the specific Dualex metrics used.

Chl displayed a nonlinear relationship with both LNC and PNC (Figures 1A, D). As Chl levels increased, both LNC and PNC rose, showing accelerated growth rates at Chl values around 30. However, a saturation effect was evident between Chl values of 35 to 40, where further increases in Chl did not lead to corresponding changes in LNC and PNC.

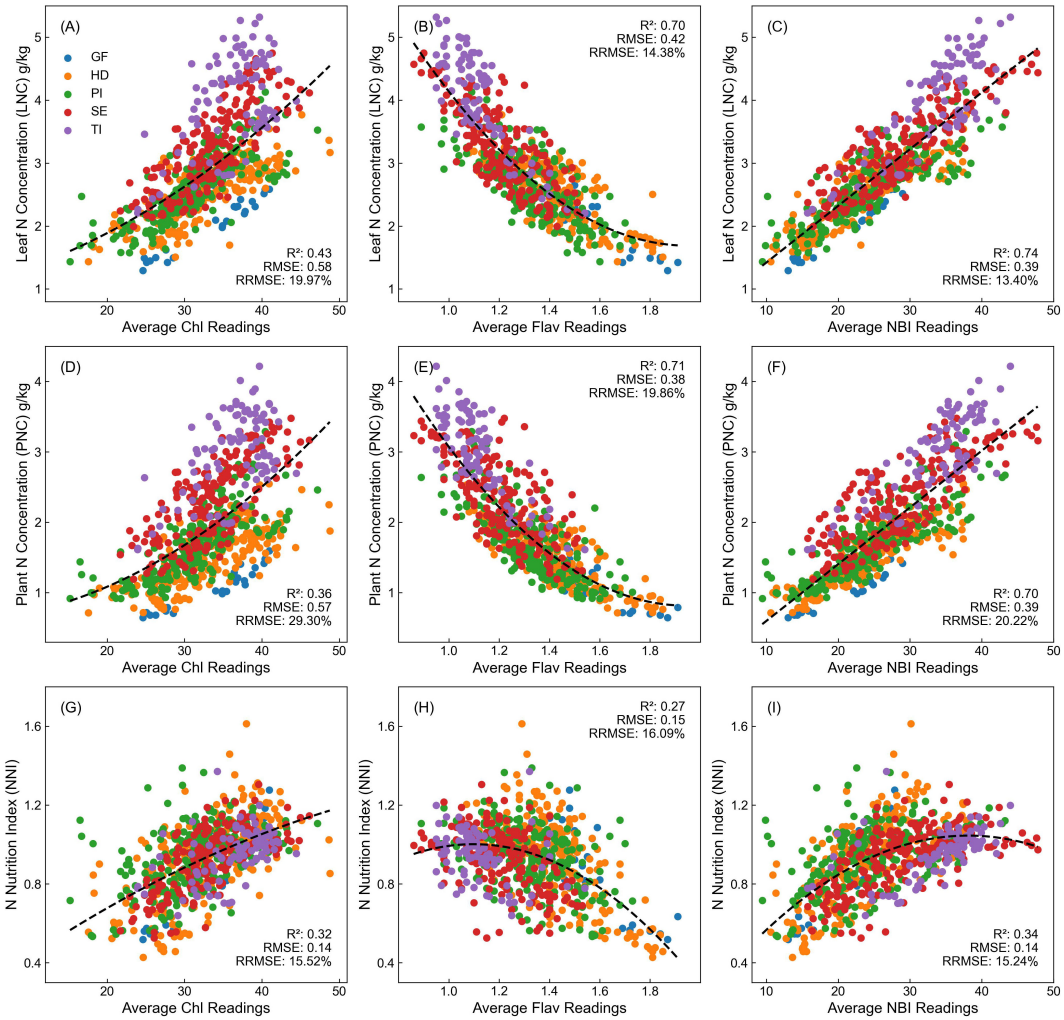


FIGURE 1
Relationship between Dualex values and rice nitrogen indicators across different varieties and growth stages. **(A–C)** show Chl, Flav, and NBI vs. LNC; **(D–F)** show Chl, Flav, and NBI vs. PNC; **(G–I)** show Chl, Flav, and NBI vs. NNI.

Conversely, Flav showed a stable and linear relationship with LNC and PNC, without notable saturation effects (Figures 1B, E). Flav levels increased steadily as nitrogen content decreased, demonstrating a strong negative correlation with LNC and PNC across the full range, with high regression fits (R^2 = 0.70 and 0.71). Even when accounting for differences across rice varieties and growth stages, Flav consistently provided an accurate assessment of nitrogen status.

NBI, defined as the ratio of Chl to Flav, exhibited a significant linear relationship with LNC and PNC (Figures 1C, F), achieving R^2 values of 0.74 and 0.70, respectively. This underscores NBI's robustness and stability in estimating nitrogen nutritional status. As a composite indicator, NBI effectively mitigates the saturation effect observed when using Chl alone, making it a highly reliable tool for nitrogen diagnosis.

For NNI estimation, which represents a relative value (the ratio of PNC to the N_c), measurements across different growth stages showed consistent NNI values, primarily ranging between 0.4 and 1.4. However, Dualex readings varied across stages, leading to lower estimation accuracy for NNI (Figures 1G–I). Although NNI tended to increase with changes in Chl, Flav, and NBI, this trend exhibited significant variability across growth stages, resulting in inconsistent patterns.

3.2 Saturation effect analysis of Dualex values in estimating nitrogen indicators

The saturation effect analysis of Dualex values in estimating rice nitrogen nutritional indicators is illustrated in Figure 2. To minimize the impact of model fitting differences when evaluating saturation effects across various Dualex indices, we applied an incremental analysis method using raw and standardized data. The standardized rate of change (ΔY/ΔX) was calculated based on

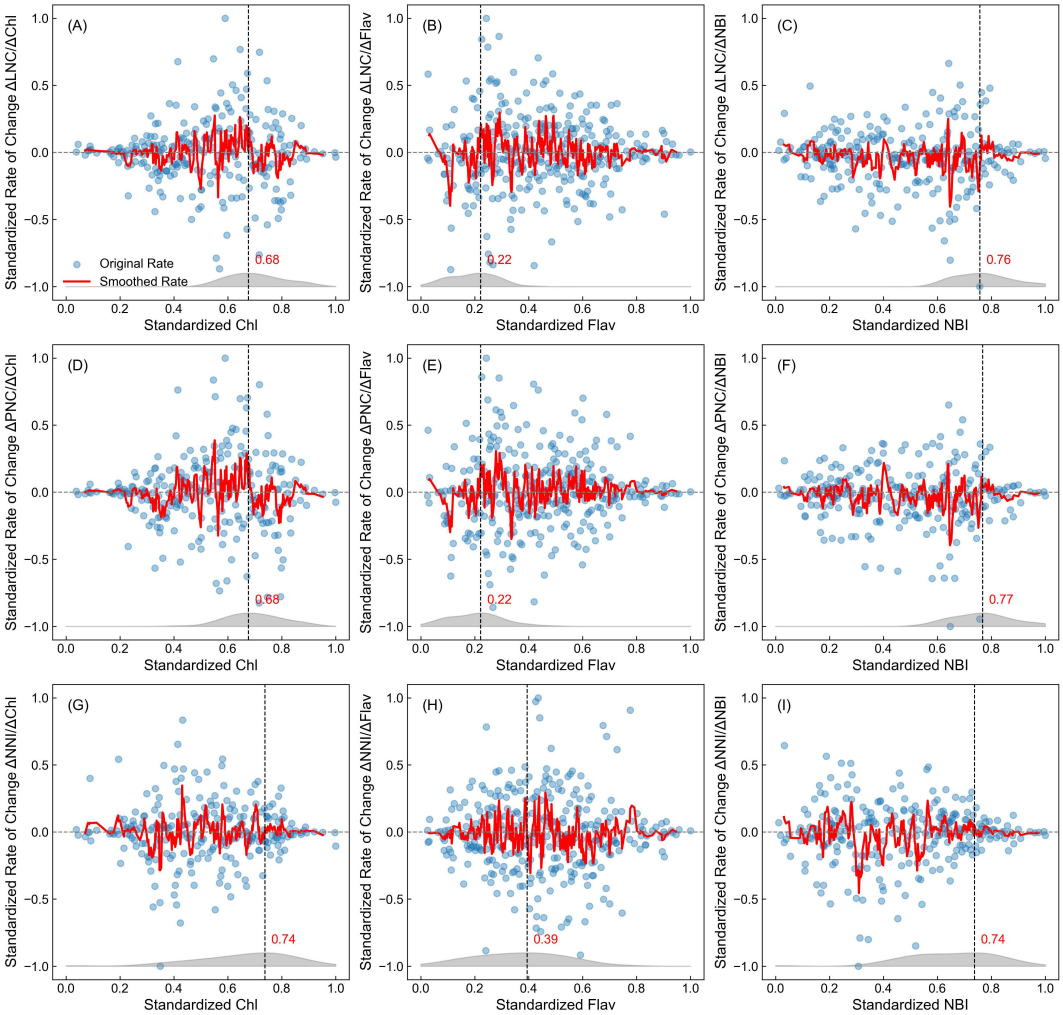


FIGURE 2
Standardized rate of change relative to Dualex values. This figure illustrates the standardized rate of change (ΔY/ΔX) of LNC **(A–C)**, PNC **(D–F)**, and NNI **(G–I)** in relation to standardized Dualex values: Chl, Flav, and NBI. In each subplot, blue scatter points represent the original calculated standardized rate of change data, while the red curve shows the smoothed trend line obtained using a 5-point moving average, reflecting the overall trend. The gray curve at the bottom indicates the density distribution of measurement values corresponding to the highest 20% of nitrogen indices (Y values). The vertical black dashed line marks the point of highest density, annotated with the corresponding X-axis value.

10.3389/fpls.2024.1518272

these original data points. Each subplot displays the rate of change of nitrogen nutritional indicators relative to standardized Dualex values. In the analysis of Chl versus nitrogen indicators (Figures 2A, D, G), an initial low rate of change was observed in the lower Chl range, followed by a gradual increase, peaking at moderate Chl values. However, as Chl values continued to rise, the rate of change began to decline, particularly in the high Chl region. In the low Chl region, the low nitrogen indicator values were primarily responsible for the reduced rate, while in the high Chl region, a declining rate suggested that further increases in Chl no longer led to corresponding rises in nitrogen indicators, indicating a clear saturation effect. The density curves at the bottom further validated this: standardized Chl values peaked around 0.68 (for LNC and PNC) and 0.74 (for NNI), showing a marked decline in estimation accuracy at these levels.

In contrast, Flav exhibited a negative correlation with nitrogen indicators (Figures 1B, E, H). The rate of change remained relatively stable in the low Flav region (Figures 2B, E, H), indicating effectiveness in estimating nitrogen indicators. The density peaks for Flav values were observed at 0.22 (for LNC and PNC) and 0.39 (for NNI), with a gradual distribution, suggesting that even at high nitrogen indicator levels, Flav measurements were evenly spread without a significant saturation effect. However, for NNI, as Flav values decreased, a rapid decline in the rate of change was noted, with the density peak at 0.39 indicating a more pronounced saturation. Comparing this to Figures 1H, NNI exhibited almost horizontal distribution in the low Flav range, indicating diminished sensitivity to changes in Flav.

As a composite indicator, NBI (Chl to Flav ratio) displayed a more consistent trend in estimating LNC, PNC, and NNI (Figures 2C, F, I). The rate of change initially increased, then decreased with rising NBI, resembling the pattern seen in Chl but with density peaks closer to boundary values (0.76, 0.77, and 0.74, respectively). This suggests that NBI effectively estimates nitrogen nutritional status even at high nitrogen levels, exhibiting significantly less saturation than Chl alone.

In summary, Chl demonstrated a clear saturation effect, particularly at high nitrogen levels, where its sensitivity diminished. Conversely, Flav and NBI retained considerable variability in their rates of change at high nitrogen levels, showing enhanced sensitivity to changes in nitrogen status and providing a more accurate reflection of variations in rice nitrogen nutrition.

3.3 Model performance comparison for LNC, PNC, and NNI estimation

Figures 3–5 illustrate the performance of five different models in estimating rice nitrogen nutritional indicators using Dualex readings. Each model utilized measurements from the 1-5 LFT, including Chl, Flav, and NBI values for individual leaves and their averages, totaling 18 input variables. The models showed distinct differences in their performance across training and validation sets, particularly when estimating different nitrogen indicators. Overall, estimation accuracy was highest for LNC, followed by PNC, with NNI being the most challenging to estimate accurately.

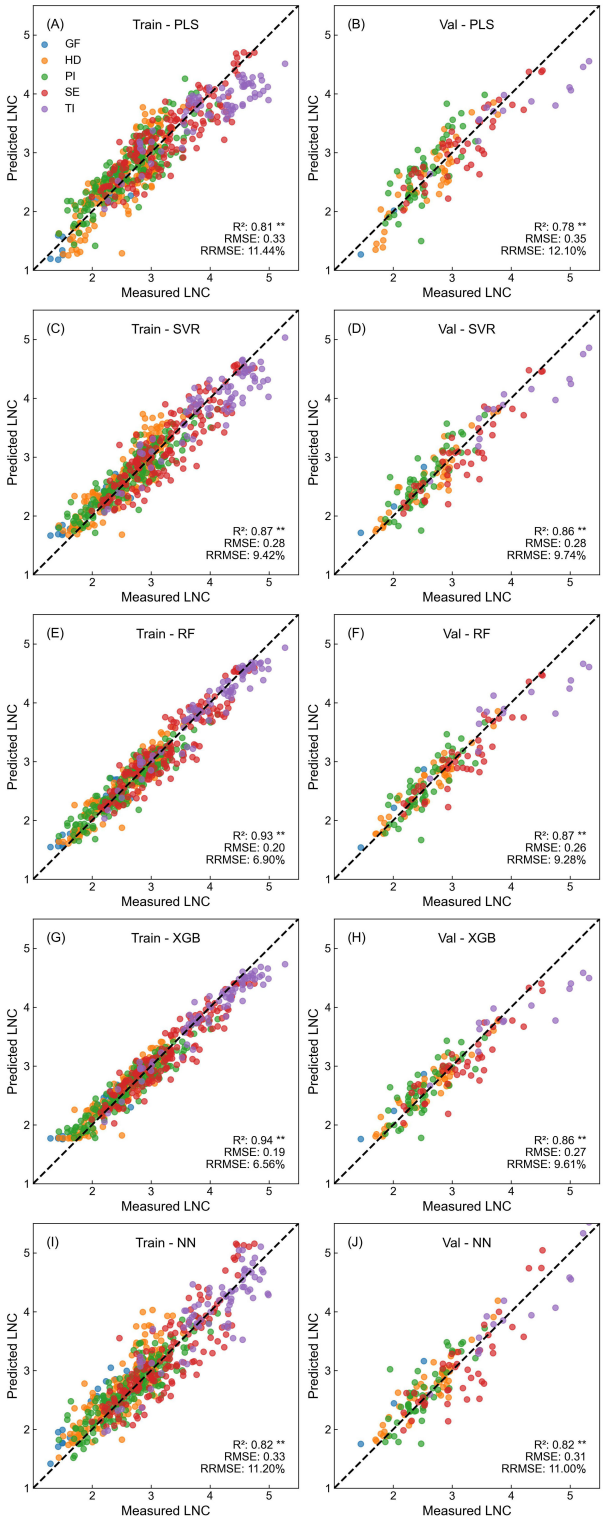


FIGURE 3
Performance of different machine learning models in predicting LNC based on Dualex readings. **(A, C, E, G, I)** show the training performance, and **(B, D, F, H, J)** show the validation performance for PLS, SVR, RF, XGB, and NN models, respectively.

For LNC estimation (Figure 3), all models demonstrated strong fitting accuracy, achieving R^2 values above 0.81 in training and 0.78–0.87 in validation datasets. SVR, RF, and XGB models

10.3389/fpls.2024.1518272

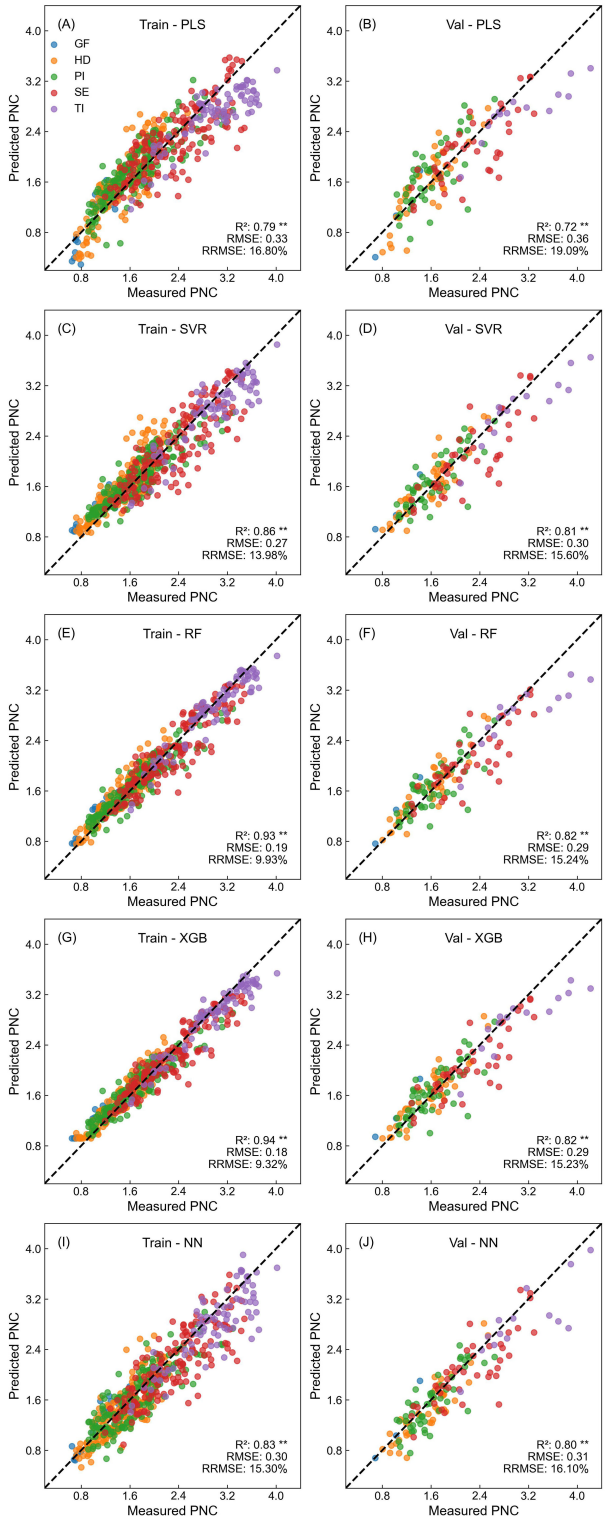


FIGURE 4
Performance of different machine learning models in predicting PNC based on Dualex readings. **(A, C, E, G, I)** show the training performance, and **(B, D, F, H, J)** show the validation performance for PLS, SVR, RF, XGB, and NN models, respectively.

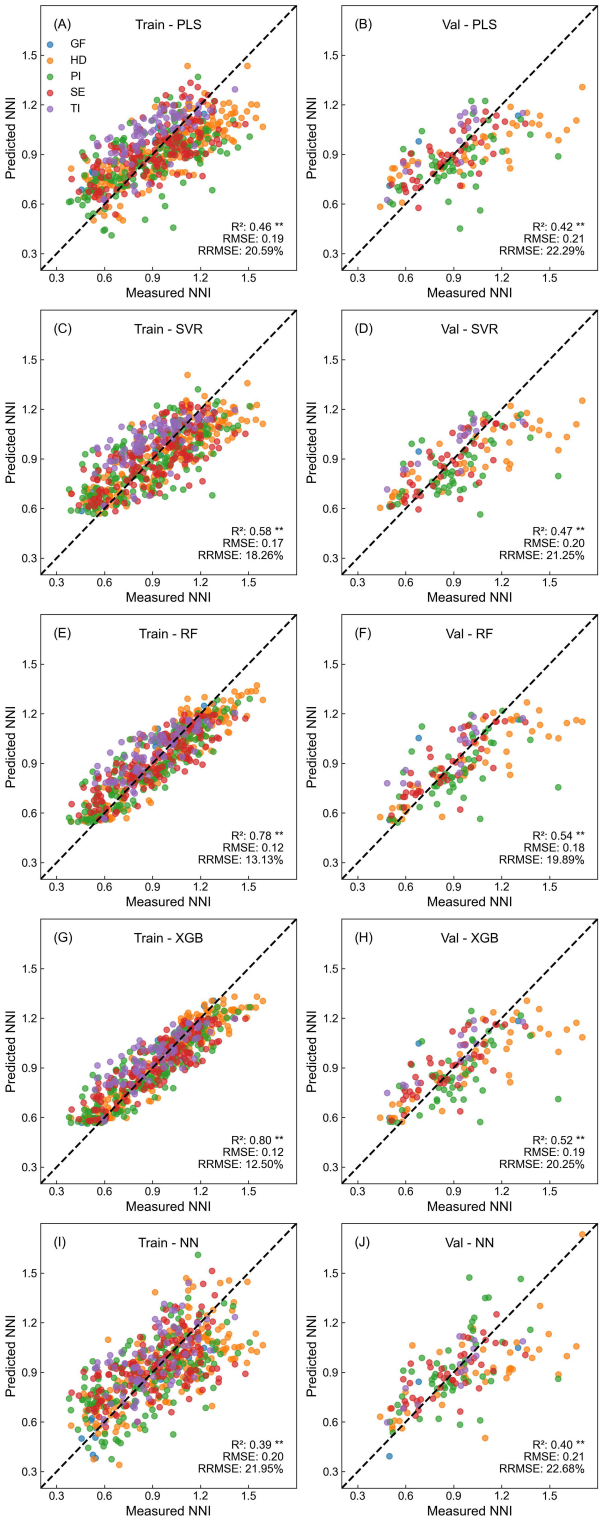


FIGURE 5
Performance of different machine learning models in predicting NNI based on Dualex readings. **(A, C, E, G, I)** show the training performance, and **(B, D, F, H, J)** show the validation performance for PLS, SVR, RF, XGB, and NN models, respectively.

demonstrated the highest validation accuracy, with R^2 values between 0.86 and 0.87, and RRMSE below 10%. Data from all sampling periods, except for minor overestimations at the tillering stage, were evenly distributed around the 1:1 line.

In estimating PNC (Figure 4), model performance was similar to that observed for LNC, despite with slightly reduced accuracy. RF and XGB remained the top performers, with validation R^2 values of 0.82 and RMSE at 0.29. The SVR model followed closely with an R^2

of 0.81. Conversely, PLS regression model demonstrated slightly lower performance.

For NNI estimation (Figure 5), all models showed a noticeable decline in fitting accuracy, which was consistent with regression analysis results from quadratic models. Despite the challenges, RF and XGB continued to lead, with validation R^2 values ranging from 0.52 to 0.54, marking a 71% improvement over QCR models. However, estimating NNI proved to be more difficult than LNC and PNC, with apparent biases in both low and high-value regions.

Radar charts were used to visualize the estimation accuracy of different models and input variable combinations on the validation dataset (Figure 6). The R^2 values decreased progressively from LNC to PNC and then to NNI (Figures 6A-C). While LNC and PNC estimations showed near-equal accuracy, R^2 values for NNI dropped substantially. Across all input variable combinations, R^2 for NNI estimation declined by an average of 57% and 55% compared to LNC and PNC, respectively.

In comparing input variable combinations, the PLS model using multi-leaf NBI value was most effective in estimating LNC and PNC. For the other models, the all-variable combination (Combination 4) and NBI combination consistently provided high accuracy. Changes in RRMSE mirrored R^2 trends inversely (Figures 6D-F); models with higher R^2 exhibited lower RRMSE. RRMSE increased sequentially when estimating LNC, PNC, and NNI. Among the input variable combinations, models using only the Chl combination had the highest RRMSE for LNC and PNC estimations, indicating lower accuracy in these cases. In contrast, for NNI estimation, models using only the Flav combination had the highest RRMSE (Figure 6F). The NBI and all-variable combinations consistently achieved similar, superior accuracy across all nitrogen indicators.

3.4 SHAP analysis of feature importance in nitrogen estimation

The feature importance analysis results based on SHAP values for the RF and XGB models in nitrogen estimation are presented in Figure 7. Both models used Combination 4 as input to assess the impact of each variable on their predictions.

For LNC prediction, Figures 7A, D show the SHAP analysis for the RF and XGB models, respectively. In the RF model (Figure 7A), the NBI value from the 2 LFT was the most influential variable, followed by the Flav value from the 1 LFT and the NBI value from the 1 LFT. The XGB model (Figure 7D) also highlighted these three variables as having the highest impact, with their SHAP values significantly surpassing those of other variables. This suggests that these variables were critical for accurate LNC prediction.

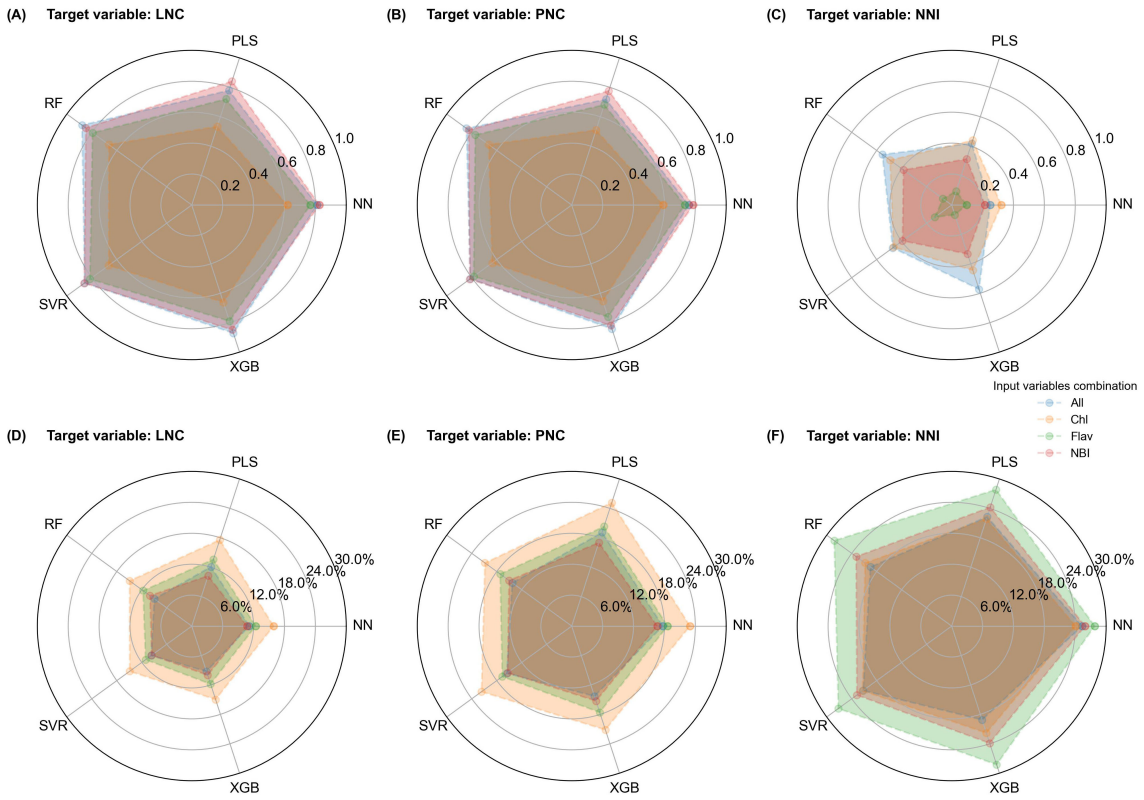


FIGURE 6
Comparison of machine learning model performance using different input variables for LNC, PNC, and NNI estimation. Each radar chart compares five models, with axes representing different input variables: All, Chl, Flav, and NBI. Charts **(A–C)** display R^2 values, and charts **(D–F)** show RRMSE for validation results.

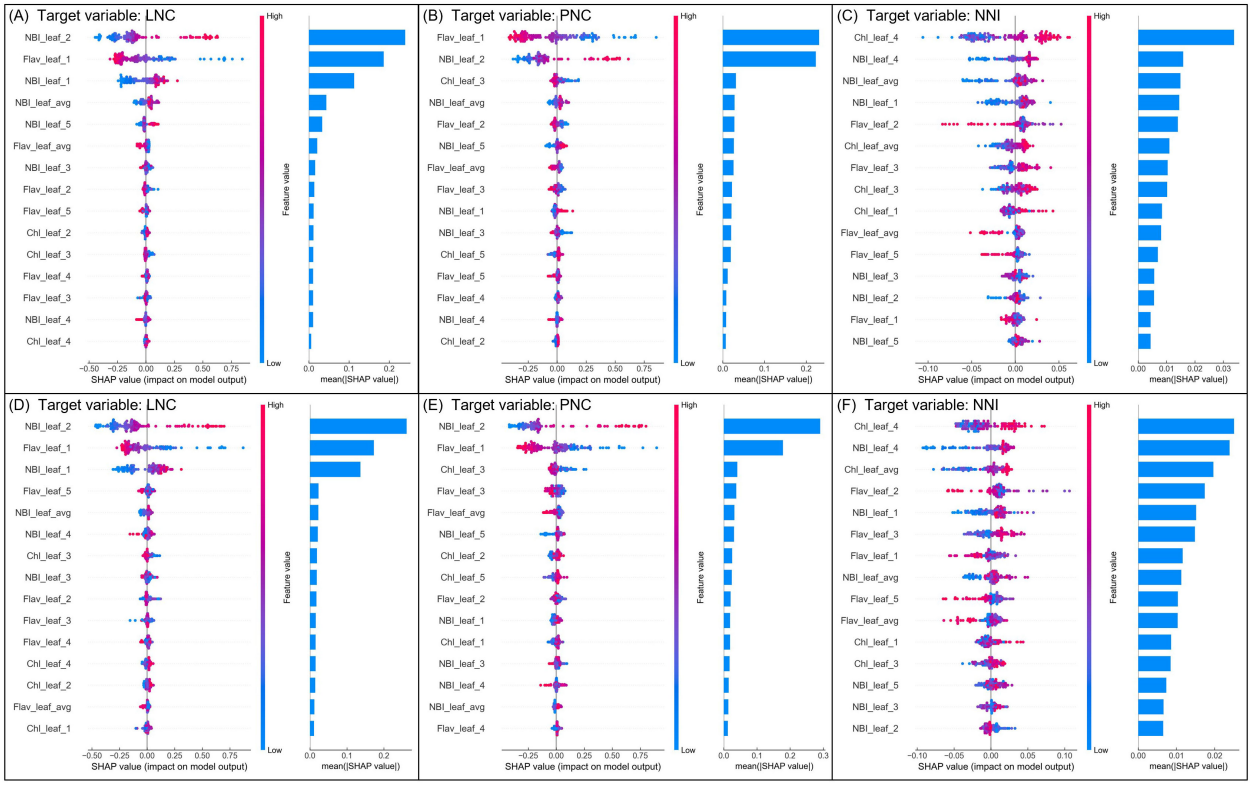


FIGURE 7
SHAP value analysis of feature importance for RF **(A–C)** and XGB **(D–F)** models in predicting rice nitrogen indicators. Each plot presents the SHAP values for the 15 most important features. In the SHAP scatter plots on the left, red indicates higher feature values and blue indicates lower feature values. Positive SHAP values indicate that the feature increases the predicted nitrogen indicator, while negative SHAP values indicate that the feature decreases it. The right side ranks feature importance based on the mean absolute SHAP value.

In the case of PNC estimation, Figures 7B, E illustrate the SHAP analysis for the RF and XGB models. The RF model (Figure 7B) indicated that the Flav value from the 1 LFT and the NBI value from the 2 LFT were the most significant variables. Similarly, in the XGB model (Figure 7E), the NBI of the 2 LFT and Flav of the 1 LFT were also identified as the most influential factors. This consistency across models underscores the importance of NBI and Flav as key indicators for both LNC and PNC estimation.

For NNI prediction, Figures 7C, F display the SHAP analysis results for the RF and XGB models. The RF model (Figure 7C) identified the Chl value from the 4 LFT as the most important variable, followed by the NBI value from the 4 LFT. The XGB model (Figure 7F) similarly showed these two variables as dominant; however, there were differences in their SHAP value distributions. In the RF model, the Chl value from the 4 LFT had a SHAP value notably higher than the other variables, whereas, in the XGB model, SHAP values of variables decreased in a stepwise manner, indicating that no single variable was overwhelmingly dominant.

Overall, the SHAP analysis results for LNC and PNC predictions revealed certain dominant variables that enabled the models to achieve high-precision estimates. However, for NNI prediction, the lack of clearly dominant variables and weaker associations between input variables and the target presented challenges in achieving satisfactory model accuracy.

4 Discussion

4.1 Saturation effect and its implications

Incremental analysis highlighted distinct differences in the saturation effects of different Dualex indices when estimating rice nitrogen indicators. The results showed that Chl experiences significant saturation at high measurement values, consistent with findings from studies using SPAD-502 (Yuan et al., 2016; Zhang et al., 2019a; Jiang et al., 2021) and chlorophyll-based multispectral/hyperspectral indices (Fu et al., 2021; Yao et al., 2023; Gao et al., 2024). Previous research has noted that SPAD values become less responsive to nitrogen at higher concentrations, complicating the differentiation of nitrogen levels (Yue et al., 2020). This saturation effect is documented across various crops, including rice (Wang et al., 2023), wheat (Jiang et al., 2021), and maize (Dong et al., 2021). Our findings reaffirm that Chl and SPAD measurements exhibit saturation under high nitrogen conditions.

When leaf chlorophyll content is low, an increase can enhance the leaf's net photosynthetic rate (Tantray et al., 2020). However, as nitrogen availability rises and chlorophyll reaches a threshold level, photosynthetic rates plateau due to limits in light absorption area (Hou et al., 2019). At this point, excess nitrogen is allocated to non-photosynthetic reserves, altering the ratio of chlorophyll to leaf nitrogen (Hou et al., 2019; Mu and Chen, 2021).

Consequently, while Chl is a sensitive indicator at low nitrogen levels, its effectiveness decreases at higher levels, introducing errors into predictive models. This trend was observed in our analysis, particularly in LNC predictions, where models relying solely on Chl values achieved lower R^2 scores, failing to capture nitrogen variations accurately. Therefore, Chl's utility as a single diagnostic variable is limited, especially under high nitrogen conditions.

The saturation effect in this study was mainly observed when rice plants exhibited high nitrogen content. Since nitrogen concentration in rice leaves or plants gradually decreases during growth—a pattern known as the nitrogen dilution effect (Yao et al., 2021; Li et al., 2022a)—with the impact most noticeable around the tillering and jointing stages. This differs from canopy-level saturation observed after canopy closure, where high leaf area can cause remote sensing indices (e.g., NDVI) to saturate due to optical signal limitations, reducing their sensitivity to nitrogen changes (Tenreiro et al., 2021). Conversely, Chl saturation is primarily driven by physiological limits on chlorophyll content within leaves, necessitating a distinction between these two types of saturation in research.

In contrast, Flav and NBI showed no significant saturation effects, maintaining their effectiveness in estimating nitrogen indicators, even at high nitrogen levels. Flav is closely associated with plant stress responses, increasing under conditions of nitrogen stress (Li et al., 2021). When nitrogen is scarce, plants produce more flavonoids as a protective response (Gitelson et al., 2017). NBI combines the divergent responses of these indicators. In nitrogen-rich conditions, Chl levels are high while Flav remains low, leading to a high NBI. Under nitrogen deficiency, Chl decreases and Flav increases, resulting in a low NBI. By integrating complementary data from both indicators, NBI provides a robust reflection of nitrogen status, offering distinct advantages over Chl alone. This aligns with studies showing that NBI can overcome saturation and retain sensitivity under high nitrogen levels (Dong et al., 2022; Song et al., 2022). Our dataset, encompassing 15 rice varieties across different growth stages, confirmed that NBI and Flav maintained high predictive accuracy across various environmental conditions, suggesting these indicators can reliably reflect nitrogen status across diverse environmental conditions.

4.2 Comparison of machine learning models

We compared the performance of five machine learning models —PLS, SVR, RF, XGB, and NN—in predicting nitrogen indicators. Overall, XGB and RF models showed the highest accuracy for LNC, PNC, and NNI predictions, whereas PLS regression and SVR models had noticeably lower performance, lagging behind the tree-based approaches. This trend is consistent with findings from other studies comparing similar models (Fan et al., 2022; Kganyago et al., 2024). While PLS regression is effective at explaining linear relationships, it struggles with complex nonlinear patterns (Yang et al., 2023). SVR, though capable of managing nonlinear data, is highly sensitive to parameter selection and less efficient with high-dimensional inputs (Ben Ishak, 2016). Conversely, tree-based ensemble models like RF and XGB excel at handling nonlinear, high-dimensional datasets (Nzekwe et al., 2024), with XGB particularly excelling due to its gradient boosting optimization, which iteratively improves model performance.

The NN model, however, was less stable than the XGB and RF models in this study. This may be due to the relatively smaller dataset size, as neural networks generally require larger datasets to train effectively and achieve good generalization. While the NN performed adequately in estimating LNC and PNC, its predictions for NNI were less accurate than those of the other models. The RF and XGB models demonstrated strong performance, although the training R^2 values were slightly higher than the validation R^2 values, especially for NNI estimation, the differences were minimal. This highlights the importance of model selection and appropriate complexity in achieving reliable predictions (Moradi et al., 2020; Yates et al., 2023).

The lower estimation accuracy for NNI compared to LNC and PNC can be attributed to several factors. NNI is a composite index involving both PNC and N_c, where N_c is dependent on plant biomass and derived from the nitrogen dilution curve. This adds layers of complexity and potential sources of error. The dynamic nature of biomass accumulation and nitrogen dilution across growth stages introduces variability that is difficult to capture with current models. Additionally, the relatively narrow range of NNI values reduces the models' ability to distinguish subtle differences. Measurement errors in biomass and nitrogen content further compound the challenge. Future research could explore incorporating growth stage information, more precise biomass measurements, and additional physiological parameters to enhance NNI prediction accuracy.

4.3 Feature importance across different input variables and leaf positions

SHAP analysis was instrumental in interpreting the models by quantifying the contribution of each input variable to the predictions (Lundberg et al., 2017). This is essential for enhancing model transparency, which is particularly valuable in agriculture, where interpretable models facilitate practical application (Ryo, 2022). The SHAP results highlighted that spectral data from different leaf positions do not contribute equally to nitrogen prediction. For LNC and PNC, the most significant contributions came from measurements taken at the 1 LFT and 2 LFT, corroborating findings from related studies where spectral reflectance from top leaves was found to be most sensitive for nitrogen assessment (Dong et al., 2021; Lu et al., 2022; Wang et al., 2023). For NNI prediction, SHAP analysis pointed to the importance of Chl measurements from the 4 LFT, suggesting that chlorophyll content in middle to lower leaves is particularly relevant. This may stem from the generally lower estimation accuracy of NNI, which tends to make the significance of variables more uniform. Additionally, nitrogen translocation in middle and lower leaves closely aligns with the plant's overall nitrogen nutritional status (Sun et al., 2018; Li et al., 2022b).

To improve NNI prediction accuracy, future strategies could include incorporating additional variables closely related to NNI, such as plant leaf area index (Tuccio et al., 2022), growth stages, and other physiological and environmental factors (Dong et al., 2022, 2024). Moreover, combining these models with crop growth simulations or radiative transfer models, as well as using multi-source data, could enhance the model's ability to account for complex interactions (Raya-Sereno et al., 2024; Wang et al., 2024). Expanding the dataset's size and diversity will also be crucial for improving generalization and predictive accuracy.

The differences in contributions from various leaf positions suggest the potential for optimizing measurement strategies in practical applications. By focusing on key leaf positions that provide the most significant input to the models, measurement efficiency can be improved. Reducing the number of measurement variables not only lowers costs but also enhances the computational efficiency of the models, making them more viable for real-world use.

5 Conclusion

This study systematically assessed the effectiveness of Dualex measurements for estimating rice nitrogen nutritional indicators. The results indicated that Chl experiences significant saturation at high nitrogen levels, limiting its utility as a nitrogen diagnostic indicator. In contrast, Flav and NBI did not exhibit saturation effects and remained highly sensitive under high nitrogen conditions, accurately reflecting rice nitrogen status. Among the machine learning models tested, RF and XGB were the most effective in predicting LNC and PNC. SHAP analysis identified the most influential feature variables and leaf positions, further enhancing model interpretability and accuracy.

Future research should focus on expanding the dataset to include more diverse rice varieties and growth stages and incorporating additional crop and environmental variables. Additionally, exploring the applicability of Flav and NBI across different crops and environmental contexts could provide broader technical support for nitrogen management in precision agriculture.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

PS: Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. YW: Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. CY: Investigation, Methodology, Resources, Writing – review & editing. KF: Investigation, Methodology, Resources, Writing – review & editing. YQ: Conceptualization, Methodology, Project administration, Resources, Supervision, Writing – review & editing. GC: Conceptualization, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Project of Jiangsu Vocational College of Agriculture and Forestry, China (No.2021kj15, 2022kj63), the National Natural Science Foundation of China (No. 32072681, 32101616, 32360530), Zhejiang Provincial Natural Science Foundation of China (No. LTGN23D010001), and the Frontier Project from the Institute of Soil Science, Chinese Academy of Sciences (ISSASIP2406).

Acknowledgments

We would like to express our sincere gratitude to all the team members and field staff who contributed significantly to the management of the field experiments and sample collection. Their collective efforts were important in the successful completion of this research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that Generative AI was used in the creation of this manuscript. GENERATIVE AI was used for the creation of this Manuscript. This article was edited and refined using generative AI.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ben Ishak, A. (2016). Variable selection using support vector regression and random forests: A comparative study. *Intell. Data Anal.* 20, 83–104. doi: 10.3233/IDA-150795

Berger, K., Verrelst, J., Féret, J.-B., Wang, Z., Wocher, M., Strathmann, M., et al. (2020). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sens. Environ.* 242, 111758. doi: 10.1016/j.rse.2020.111758

Cai, S., Zhao, X., Pittelkow, C. M., Fan, M., Zhang, X., and Yan, X. (2023). Optimal nitrogen rate strategy for sustainable rice production in China. *Nature* 615, 73–79. doi: 10.1038/s41586-022-05678-x

Cerovic, Z. G., Ghozlen, N. B., Milhade, C., Obert, M., Debuisson, S., and Le Moigne, M. (2015). Nondestructive diagnostic test for nitrogen nutrition of grapevine (Vitis vinifera L.) based on dualex leaf-clip measurements in the field. *J. Agric. Food Chem.* 63, 3669–3680. doi: 10.1021/acs.jafc.5b00304

Cerovic, Z. G., Masdoumier, G., Ghozlen, N. B., and Latouche, G. (2012). A new optical leaf-clip meter for simultaneous non-destructive assessment of leaf chlorophyll and epidermal flavonoids. *Physiol. Plant* 146, 251–260. doi: 10.1111/j.1399-3054.2012.01639.x

Chang, G. J., Oh, Y., Goldshleger, N., and Shoshany, M. (2022). Biomass estimation of crops and natural shrubs by combining red-edge ratio with normalized difference vegetation index. *J. Appl. Remote Sens.* 16, 14501. doi: 10.1117/1.JRS.16.014501

Chang, T., Zhao, H., Wang, N., Song, Q., Xiao, Y., Qu, M., et al. (2019). A three-dimensional canopy photosynthesis model in rice with a complete description of the canopy architecture, leaf physiology, and mechanical properties. *J. Exp. Bot.* 70, 2479–2490. doi: 10.1093/jxb/ery430

Dong, R., Miao, Y., Wang, X., Chen, Z., and Yuan, F. (2021). Improving maize nitrogen nutrition index prediction using leaf fluorescence sensor combined with environmental and management variables. *Field Crops Res.* 269, 108180. doi: 10.1016/j.fcr.2021.108180

Dong, L., Miao, Y., Wang, X., Kusnierek, K., Zha, H., Pan, M., et al. (2024). In-season dynamic diagnosis of maize nitrogen status across the growing season by integrating proximal sensing and crop growth modeling. *Comput. Electron. Agric.* 224, 109240. doi: 10.1016/j.compag.2024.109240

Dong, R., Miao, Y., Wang, X., Yuan, F., and Kusnierek, K. (2022). Combining leaf fluorescence and active canopy reflectance sensing technologies to diagnose maize nitrogen status across growth stages. *Precis. Agric.* 23, 939–960. doi: 10.1007/s11119-021-09869-w

Dong, T., Shang, J., Chen, J. M., Liu, J., Qian, B., Ma, B., et al. (2019). Assessment of portable chlorophyll meters for measuring crop leaf chlorophyll concentration. *Remote Sens.* 11, 2706. doi: 10.3390/rs11222706

Fan, K., Li, F., Chen, X., Li, Z., and Mulla, D. J. (2022). Nitrogen balance index prediction of winter wheat by canopy hyperspectral transformation and machine learning. *Remote Sens.* 14, 3504. doi: 10.3390/rs14143504

Fu, Y., Yang, G., Pu, R., Li, Z., Li, H., Xu, X., et al. (2021). An overview of crop nitrogen status assessment using hyperspectral remote sensing: Current status and perspectives. *Eur. J. Agron.* 124, 126241. doi: 10.1016/j.eja.2021.126241

Gao, S., Yan, K., Liu, J., Pu, J., Zou, D., Qi, J., et al. (2024). Assessment of remote-sensed vegetation indices for estimating forest chlorophyll concentration. *Ecol. Indic.* 162, 112001. doi: 10.1016/j.ecolind.2024.112001

Gitelson, A., Chivkunova, O., Zhigalova, T., and Solovchenko, A. (2017). *In situ* optical properties of foliar flavonoids: Implication for non-destructive estimation of flavonoid content. *J. Plant Physiol.* 218, 258–264. doi: 10.1016/j.jplph.2017.08.009

Guo, Y., Chen, Y., Searchinger, T. D., Zhou, M., Pan, D., Yang, J., et al. (2020). Air quality, nitrogen use efficiency and food security in China are improved by cost-effective agricultural nitrogen management. *Nat. Food* 1, 648–658. doi: 10.1038/s43016-020-00162-z

Hamdane, Y., Segarra, J., Buchaillot, M. L., Rezzouk, F. Z., Gracia-Romero, A., Vatter, T., et al. (2023). Using ground and UAV vegetation indexes for the selection of fungal-resistant bread wheat varieties. *Drones* 7, 454. doi: 10.3390/drones7070454

Hou, W., Tränkner, M., Lu, J., Yan, J., Huang, S., Ren, T., et al. (2019). Interactive effects of nitrogen and potassium on photosynthesis and photosynthetic nitrogen allocation of rice leaves. *BMC Plant Biol.* 19, 302. doi: 10.1186/s12870-019-1894-8

Huang, C., Zhang, Z., Lu, Z., Zhang, X., Zhu, L., and Yang, W. (2023). Leaf grading for cotton verticillium wilt based on VFNet-Improved and Deep Sort. *J. Intell. Agric. Mech.* 4, 12–21. doi: 10.12398/j.issn.2096-7217.2023.02.002

Jiang, J., Wang, C., Wang, H., Fu, Z., Cao, Q., Tian, Y., et al. (2021). Evaluation of three portable optical sensors for non-destructive diagnosis of nitrogen status in winter wheat. *Sensors* 21, 5579. doi: 10.3390/s21165579

Justes, E., Mary, B., Meynard, J. M., Machet, J. M., and Thelier-Huche, L. (1994). Determination of a critical nitrogen dilution curve for winter wheat crops. *Ann. Bot.* 74, 397–407. doi: 10.1006/anbo.1994.1133

Karaca, C., Thompson, R. B., Peña-Fleitas, M. T., Gallardo, M., and Padilla, F. M. (2023). Evaluation of absolute measurements and normalized indices of proximal optical sensors as estimators of yield in muskmelon and sweet pepper. *Remote Sens.* 15, 2174. doi: 10.3390/rs15082174

Kganyago, M., Adjorlolo, C., Mhangara, P., and Tsoeleng, L. (2024). Optical remote sensing of crop biophysical and biochemical parameters: An overview of advances in sensor technologies and machine learning algorithms for precision agriculture. *Comput. Electron. Agric.* 218, 108730. doi: 10.1016/j.compag.2024.108730

Li, X., Ata-UI-Karim, S. T., Li, Y., Yuan, F., Miao, Y., Kato, Y., et al. (2022a). Advances in the estimations and applications of critical nitrogen dilution curve and nitrogen nutrition index of major cereal crops. A review. *Comput. Electron. Agric.* 197, 106998. doi: 10.1016/j.compag.2022.106998

Li, Z., Jiang, H., Qin, Y., Yan, H., Jiang, X., and Qin, Y. (2021). Nitrogen deficiency maintains the yield and improves the antioxidant activity of Coreopsis tinctoria Nutt. *Biosci. Biotechnol. Biochem.* 85, 1492–1505. doi: 10.1093/bbb/zbab048

Li, Y. Y., Ming, B., Fan, P. P., Liu, Y., Wang, K. R., Hou, P., et al. (2022b). Effects of nitrogen application rates on the spatio-temporal variation of leaf SPAD readings on the maize canopy. *J. Agric. Sci.* 160, 32–44. doi: 10.1017/S0021859621001052

Li, S., Xu, C., Wang, N., Cao, H., and Yu, F. (2024). Research on rice nitrogen unmanned aerial vehicle hyperspectral inversion based on BWO-ELM. *J. Intell. Agric. Mech.* 5, 14–21. doi: 10.12398/j.issn.2096-7217.2024.03.002

Liu, Y., Wang, H., Jiang, Z., Wang, W., Xu, R., Wang, Q., et al. (2021). Genomic basis of geographical adaptation to soil nitrogen in rice. *Nature* 590, 600–605. doi: 10.1038/s41586-020-03091-w

Lu, J., Lu, J., Nie, W., Song, J., Zhan, Q., Wang, M., et al. (2022). Estimation of nitrogen nutrition index in chrysanthemum using chlorophyll meter readings. *Ind. Crops Prod.* 187, 115459. doi: 10.1016/j.indcrop.2022.115459

Lundberg, S. M., and Lee, S. I. (2017). "A unified approach to interpreting model predictions," in *Advances in neural information processing systems*. Eds. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et al (Curran Associates, Inc, Long Beach, CA, USA). Available at: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf (accessed September 20, 2024).

Luo, X., Keenan, T. F., Chen, J. M., Croft, H., Colin Prentice, I., Smith, N. G., et al. (2021). Global variation in the fraction of leaf nitrogen allocated to photosynthesis. *Nat. Commun.* 12, 4866. doi: 10.1038/s41467-021-25163-9

Ma, X., Huete, A., Tran, N. N., Bi, J., Gao, S., and Zeng, Y. (2020). Sun-angle effects on remote-sensing phenology observed and modelled using Himawari-8. *Remote Sens.* 12, 1339. doi: 10.3390/rs12081339

Mandal, D., Siqueira, R., de, Longchamps, L., and Khosla, R. (2024). Machine learning and fluorosensing for estimation of maize nitrogen status at early growth-stages. *Comput. Electron. Agric.* 225, 109341. doi: 10.1016/j.compag.2024.109341

Moradi, R., Berangi, R., and Minaei, B. (2020). A survey of regularization strategies for deep models. *Artif. Intell. Rev.* 53, 3947–3986. doi: 10.1007/s10462-019-09784-7

Mu, X., and Chen, Y. (2021). The physiological response of photosynthesis to nitrogen deficiency. *Plant Physiol. Biochem.* 158, 76–82. doi: 10.1016/j.plaphy.2020.11.019

Nzekwe, C. J., Kim, S., and Mostafa, S. A. (2024). Interaction selection and prediction performance in high-dimensional data: A comparative study of statistical and tree-based methods. *J. Data Sci.* 22, 259–279. doi: 10.6339/24-JDS1127

Ouyang, W., Yin, X., Yang, J., and Struik, P. C. (2021). Roles of canopy architecture and nitrogen distribution in the better performance of an aerobic than a lowland rice cultivar under water deficit. *Field Crops Res.* 271, 108257. doi: 10.1016/j.fcr.2021.108257

Qian, B., Ye, H., Huang, W., Xie, Q., Pan, Y., Xing, N., et al. (2022). A sentinel-2-based triangular vegetation index for chlorophyll content estimation. *Agric. For. Meteorol.* 322, 109000. doi: 10.1016/j.agrformet.2022.109000

Raya-Sereno, M. D., Camino, C., Pancorbo, J. L., Alonso-Ayuso, M., Gabriel, J. L., Beck, P. S. A., et al. (2024). Assessing wheat genotype response under combined nitrogen and water stress scenarios coupling high-resolution optical and thermal sensors with radiative transfer models. *Eur. J. Agron.* 154, 127102. doi: 10.1016/j.eja.2024.127102

Röll, G., Hartung, J., and Graeff-Hönninger, S. (2019). Determination of plant nitrogen content in wheat plants via spectral reflectance measurements: Impact of leaf number and leaf position. *Remote Sens.* 11, 2794. doi: 10.3390/rs11232794

Rubo, S., and Zinkernagel, J. (2022). Exploring hyperspectral reflectance indices for the estimation of water and nitrogen status of spinach. *Biosyst. Eng.* 214, 58–71. doi: 10.1016/j.biosystemseng.2021.12.008

Ryo, M. (2022). Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. *Artif. Intell. Agric.* 6, 257–265. doi: 10.1016/j.aiia.2022.11.003

Shi, P., Wang, Y., Xu, J., Zhao, Y., Yang, B., Yuan, Z., et al. (2021). Rice nitrogen nutrition estimation with RGB images and machine learning methods. *Comput. Electron. Agric.* 180, 105860. doi: 10.1016/j.compag.2020.105860

Singh, V., Kunal,, Singh, M., and Singh, B. (2022). Spectral indices measured with proximal sensing using canopy reflectance sensor, chlorophyll meter and leaf color chart for in-season grain yield prediction of basmati rice. *Pedosphere* 32, 812–822. doi: 10.1016/j.pedsph.2022.06.015

Song, X., Yang, G., Xu, X., Zhang, D., Yang, C., and Feng, H. (2022). Winter wheat nitrogen estimation based on ground-level and UAV-mounted sensors. *Sensors* 22, 549. doi: 10.3390/s22020549

Sun, Y., Zhu, S., Yang, X., Weston, M. V., Wang, K., Shen, Z., et al. (2018). Nitrogen diagnosis based on dynamic characteristics of rice leaf image. *PloS One* 13, e0196298. doi: 10.1371/journal.pone.0196298

Tan, L., Zhou, L., Zhao, N., He, Y., and Qiu, Z. (2021). Development of a low-cost portable device for pixel-wise leaf SPAD estimation and blade-level SPAD distribution visualization using color sensing. *Comput. Electron. Agric.* 190, 106487. doi: 10.1016/j.compag.2021.106487

Tantray, A. Y., Bashir, S. S., and Ahmad, A. (2020). Low nitrogen stress regulates chlorophyll fluorescence in coordination with photosynthesis and Rubisco efficiency of rice. *Physiol. Mol. Biol. Plants* 26, 83–94. doi: 10.1007/s12298-019-00721-0

Tenreiro, T. R., García-Vila, M., Gómez, J. A., Jiménez-Berni, J. A., and Fereres, E. (2021). Using NDVI for the assessment of canopy cover in agricultural crops within modelling research. *Comput. Electron. Agric.* 182, 106038. doi: 10.1016/j.compag.2021.106038

Tuccio, L., Massa, D., Cacini, S., Iovieno, P., and Agati, G. (2022). Monitoring nitrogen variability in two Mediterranean ornamental shrubs through proximal fluorescence-based sensors at leaf and canopy level. *Sci. Hortic.* 294, 110773. doi: 10.1016/j.scienta.2021.110773

Wang, Z., Li, H., Xia, Y., Fan, M., Zhao, H., Lu, J., et al. (2023). Relationship between blade SPAD value and protein content of rice in cold region. *Agric. Res. Arid Areas* 41, 150–161. doi: 10.7606/j.issn.1000-7601.2023.05.16

Wang, D., Struik, P. C., Liang, L., and Yin, X. (2024). Developing remote sensing- and crop model-based methods to optimize nitrogen management in rice fields. *Comput. Electron. Agric.* 220, 108899. doi: 10.1016/j.compag.2024.108899

Waqas, M., Wang, L., Jones, J. J., Josef Turetschek, R., Engelmeier, D., Geilfus, C.-M., et al. (2023). Short-term phosphorus deficiency induces flavonoid accumulation in the lamina of pak choi: A finishing treatment that influences inner quality. *Sci. Hortic.* 314, 111953. doi: 10.1016/j.scienta.2023.111953

Wu, B., Zhang, M., Zeng, H., Tian, F., Potgieter, A. B., Qin, X., et al. (2023). Challenges and opportunities in remote sensing-based crop monitoring: a review. *Natl. Sci. Rev.* 10, nwac290. doi: 10.1093/nsr/nwac290

Yang, H., Yin, H., Li, F., Hu, Y., and Yu, K. (2023). Machine learning models fed with optimized spectral indices to advance crop nitrogen monitoring. *Field Crops Res.* 293, 108844. doi: 10.1016/j.fcr.2023.108844

Yao, B., Ata-Ul-Karim, S. T., Li, Y., Ye, T., Zhu, Y., Cao, W., et al. (2023). Plant nitrogen status at phenological stages can well estimate wheat yield and its components. *Field Crops Res.* 297, 108950. doi: 10.1016/j.fcr.2023.108950

Yao, B., Wang, X., Lemaire, G., Makowski, D., Cao, Q., Liu, X., et al. (2021). Uncertainty analysis of critical nitrogen dilution curves for wheat. *Eur. J. Agron.* 128, 126315. doi: 10.1016/j.eja.2021.126315

Yates, L. A., Aandahl, Z., Richards, S. A., and Brook, B. W. (2023). Cross validation for model selection: A review with examples from ecology. *Ecol. Monogr.* 93, e1557. doi: 10.1002/ecm.1557

Yuan, Z., Cao, Q., Zhang, K., Ata-Ul-Karim, S. T., Tian, Y., Zhu, Y., et al. (2016). Optimal leaf positions for SPAD meter measurement in rice. *Front. Plant Sci.* 7, 719. doi: 10.3389/fpls.2016.00719

Yue, X., Hu, Y., Zhang, H., and Schmidhalter, U. (2020). Evaluation of both SPAD reading and SPAD index on estimating the plant nitrogen status of winter wheat. *Int. J. Plant Prod.* 14, 67–75. doi: 10.1007/s42106-019-00068-2

Zhang, K., Liu, X., Ma, Y., Zhang, R., Cao, Q., Zhu, Y., et al. (2019a). A comparative assessment of measures of leaf nitrogen in rice using two leaf-clip meters. *Sensors* 20, 175. doi: 10.3390/s20010175

Zhang, L., Shao, Z., Liu, J., and Cheng, Q. (2019b). Deep learning based retrieval of forest aboveground biomass from combined LiDAR and Landsat 8 data. *Remote Sens.* 11, 1459. doi: 10.3390/rs11121459

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 December 2024
DOI 10.3389/fpls.2024.1519200

Check for updates

OPEN ACCESS

EDITED BY
Liujun Xiao,
Nanjing Agricultural University, China

REVIEWED BY
Yanxi Zhao,
Zhejiang University, China
Guocheng Wang,
Beijing Normal University, China
Ziyin Shang,
Chinese Academy of Agricultural Sciences (CAAS), China

*CORRESPONDENCE
Qingsong Jiang
qingsongjiang219@126.com

RECEIVED 29 October 2024
ACCEPTED 03 December 2024
PUBLISHED 20 December 2024

CITATION
Gao Y, Chang L, Zeng M, Hu Q, Hui J and Jiang Q (2024) Quantitative relationship model between soil profile salinity and soil depth in cotton fields based on data assimilation algorithm: forecasting cotton field yields and profits.
Front. Plant Sci. 15:1519200.
doi: 10.3389/fpls.2024.1519200

COPYRIGHT
© 2024 Gao, Chang, Zeng, Hu, Hui and Jiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Quantitative relationship model between soil profile salinity and soil depth in cotton fields based on data assimilation algorithm: forecasting cotton field yields and profits

Yang Gao[1,2], Lin Chang[1,2], Mei Zeng[1,2], Quanze Hu[1,2], Jiaojiao Hui[1,2] and Qingsong Jiang[1,2]*

[1]College of Information Engineering, Tarim University, Alar, China, [2]Key Laboratory of Tarim Oasis Agriculture (Tarim University), Ministry of Education, Alar, China

Soil salinization seriously affects the efficiency of crops in absorbing soil nutrients, and the cotton production in southern Xinjiang accounts for more than 60% of China's total. Therefore, it is crucial to monitor the dynamic changes in the salinity of the soil profile in cotton fields in southern Xinjiang, understand the status of soil salinization, and implement effective prevention and control measures. The drip-irrigated cotton fields in Alaer Reclamation Area were taken as the research objects. The multivariate linear regression model was used to study the relationship between soil salinity and soil depth in different periods, and the Kalman filter algorithm was used to improve the model accuracy. The results showed that the month with the highest improvement in model accuracy was July, with the model accuracy R^2 increasing by 0.26 before and after calibration; followed by June and October, with the model accuracy R^2 increasing by 0.19 and 0.18 respectively; the lowest improvement was in March, which was only 0.01. After the model was calibrated by the Kalman filter algorithm, the fitting accuracy (R^2) between the predicted value and the actual value was as high as 0.79, and the corresponding RMSE was only 96.17 μS cm^{-1}, and the measured value of soil salinity was consistent with the predicted value. Combined with the predicted conductivity data of each soil layer, the total yield of the study area was predicted to be 5,203-5,551 kg hm^{-2}, and the income was about 4,953-7,441 RMB hm^{-2}. It can be seen that Kalman filtering can improve the prediction accuracy of the model and provide a theoretical basis for studying the mechanism of soil salt migration in drip-irrigated cotton fields at different stages. It is of great significance for evaluating the potential relationship between cotton yield and deep soil salinity and guiding the efficient prevention and control of saline soil in cotton fields.

KEYWORDS

salinization, apparent conductivity, soil conductivity, multivariate linear algorithm, Kalman filter

1 Introduction

Cotton is a crucial pillar industry in Xinjiang, particularly in the southern region, with factors such as climatic conditions, seed selection, cultivation management, soil fertility and soil salinity content significantly influencing cotton yields (Zhu et al., 2023). Soil salinity is a primary impediment hindering the growth and development of cotton in this area (Zhou et al., 2021). Excessive soil salinity will accumulate too much sodium and chloride ions, which will antagonize the absorption of essential nutrients such as K^+, Ca^{2+} and Mg^{2+}, leading to nutritional imbalance. Excessive salt accumulation will cause salt stress on the cotton root system, inhibit its use of water, and affect the growth and development of cotton. Under salt stress, cotton leaves will turn yellow or brown, chlorophyll content will decrease, and photosynthesis efficiency will decrease. This directly affects the carbon assimilation capacity of cotton, thereby inhibiting growth and yield. Soil salinization occurs when salts and water in the soil are transported and redistributed under certain natural conditions, such as high temperatures causing evaporation or heavy rainfall, leading to an excessive accumulation of salts in certain areas. Soil moisture is one of the main driving forces of salt migration. Water in the soil drives salt migration through capillary action, osmosis and evaporation. In the linear model, soil salt concentration fluctuates with changes in soil moisture content. El-Naggar et al. (El-Naggar et al., 2021) imaged the conductivity of the soil profile and its relationship with soil moisture patterns and drainage characteristics. Zare et al. (Zare et al., 2020) used EM38 data to perform two-dimensional time-lapse imaging of the soil wetting and drying cycles of flood-irrigated cotton fields. It can be seen that soil moisture has a certain correlation with the conductivity of the soil profile, and soil moisture is an important factor that cannot be ignored in the study of soil salinization.Understanding the salinity content of soil profiles at different depths is critical for studying soil salinization phenomena (Ren et al., 2019; Corwin, 2021; Kalinitchenko et al., 2021). The distribution of salt in the soil profile usually shows obvious depth changes. Especially in the arid areas of southern Xinjiang, when evaporation is strong, the evaporation of surface water drives the migration of salt and accumulates on the surface; and the unique sandy soil has strong permeability, so salt easily migrates to the deep layer with the water flow.Currently, small-area studies can obtain accurate soil salinity profiles through field sampling and indoor analysis (Gits-Muselli et al., 2020; Rahman et al., 2020). However, this method is impractical for medium and large-scale regional studies due to its time-consuming and labor-intensive. Smart agriculture technologies, such as remote sensing satellites, hyperspectral imaging and near-earth sensing, enable the acquisition of spectral information or conductivity values indicative of soil salinity content. By correlating this spectral data with local sampling and laboratory analyses, an inverse model can be constructed to estimate soil salinity in the whole profiles (Schmidt and Bijmolt, 2020; Tancik et al., 2020). This approach significantly improves the efficiency of monitoring soil salinity over larger scales. However, there is a current technology face limitations in achieving high-precision monitoring, particularly for deep soil layers, where remote sensing proves inadequate (Gorji et al., 2019; Wang et al., 2020). Remote sensing technology mainly relies on electromagnetic wave signals, which have limited penetration depth. Optical remote sensing technology, in particular, can only obtain soil information within the range of 0-5 cm soil layer. Even microwave remote sensing, which has slightly stronger penetration ability, generally does not exceed 10 cm. This poses a great challenge to monitoring soil salinity above 20 cm. Thus, there is a pressing need for developing a new approach to accurately estimate and analyze deep soil salinity.

Data assimilation introduced by Charney et al. (Desamsetti et al., 2019) focuses on the distribution of the data and the observation and background field errors, continuously incorporating observations into the model to enhance its accuracy (Geer, 2021). By substituting acquired observations into the model, and the accuracy of predictions is iteratively corrected, resulting in a close correlation with the true value (Ahmed et al., 2020). The primary objective of the data assimilation is to continuously adjust the data model based on the existing data and spatio-temporal distributions, so that the model predicted values align more closely with the actual situation, thus improving prediction accuracy (Zhang et al., 2020; Buizza et al., 2022). Lei et al. (Lei et al., 2020) performed data assimilation of high-resolution thermal and radar remote sensing retrievals for soil moisture monitoring in drip-irrigated vineyards, improving water use assessment and irrigation management. Both surface and root zone soil moisture predicted by the Soil-Vegetation-Atmosphere Transport model were improved by data assimilation, validating its effectiveness for agricultural sub-field scale irrigation management decision. Similarly, Lu et al. (Lu et al., 2021) proposed a data assimilation framework for soil moisture and canopy cover observations to reduce the requirement for parameter calibration in maize simulation using AquaCrop. Their findings indicated data assimilation effectively improved model performance, with joint assimilation outperforming univariate assimilation for accurate maize yield estimation and prediction. Ziliani et al. (Ziliani et al., 2022) incorporated CubeSat data (A small, standardized satellite, typically used for scientific research, technology validation and educational purposes. They can carry a variety of sensors and equipment and collect many types of data) into a crop model to predict crop yields in the early season., which is quite informative for enhancing digital agriculture objectives and improving end-of-season yield prediction. Kalman filter-based algorithms and their variants are extensively used in data assimilation. For instance, Kivi et al. (Kivi et al., 2022) employed the ensemble Kalman filter algorithm to overcome the limitations of process-based models and observational data to improves the accuracy and precision of agricultural forecasts. Huerta et al. (Huerta-Bátiz et al., 2022) used the Soil-Vegetation-Atmosphere Transport (SVAT) model based on an assimilation algorithm with an ensemble Kalman filter to estimate root zone soil moisture throughout a maize growing season in central Mexico. The simulated soil moisture in the top 10 cm of soil is particularly valuable for monitoring the impact of climate change on agricultural production in rain-fed areas, especially in developing countries. Huang et al. (Huang et al., 2017b) monitored

10.3389/fpls.2024.1519200

soil salinity dynamics at a specific depth (vertical direction) in different soil textures using ensemble Kalman filtering over a 480 m sample zone area, finding that the model performs better in sandy, clayey, and biphasic (sand over clay) soils. Although scholars have achieved good research results, current research on agricultural complex dynamic systems still has certain limitations, and aspects such as multi-time domain, multi-variable situations and regional-scale dynamic prediction have not been explored. For multivariate linear models, the Kalman filter algorithm, as a recursive optimal estimation method, has significant advantages. When model parameters or input data change, Kalman filtering can adjust the estimation results in time without re-adjusting the entire model. It is especially suitable for dealing with multi-variable problems affected by environmental changes, such as prediction of changes in soil salinity over time and moisture.

In this paper, the irrigated cotton fields in the Science and Technology Park of the Twelfth Regiment in Alar City, Xinjiang were used as the study area. The apparent conductivity at different effective depths was used to predict the measured conductivity at different periods and soil depths, and then a multiple linear regression model of measured conductivity and soil depth at each period was established by using the dummy variable processing method, and the model was corrected by the Kalman filtering algorithm to improve the accuracy of the model. In this way, the measured conductivity data can be used to predict the conductivity data at different depths of the soil layer, and then the Kalman filter algorithm can be used to correct the multivariate linear model of conductivity and soil depth. With the help of the above methodology, the relationship between soil depth and profile salinity is investigated, and the soil profile salinity is monitored and predicted in different periods, which provides a reference for planting and quantitative irrigation in cotton fields. This paper focuses on answering the questions of whether the data assimilation algorithm improves the accuracy of the prediction model and how much the accuracy of the prediction model is improved.

2 Materials and methods

2.1 Study area overview

The study area is located within the Twelfth Regiment of Alar Reclamation District, Xinjiang Uygur Autonomous Region. This region (see Figure 1) is situated on the south bank of the upper Tarim River, on the northern edge of the Taklamakan Desert, at the center of the Tarim Basin, and directly across the Tarim River from Alar City, the central city of the Corps in southern Xinjiang. The area of the district is 524.8 km^2, with 318,000 acres of cultivated land, of which 240,000 acres are dedicated to cotton and 78,000 acres to forestry and fruit industry. The elevation of this area ranges from 990 to 1040 m, with the water table fluctuating between 1.4-11.7 m. The primary rainfall period is June through August, the region experiences anthe evapotranspiration ratio of 40:1, characteristic of an extremely arid climate. Rich in light and heat resources, and the frost-free period extends from206 to 220 days. The geographic location of the study area is between 81°29′-81°35′E and 40°48′-40°51′N, covering a total area of 4.45 km^2. The soil texture type is predominantly sandy loam, and the primary crop cultivated is cotton. A total of 15 hm^2 of mechanically harvested cotton were selected, located between 40°50′-40°51′N and 80°34′-81°35′E. The dominant cotton variety is Xinluzhong 78, and irrigation water is mainly from the Tarim River and upstream reservoirs. The cotton fields in the study area were drip irrigated five times throughout the growth cycle on 20 June, 8 July, 27 July, 14 August and 28 August, with additional diffuse irrigation on 21 November. The organic matter content is approximately 5.34 g kg^{-1}.

2.2 Data collection and sampling

An EM38-MK2 geodesic conductivity meter manufactured by Geonics was used to collect apparent conductivity data

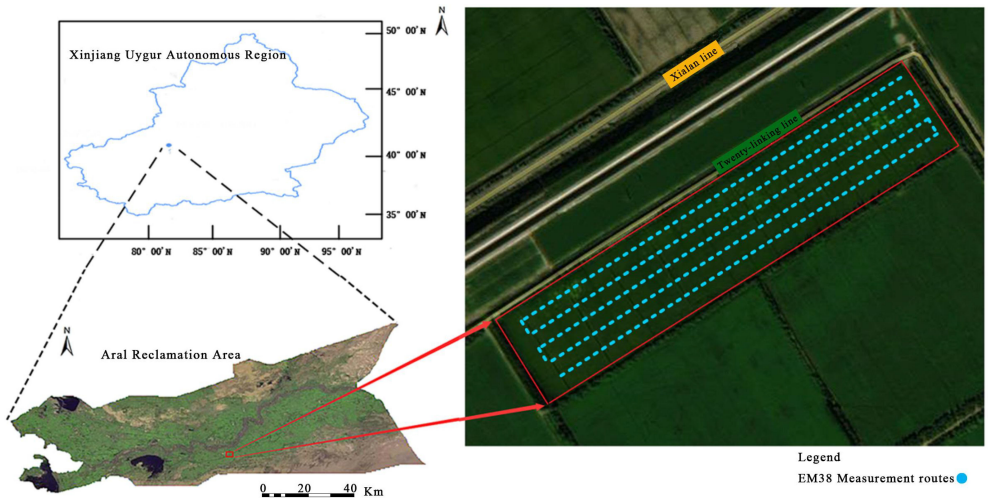


FIGURE 1
Geographical location map of the study area.

(see Figure 2) (Kasim et al., 2020). Apparent conductivity and soil samples were collected in March, June, July and October 2021, respectively. The instrument has a transmitting coil (Tx) and two receiving coils (Rx). The distance between Tx and the two Rx is 0.5 m and 1.0 m respectively, so the instrument can measure two kinds of apparent conductivity data under the same model ($EC_{x0.5}$ and $EC_{x1.0}$, X represents mode). The EM38-MK2 instrument can collect apparent conductivity measurement results from both vertical and horizontal directions, generating four different data sets, namely $EC_{h0.5}$, $EC_{h1.0}$, $EC_{v0.5}$, and $EC_{v1.0}$. The instrument recording mode record data automatically with the frequency of one point per second from both vertical and horizontal directions. Approximately 1000 sets of apparent conductivity data were collected for each period, with a sampling interval of 1 m. The apparent conductivity data collected in the 'S' linear test area were used to determine the best model apparent conductivity data set for different soil layers for linear inverse modeling of soil salinity and its quantitative relationship with soil depth.

Soil samples were systematically collected in compliance with principles of typicality and representativeness, corresponding to the same timeframes as the conductivity data. Employing a stratified random sampling (SRS) approach, samples were collected using soil auger at depths of 0-20, 20-40, 40-60, 60-80 and 80-100 cm. Each sample was promptly placed in a sealed bag, labeled, and transported to the laboratory for soil salinity test.

2.3 Soil moisture content and soil electrical conductivity

The soil moisture was determined using the drying method, which involved the following steps: (1) Soil samples were first placed in small aluminum containers and then baked in an oven at 105°C for 2 hours, followed by cooling to room temperature (25°C) in a desiccator. The samples were subsequently weighed, ensuring consistent weightings with an accuracy of 0.001 g). (2) The mixed soil sample was transferred to an aluminum container using an angular spoon, and its mass was recorded using a precision balance. (3) The aluminum container was opened, and the soil was baked in an oven at 105 ± 2°C for 6 h. After removing the lid, the container was cooled in a desiccator to cool to room temperature and immediately weighed.

The soil electrical conductivity was measured using the conductivity method (Rooij and Schulz, 2020). The generally high soil conductivity in the study area, such as the 1:1 soil-to-water extracts, often exceeded the measurement range of the assay instrument, rendering direct measurement impractical. Therefore, a water-to-soil mass ratio of 5:1 was typically employed. The mixture was thoroughly shaken, allowed to settle, and the leachate was filtered before determining the electrical conductivity (ECe) using a conductivity meter. To minimize the measurement errors, duplicate samples were prepared for parallel determinations of both moisture content and electrical conductivity. For the five sampling sites in March, measured soil conductivity (ECe), PH, soil sample moisture content (θ) and salinity content (α) were determined, and the nearest-neighbor interpolation method in the MATLAB R2020a Griddata function was utilized to map the depth profile display of the 0.2-1.0 m soil layer at the sampling sites (see Figure 3), whereas for the other soil samples only ECe and soil sample moisture content of the soil samples.

2.4 Modelling approach

Time series, also known as dynamic series, is a numerical sequence of indicators of a phenomenon arranged in chronological order. Time series analysis can be divided into three major parts, namely, describing the past, analyzing the law and predicting the future (Debella-Gilo and Gjertsen, 2021). Because the time series is a long-term change in the numerical value of an indicator of numerical performance, so the time series of numerical changes behind the inevitable numerical changes in the numerical transformation of the regularity of these regularities is the entry point of time series analysis. In general, the law of change of the value of the time series has the following four kinds of change: (1) long-term trend; (2) seasonal changes in the law; (4) cyclical changes in the law; (4) irregular changes (random perturbation term) (Reddy and Henze, 2011). A time series is often a superposition of the above four types of variability.

The sample series used in this paper are affected by soil depth, and the soil depth adjustment method with dummy variables is

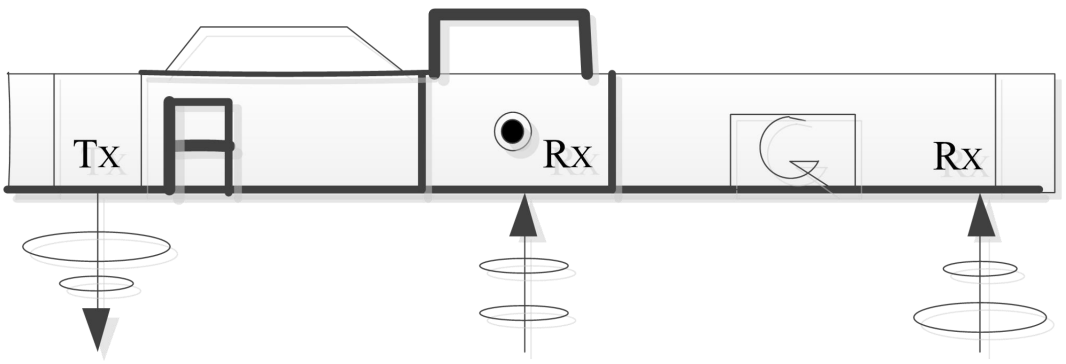


FIGURE 2
EM38-MK2 simple diagram.

10.3389/fpls.2024.1519200

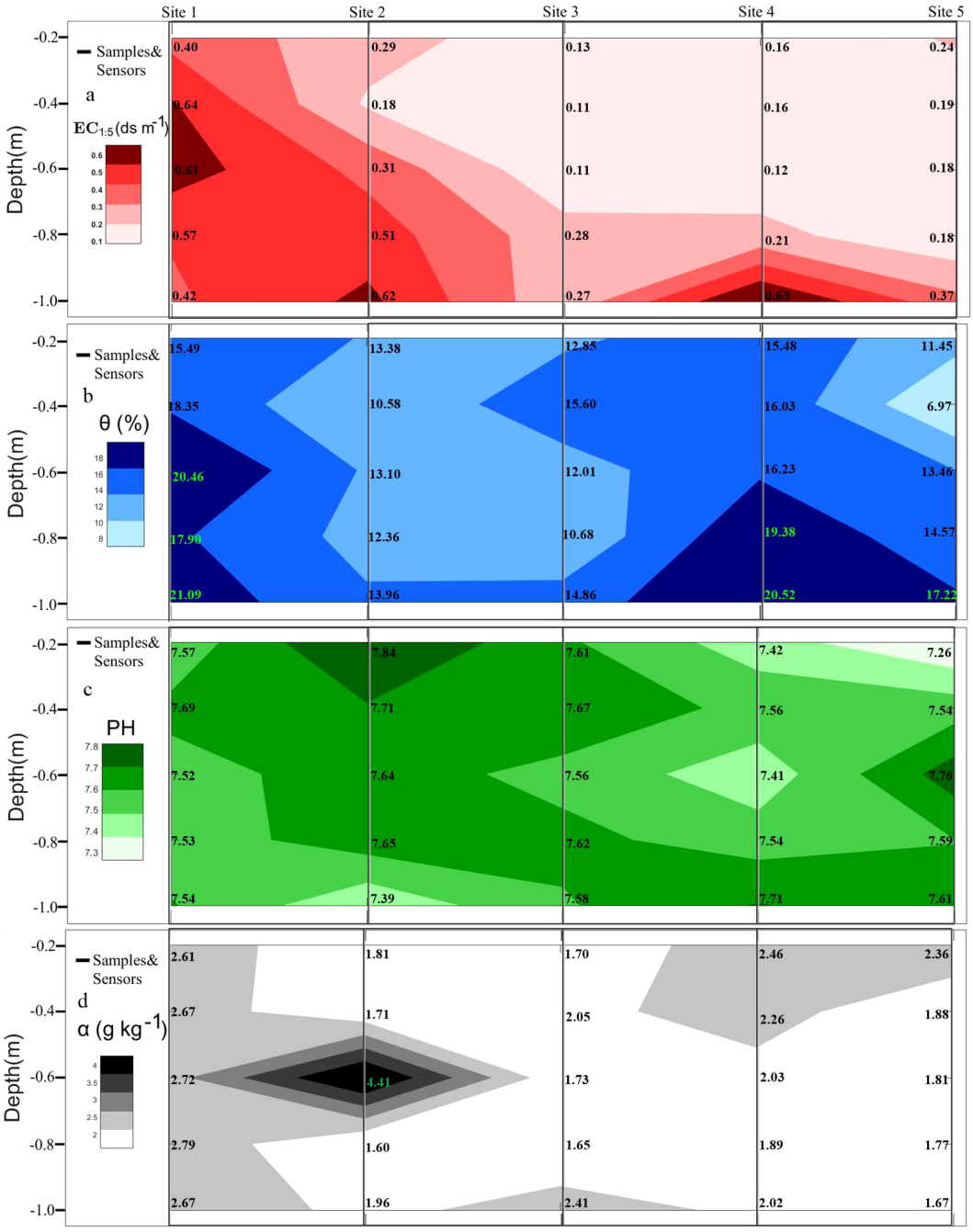


FIGURE 3
Contour plots of various soil properties measured at five sampling points, including: **(A)** measured electrical conductivity of the soil (EC1:5, dS m^{-1}), **(B)** PH value, **(C)** soil water content (θ), and **(D)** salinity content (α g kg^{-1}). Data are for reference only and are subject to some uncertainty.

utilized to eliminate the effect with soil depth. This soil layer adjustment method is based on the fact that the soil layer depth obeys an additive model $Y = T + S + C + I$, where Y, T represent aggregate indicators with the same units; and S, C, I produce either positive or negative deviations from the long-term trend. A multiple linear regression prediction model was then developed using soil depth x_i and sample T as variables:

$$y = \beta_0 t + \beta_i x_i + C$$

Where, β_i and C for constant coefficient; y represents the predicted value of ECe; t represents the number of samples; x_i represents apparent conductivity.

For sample selection, 12 groups were randomly selected from the 18 groups of measured conductivity data in each period at a ratio of 1:2 as the modeling set, and the remaining 6 groups were used as the validation set. In the above way, time series analysis can help us to understand the change of ECe with soil depth and capture the pattern and trend of observations over time. And it also makes the data

smoother and reduces the interference of random fluctuations in the analysis, providing a strong guarantee for the later work.

For the prediction of yield, sample T is not considered. The seed cotton yield of the first three years (2016-2018) is used as the modeling set, and the data of the last two years (2019-2020) is used as the validation set. The different soil conductivity data at the bud stage (June) and the boll stage (July) are used as modeling factors to establish a multivariate linear model:

$$y = \beta_i x_i + C$$

Where, C and β represents a constant; x_1 represents 0-20 cm conductivity data.; x_2 represents 20-40 cm conductivity data; x_3 represents 40-60 cm conductivity data; x_4 represents 60-80 cm conductivity data; x_5 represents 80-100 cm conductivity data. In this way, the cotton production of this year can be predicted based on the predicted conductivity data.

2.5 Data assimilation

Data assimilation algorithms can generally be classified into two mains: sequential assimilation and continuous data assimilation. The earliest of the sequential data assimilation algorithms is the Kalman filter algorithm, and its idea lays a theoretical foundation for other sequential data assimilation algorithms. The Kalman filter algorithm takes the observed data and the statistical characteristics of the error between the model and the data as the basis, and estimates the state quantities with the restriction of obtaining the minimum error of the state estimation value (Akram et al., 2019). The specific modeling is as follows.

2.5.1 Forecasting

$$X_{k+1}^{f} = M_{k,k+1} X_k^a \tag{1}$$

where k and $k+1$ are the two states and the k+1 is the next state of k. f and a represent the predicted and analyzed values, respectively; X_{k+1}^{f} represents the predicted value of k+1, which the $sM_{k,k+1}$ represents the linear relationship from k state change to k+1 state; the X_k^a represents the analyzed value of k.

$$P_{k+1}^{f} = M_{k,k+1} P_k^a M_{k,k+1}^T + Q_k \tag{2}$$

where P_{k+1}^{f} and P_k^a represent the error covariance matrix between the predicted and analyzed values; Q_K is the model error variance matrix. From Equations 1 to 2, it completes the calculation of the pair from k state to $k+1$ state.

2.5.2 Updates

After the observations are acquired in the $k+1$ state, the corresponding analyzed values and error covariances are updated based on the observations at this moment in time, see Equations 3–5.

$$X_{k+1}^a = X_{k+1}^f + X_{k+1}(Y_{k+1}^o - H_{k+1} X_{k+1}^f) \tag{3}$$

where the superscript o represents the observation; X_{k+1}^{f} is calculated from equation 1; X_{k+1}^a represents the analyzed value at the k+1 state; K_{k+1} represents the Kalman gain solved at the k+1 state; Y_{k+1}^O represents the observation acquired at the k+1 state; H_{k+1} represents the observation operator solved at the k+1 state.

$$K_{k+1} = (H_{k+1} P_{k+1})^T [H_{k+1} (H_{k+1} P_{k+1}^f)^T + R_{k+1}]^{-1} \tag{4}$$

$$P_{k+1}^a = (I - K_{k+1} H_{k+1}) P_{k+1}^f \tag{5}$$

where P_{k+1}^{f} are calculated from Equation 2, respectively; P_{k+1}^a represents the error covariance matrix thatis solved for the analyzed values at the k+1 state. Equations 3–5 are based on the observations acquired in the k+1 state and adjusted to the background field to obtain the analytical values X_{k+1}^a and the error covariance matrix P_{k+1}^a at the state and provide the data base for the operations in the next $k+2$ state.

In this paper, the mean value of 0-20 cm soil sample data is set as the initial state; the difference between the first sample value and the mean value is used as the initial error value. The process noise and observation noise are set to white noise, which means that the noise has no dependence on time and is statistically independent of each other. Regarding data preprocessing, soil salinity data generally have significant trends and fluctuate with seasonality, which may cause the data to deviate from Gaussian distribution. In this case, a moving average method is used to remove these components, making the data more stationary and consistent with Gaussian assumptions. For the processing of outliers, use statistical methods to detect and remove outliers to ensure that the data set more closely conforms to the Gaussian distribution. Draw technology flow diagrams (see Figure 4).

In contrast to the optimal interpolation algorithm, the Kalman filter algorithm takes into account the state of the model and its changes, and its prediction error varies with the model dynamics process. Compared with the variational algorithm, the Kalman filtering algorithm is easier to implement and its arithmetic process does not need to take into account the accompanying patterns of the model. However, for the use of Kalman filtering algorithms, it is also important to consider whether the data conforms to a Gaussian distribution and whether the state changes conform to linear changes so that the optimal solution can be obtained (Huang et al., 2019). Therefore, whether the Kalman filtering algorithm can be applied in real-world problems requires further model validity tests to be carried out.

2.6 Model validation

The quantitative response model of soil depth to profile salinity was developed by firstly establishing a multiple linear regression equation between measured conductivity and soil depth using dummy variable treatment and least squares method, and then calibrating the model using Kalman filter algorithm. After dividing the sample data into modelling and validation sets, they can be validated using cross-validation. Secondly the coefficient of

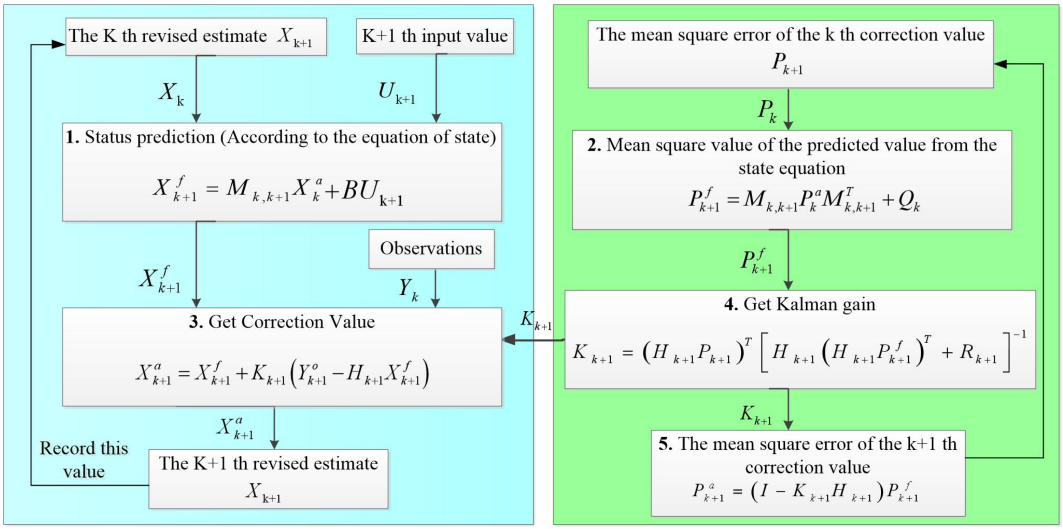


FIGURE 4
Technical flow chart.

determination R^2 and the root mean square error RMSE can be used to further judge the effectiveness of the fit between the predicted values and the actual measurements.

$$R^2 = \sum(\hat{y}_i - \bar{y})^2 / \sum(y_i - \bar{y})^2 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N}\sum(\hat{y}_i - y_i)^2} \quad (7)$$

Where $\hat{y}_i$, y_i and $\bar{y}_i$ displays the predicted value, true value and average value of the variable, respectively. And the larger the R^2 value, the smaller the RMSE, the higher the model accuracy, the stronger the predictive ability and the better the stability. On the contrary, the smaller the R^2 value and the larger the RMSE, the worse the model accuracy and the weaker its predictive ability.

3 Results and analysis

3.1 Inverse modelling of conductivity in soil layers at different depths

The soil conductivity data (ECe) was measured primarily using the saturated paste method on soil samples collected during various periods. Since the detection range of the data collected through the EM38-MK2 is limited to the free state electrolyte content in the deep soil space, the conductivity ECe was used to provide an accurate expression of the salinity status of the agricultural soil, and there was a high correlation between the conductivity EC_e collected using the saturated slurry method and the soil salinity. Referring to the research idea of Wang et al., descriptive statistics of soil properties established in the study area for 75 sampling points were carried out considering the sensing range of EM38-MK2 and its degree of response under the change of soil depth under different modes, which can be accessed from the horizontal and vertical aspects of the modelling factors, respectively (Wang et al., 2021). Taking the data collected in June 2021 as an example, the apparent conductivities of 0.5 m and 1.0 m in the horizontal mode and 0.5 m and 1.0 m in the vertical mode were selected as the independent variables, which were denoted by $EC_{h0.5}$, $EC_{h1.0}$, $EC_{v0.5}$ and $EC_{v1.0}$, respectively. With the measured conductivity of different soil depths collected, a linear model was established and the accuracy was compared (see Figure 5).

The R^2 of $EC_{h0.5}$ in the horizontal mode reached a maximum value of 0.87 at 20-40 cm, and thereafter, showed a decreasing trend with increasing soil depth. And the R^2 of $EC_{v1.0}$ in vertical mode reaches its highest value of 0.76 at 80-100 cm. From the above table, it can be seen that $EC_{h0.5}$ can be chosen as the modelling factor for the 0-40 cm range. Whereas after 40-80 cm, either $EC_{h1.0}$ in the horizontal mode or $EC_{v0.5}$ in the vertical mode can be chosen as the modelling factor. In this paper, $EC_{v0.5}$ in the vertical mode is chosen as the modelling factor. At 80-100 cm, only $EC_{v1.0}$ can be selected as a modelling factor. The results of the analyses in the above table, which are the same as those analyzed by Li et al, are consistent with the response function model between the conductivity data and the change in soil depth in this measurement model (Li et al., 2020).

3.2 Effect of soil moisture content on the accuracy of conductivity inversion models

The EM38-MK2 instrument measures the free conductive medium in the soil, and its accuracy is influenced primarily by soil moisture and salinity. To assess the effect of moisture content on model accuracy, soil samples were categorized into three moisture gradients: 0-10%, 10-20%, and >20%. The samples were divided into modeling and prediction sets in a 2:1 ratio, ensuring that the prediction set covered the full range of moisture content. Linear models were constructed to analyze the relationship between

10.3389/fpls.2024.1519200

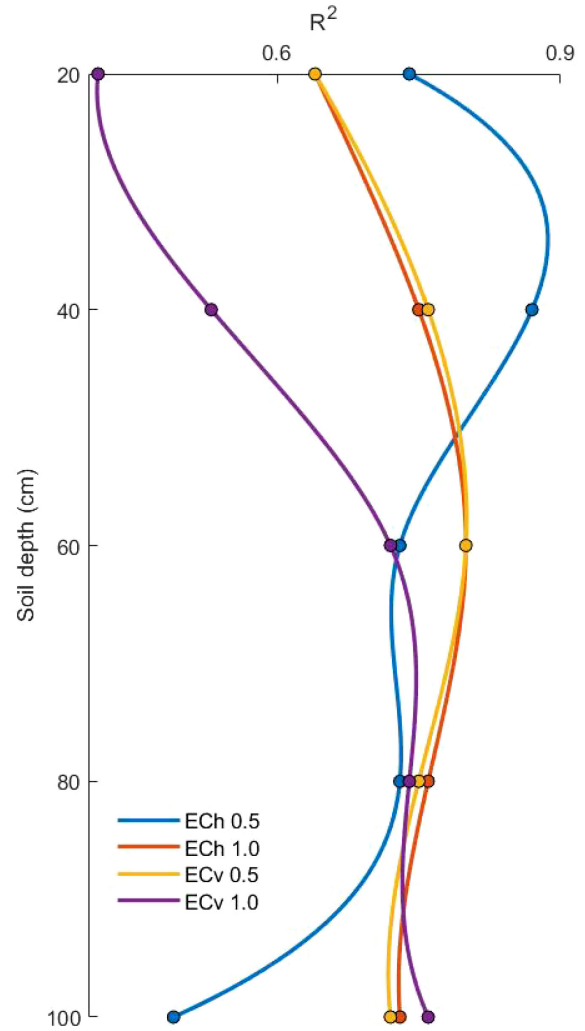


FIGURE 5
Soil depth determination factor R^2 for different modelling factors.

apparent conductivity and measured soil conductivity at each gradient.

From Figure 6, it can be observed that the model's R^2 was 0.72, and the RMSE was largest when soil moisture was in the 0–10% range, indicating poor model accuracy. In contrast, for soil moisture levels between 10–20% and above 20%, the R^2 values were 0.84 and 0.86, respectively, with smaller RMSE values. These findings suggest that soil moisture levels above 10% have less impact on model accuracy. To ensure model accuracy, soil moisture should be maintained above 10% during modeling. The inverse model was further developed for different periods, and the model accuracy for each period is shown in Tables 1 and 2.

The accuracy of the conductivity interpretation model was compared across different time periods. For the 0–40 cm soil depth range, $EC_{h0.5}$ in the horizontal mode was selected as the modeling factor, while $EC_{v0.5}$ and $EC_{v1.0}$ in the vertical mode were chosen for the 60–100 cm range. The results show that the model's accuracy and stability were higher in July and October, when soil moisture levels were above 10%, compared to March and June when moisture levels were lower.

3.3 Kalman filter algorithm and implementation

In March, June, July and October 2021, a total of four soil sample collections were carried out, each time 18 fixed points were selected to sample the soil profile at different depths, and measured conductivity data were measured for each period. The optimal sampling scheme is selected by comparing the effect of different isotopes on the accuracy of the inversion model, so that the measured conductivity of soil layers at different depths over time can be predicted with the help of the apparent conductivity of soil layers representing different effective depths in different measurement modes. In order to further explore the relationship between soil salinity in the profile and soil depth, the relationship between soil depth and soil salinity in the profile was further investigated with the help of measured conductivity EC_e data measured at different depths of the soil layer in each period. Firstly, 12 sets of the 18 sets of measured conductivity data for each period were randomly selected on a 1:2 basis as the modelling set. The sample series used were affected by soil depth, and the effect with soil depth was removed using a soil depth adjustment method with dummy variables, followed by a multiple linear regression prediction model using the least squares method.

From the Kalman filter and its variant forms, it can be seen that the Kalman filter algorithm is more suitable for linear problems; while the extended Kalman filter algorithm can be processed through the Taylor's formula, which can convert the nonlinear problem into a linear problem to be processed; and the ensemble Kalman filter algorithm is more suitable for the multi-dimensional space which is strongly nonlinear (Khodarahmi and Maihami, 2023). In view of the strong linear correlation between the salt content of the soil profile and the depth of the soil layer at different times and at different depths, the Kalman filter algorithm was chosen to further study the relationship between the salt content of the profile and the depth of the soil layer.

Using the Kalman filter algorithm, the multivariate linear inverse prediction models of measured conductivity and soil depth for different periods were calibrated, and the accuracy of the models was all improved to some extent. The best correlation between predicted and measured conductivity data was found in March, with an R^2 of 0.79. It is significantly higher than the other three months. However, the model accuracy before calibration reached 0.78, with the lowest growth rate. The model-corrected determination coefficients in June and October were slightly higher than that in July, with R^2 being 0.59 and 0.60 respectively; however, the model-corrected determination coefficient R^2 in July increased the most, from 0.32 to 0.58, with an increase rate of 81.3%. Therefore, using the dummy variable processing method and the least squares method to establish the multivariate linear inversion model of measured conductivity and soil depth in each period, and then using the Kalman filter algorithm to correct the model, the correlation between the model predicted value and the measured value was obviously improved, and the model prediction effect became better. Comparison of the accuracy of the corrected predicted and measured values for different periods (see Figure 7).

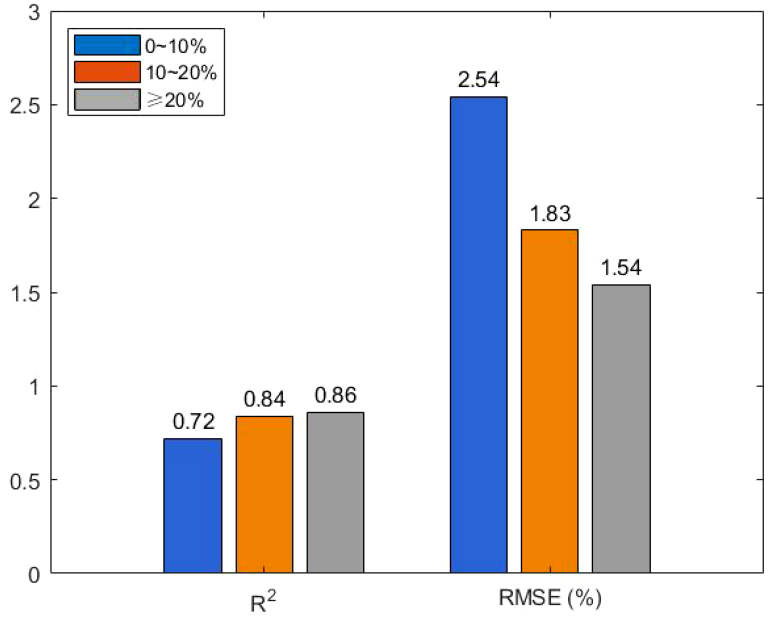


FIGURE 6
Comparison diagram of model accuracy under different soil moisture gradient.

In Figure 7, it can be seen that the best fit between the corrected best predicted values and the measured values for the month of March was the best, with a fitting accuracy (coefficient of determination R^2) as high as 0.79 and the smallest root mean square error (RMSE) of 96.17 μS cm^{-1}; The fit was relatively worst in July, with a fitting accuracy R^2 of only 0.58 and a root mean square error (RMSE) of 53.74 μS cm^{-1}. The fit is also better in October, and the fitting accuracy R^2 also reaches 0.60, while its root mean square error (RMSE) is also lower at 105.58 μS cm^{-1}. The fit between the corrected best predictions and the measured values shows some variability across time, which is due to a number of reasons. Measurement errors caused by objective reasons such as ambient temperature, soil moisture and instrument sensitivity all have a certain impact on the experimental results and predicted values, which in turn cause some variability in the measured data and predicted results in different periods.

For the implementation of the Kalman filtering algorithm, the observation equations are used as the basis for further determining the state function, setting up the number of simulations (there are a total of 90 points in each period, so it is set to be 90, and the period is 5), the simulation time, the system state noise, and the observation noise, as well as further initializing the state equations, the output equations, and the covariance matrix. The initial value of the Kalman filter state is given and the error value is initialized and set to 0. On this basis, the operation can be carried out according to the idea of Kalman filter algorithm. In the operation process, the prediction of the state is needed first, and then the prediction error autocorrelation matrix is derived; Next, the Kalman gain solution is performed; Finally, state estimation and observation prediction can be obtained. After several iterative MATLAB operations, the corrected best predicted values were obtained. The simulated comparison between the measured values and the corrected best predicted values for different periods of time, and the correlation control plot between the two (see Figure 8).

As can be seen in Figure 8, overall, the fit between the corrected best predicted and measured values for different periods is better, with a high degree of approximation between the two. The Kalman

TABLE 1 Model of conductivity interpretation in different periods.

	$ECe=a+b\times EC_{h0.5}$				$ECe=A+B\times EC_{vx}$					
	0-20(cm)		20-40(cm)		40-60(cm)		60-80(cm)		80-100(cm)	
Month	a	b	a	b	A	B	A	B	A	B
March	-0.679	0.907	-0.665	0.834	-0.083	0.503	0.274	0.454	0.325	0.476
June	-0.645	0.769	0.843	0.694	0.264	0.456	0.283	0.463	0.304	0.473
July	-1.566	1.566	-1.424	1.011	-0.675	0.785	0.301	0.496	-0.079	0.521
October	-0.671	0.863	-1.347	1.432	-0.372	0.534	-0.405	0.573	0.353	0.479

a, b, A, B represent constants; ECe represents measured conductivity data; $EC_{h0.5}$ represents 0.5m apparent conductivity data in horizontal mode; EC_{vx} represents 0.5 m and 1.0 m apparent conductivity data in vertical mode.

10.3389/fpls.2024.1519200

TABLE 2 Differences in model accuracy for different soil moisture content gradients.

Date	Judging Indicators	0-20(cm)	20-40(cm)	40-60(cm)	60-80(cm)	80-100(cm)
2021/03/10	R^2	0.81	0.78	0.89	0.86	0.77
	RMSE	0.0673	0.1284	0.0627	0.0671	0.1307
2021/06/03	R^2	0.71	0.82	0.81	0.80	0.72
	RMSE	2.3875	0.9087	1.9132	0.9284	2.3894
2021/07/07	R^2	0.79	0.89	0.87	0.88	0.85
	RMSE	0.1387	0.0723	0.0834	0.0915	0.0882
2021/10/19	R^2	0.77	0.84	0.86	0.85	0.86
	RMSE	0.1289	0.0923	0.0627	0.0804	0.0757

filter algorithm was used to calibrate the multivariate linear model of measured conductivity and soil depth for each period, to find out the correlation between the measured and predicted values of the model for each period after calibration, and to compare the accuracy of the model with the accuracy of the model before calibration. The results of the comparison of the fitting accuracy between the corrected best prediction and the measured values are obtained for different periods.

In order to further investigate the relationship between soil profile salinity and soil depth, the results of this research were combined with the results of previous studies. The predicted values obtained in each period were averaged according to the depth of the soil layer. In order to facilitate the comparison of the results, the overall effect of soil salinity accumulation between different months corresponded to July, June, March and October in descending order, and the related soil salinity accumulation process was plotted (see Figure 9).

As can be seen in Figure 9, soil conductivity showed a gradual accumulation trend at different periods. From the month of March, we can see that salts appeared to accumulate significantly in the soil layer from 40-80 cm and reached a peak value of 0.76 dS m^{-1} at 60-80 cm. Whereas, soil conductivity in October is peaked at 40-60 cm

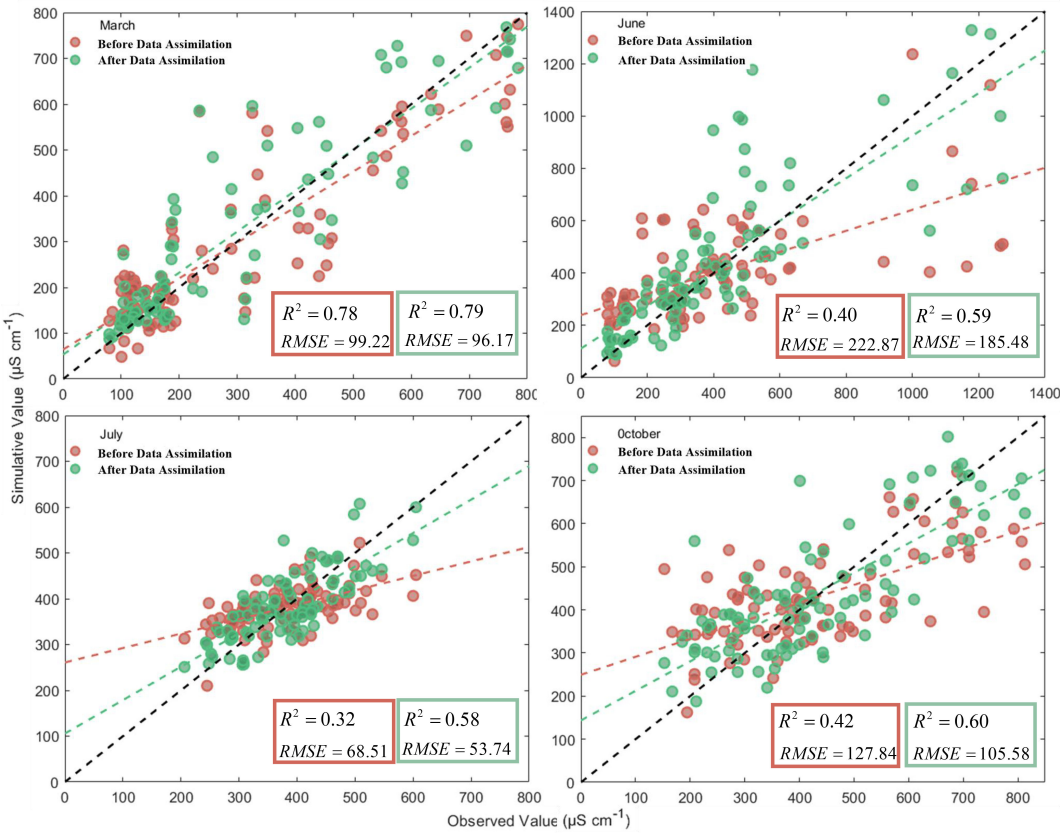


FIGURE 7
Comparison of observed and simulated values before and after assimilation at different time periods.

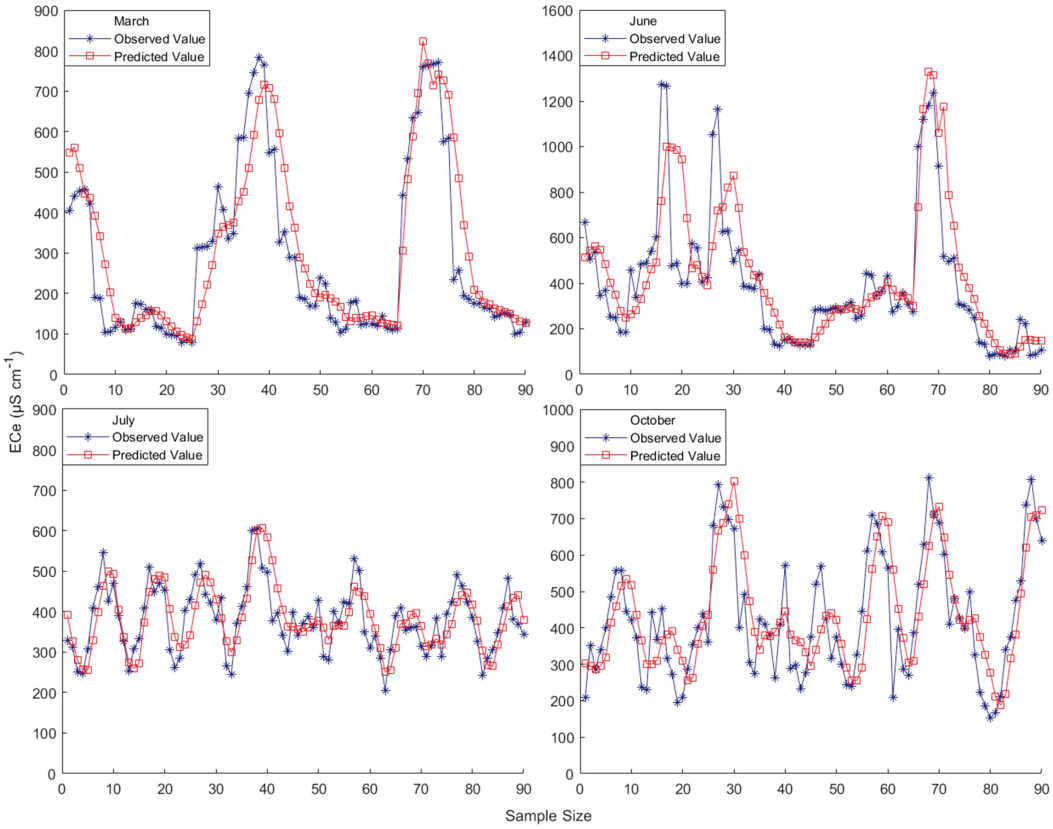


FIGURE 8
Measured and predicted values by period.

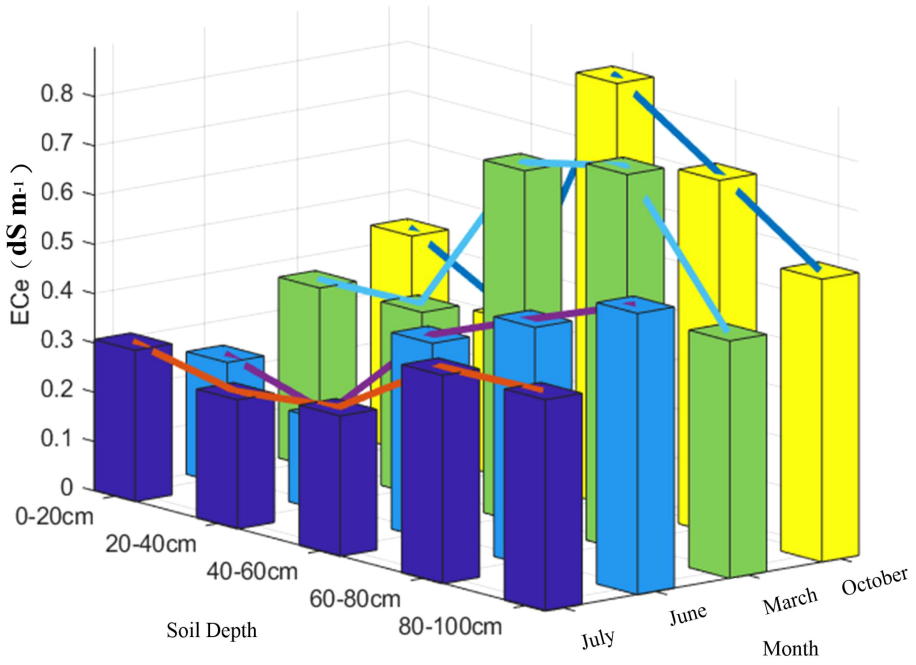


FIGURE 9
Salt accumulation process in soil layers in different months.

with a value of 0.86 dS m^{-1}. However, from 40-80 cm, it is still a highly aggregated zone of salts at soil depth. In June and July salinity showed an aggregation with increasing soil depth, and this trend remained significant from 40-100 cm. Combined with the actual growth cycle of cotton and planting methods in South Xinjiang, the months of March, June, July and October can be defined as: pre-sowing, bud stage, boll stage and harvesting period, respectively. The months of June and July are the growing period of cotton, during which cotton is grown using drip irrigation, and it can be seen that the conductivity at 0-60 cm is significantly lower than in the other two months. However, as can be seen in Figure 4, soil salinity appeared to accumulate significantly in the 60-100 cm soil layer, both in June and July (growing period) and in March and October (non-growing period). Further explanation, although drip irrigation can wash the salt in all directions, but in the vertical direction, with the deepening of the soil layer, the washing effect will gradually weaken, and the salt will eventually accumulate in the deep soil. After the cotton harvest, it is necessary to carry out pre-winter deep ploughing, irrigation and alkali pressure, drainage and irrigation support, in order to achieve the purpose of salt washing and desalination. From Figure 4, it can be seen that before sowing in March, the conductivity from 0-40 cm was lower and less variable than that in October (harvesting period), with a mean value of 0.47 dS m^{-1}. This further indicates that, after winter irrigation, the salinity of the soil surface was uniformly distributed, which is basically in line with the findings of Feng J et al. (Feng et al., 2021).

From the above research results, it can be seen that using the Kalman filter algorithm, the multivariate linear model can be corrected in order to improve the model prediction accuracy. Although the corrected model accuracy using Kalman filter algorithm facilitates the search and discovery of intrinsic patterns between soil profile salinity and soil depth. However, the initial model and its accuracy before correction are still important when using soil profile conductivity to study the real relationship between soil profile salinity and soil depth. This is because in Kalman filtering, the initial model and its accuracy determine the filter's initial performance, convergence speed, and the benchmark for subsequent corrections. If the initial state estimate deviates too much from the true value, the filter will need more time or iteration steps to converge to the correct state, and it may also directly affect the filter's performance and cause the system to diverge. Therefore, reasonably setting the initial state, error covariance, and model parameters can not only ensure the stability of the filter, but also provide a reliable basis for subsequent correction and optimization.

3.4 Application of data assimilation models in cotton yield and economic efficiency forecasting

The whole life span of cotton is the total number of days from the cotton seeding to the end of the flower harvest, about 210 days. It can be divided into five reproductive stages: seedling emergence, seedling stage, bud stage, boll stage and harvest stage (Lu et al., 2022). And Xinjiang because of the special geographic location, the use of under-membrane drip irrigation for planting, the study area has been used 1 membrane 2 tubes 4 rows of under-membrane drip irrigation technology planting cotton, in which the membrane width of 1.25 m, wide rows, narrow rows, as well as the distance between the membrane were 0.65 m, 0.12 cm and 0.2 cm, respectively, the drip irrigation belt laying (see Figure 10).

Cotton goes through four key periods from nutrient growth to reproductive growth: seedling stage (April), bud stage (June), boll stage (July), and fluffing stage (August). Considering the sampling date, the information collected in the previous five years was utilized, so the bud stage and boll stage were selected to determine the best yield estimation factor for cotton (see Figure 11). The relevant information of Cotton Plant Height, Chlorophyll and Production are shown in Table 3.

In this study, the seed cotton yield of the first three years was used as the modeling set and the data of the last two years as the validation set, respectively. Different soil conductivity data at bud stage (June) and boll stage (July) were used as modeling factors to establish multivariate linear models.

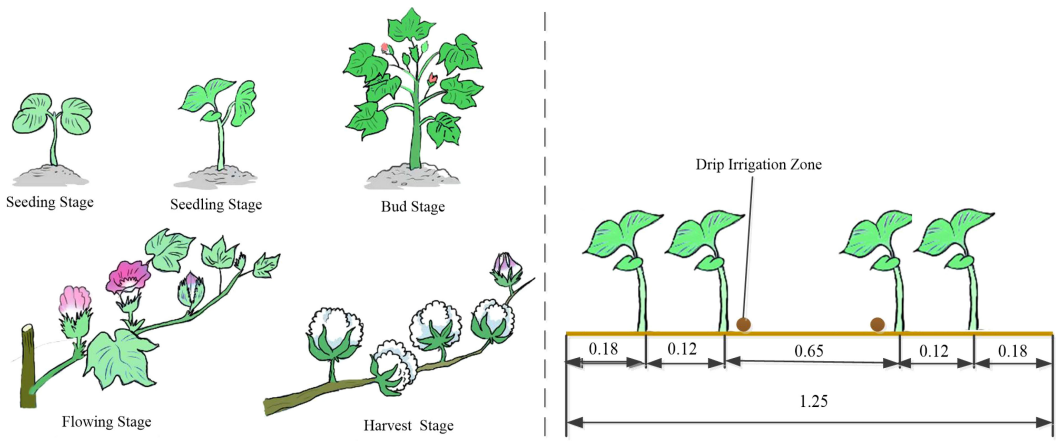


FIGURE 10
Cotton fertility period and drip irrigation belt arrangement (m).

TABLE 3 2016-2020 cotton plant height, chlorophyll and seed cotton production.

Year	Plant height/cm	Chlorophyll/ SPAD	Cotton seed yield/ (kg·hm^{-2})
2016	84.65 ± 9.03	36.49 ± 1.74	4 984.37 ± 112.95
2017	83.31 ± 6.63	40.39 ± 4.35	5 355.61 ± 98.13
2018	82.63 ± 2.5	42.29 ± 3.41	5 229.44 ± 179.02
2019	67.31 ± 5.01	32.85 ± 2.0	4344.70 ± 115.10
2020	59.65 ± 2.51	31.78 ± 2.41	4046.56 ± 243.63

SPAD stands for Soil Plant Analysis Development.

Table 4 shows the feasibility of multivariate linear modeling using different soil conductivities in June and July to predict seed cotton yield. Bringing the conductivity data predicted using data assimilation into the model yields 5203-5551 kg hm^{-2}, while the study area is 15 hm^{-2}, so the total production should be: 78048-83269 kg. According to the survey statistics of cotton purchase price of 2020 face of each regiment of Aral Reclamation Area, the purchase price of cotton is in the range of 4.88 - 5.34 RMB kg^{-1}, the government subsidized price of 2.26 RMB kg^{-1}, but the more frequent cotton purchase activities in the early and middle stage, the final farmers per kilogram of cotton revenue average price of about 7.15 RMB. And seeds, fertilizers, pesticides, mulch, drip irrigation belt, water, electricity, machine power, machine picking fee and other planting costs of about 32,250 RMB hm^{-2}. Based on the statistical results of cotton unit price and planting costs in Alar Reclamation Area, combined with the cotton production in the study area, the study area in 2021, the revenue should be: 74,296.67-111,621.54 RMB.

4 Discussion and conclusions

4.1 Effectiveness of modeling factors

This study evaluated the suitability of different apparent conductivity modeling factors across various soil depths, using the EM38-MK2 geophysical conductivity meter. By comparing the performance of different modeling factors in multiple measurement modes, the most effective factors were identified for each depth range. The findings of the present study are in agreement with those of some previous studies using conductivity instruments for soil salinity prediction, while also presenting new insights. Although, Song and Zhao et al. (Zhao et al., 2013; Song et al., 2019) have used EM38 and combined with soil measured conductivity data (ECe) to analyze salinization at different soil depths and achieved better research results. However, it is modeled by selecting the apparent conductivity data (ECa) at a specific soil depth, at the same depth, and in different measurement modes (horizontal and vertical modes), in order to determine which mode has applicability for ECa. Instead, this paper is not limited to the measurement model, but is modeling ECe at different soil depths using ECa at different depths and in different modes, respectively, to determine the most effective factor in each depth range. Compared to Song and Zhao et al, this study does not need to re-collect ECa at a specific soil depth, and extends the scope of consideration, which has the advantage of rapidity and simplicity, and saves financial and material resources. As for the research results, Sun et al. (Sun et al., 2012) had used EM38 to identify heavily saline soils and found that EM38 horizontal mode measurements were more sensitive to shallow soils, while vertical mode measurements were more sensitive to deep soils, which was further verified by the results of the present study, especially in the depth ranges of 0-40 cm and 80-100 cm, with $EC_{h0.5}$ in the horizontal mode and $EC_{v1.0}$ in the vertical mode The performance was superior and the results of this study were reliable. The reasons for the consistency are as follows: First, the research areas of this article and the scholars are both in the Tarim Basin in southern Xinjiang, where rainfall is scarce, evaporation is strong, and soil salinization is serious. Secondly, EM38 is used to predict soil salinity in different soil layers with the help of apparent conductivity. The selected methods are all linear modeling, and the fitting effect is significant.

In this paper, by optimizing the sampling strategy based on the performance of different modeling factors, the accuracy of the soil salinity prediction model has been significantly improved, reducing the need for labor-intensive field sampling, but there are some limitations. The data were mainly derived from samples under a single climatic condition, so the broad applicability of the model may be limited. The results of this study have yet to be validated under a wider range of regional and climatic conditions than those of Mello et al. (Mello et al., 2022) who used datasets with different lithologies, landforms and soil types. Future research should focus on improving the accuracy of global models for large-scale applications, potentially through the integration of other methods, such as multiple regression and covariance modeling, to enhance prediction accuracy and extend these findings to broader geographic regions.

TABLE 4 Cotton yield prediction model based on different critical periods.

Month	Model	R^2	RMSE (kg·hm^{-2})
June	. $y = 5630.70 - 157.7125x_3 - 33.93x_4$.	0.98	1023.23
July	$y = 5280.58 + 259.50x_2 - 338.78x_5$	0.99	1086.03

x_2 represents 20-40 cm conductivity data; x_3 represents 40-60 cm conductivity data; x_4 represents 60-80 cm conductivity data; x_5 represents 80-100 cm conductivity data; y represents cotton production.

4.2 The advantages of the Kalman filter algorithm

The Kalman filtering algorithm was called LLMS (Linear Least Mean Squares) when it was first proposed, because it optimizes a linear stochastic system with noise-containing sensor measurements in a least-square manner to obtain an optimal solution. In other words, it optimizes the collected data that do not contribute or interfere with the system. As for the application of

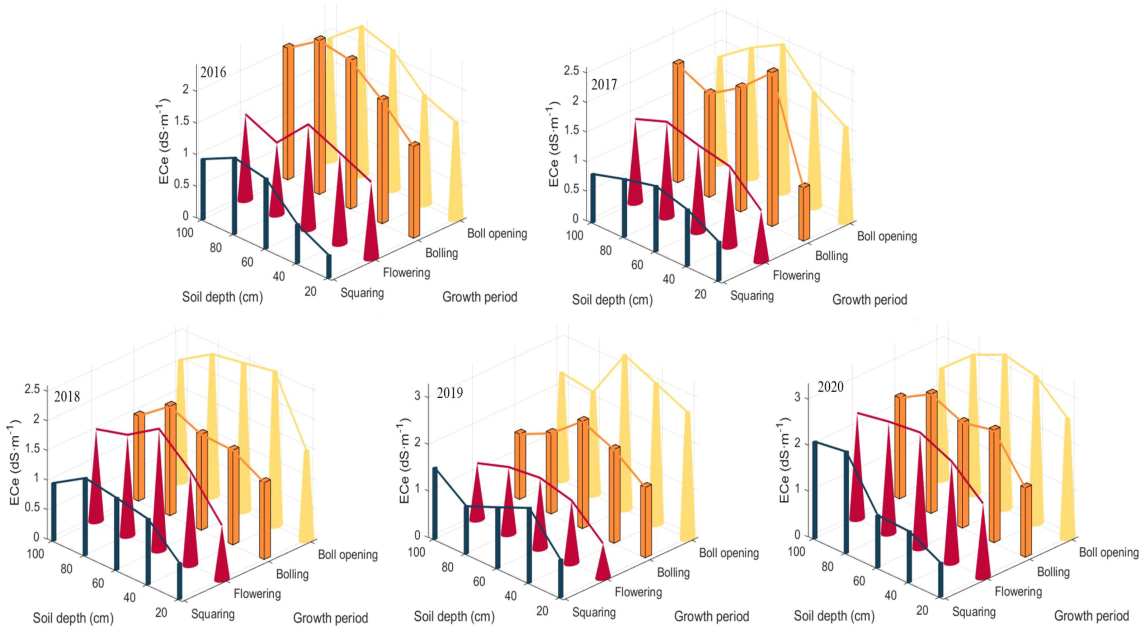


FIGURE 11
Salt accumulation processes in the 0-100 cm soil layer, 2016-2020.

Kalman filter algorithm, some scholars have long applied it to soil moisture forecasting and achieved good research results (Gruber et al., 2019). The vast majority of scholars use Kalman filter algorithm research object in soil moisture, for with the help of soil apparent conductivity data, the study of soil salinization results is less. Huang et al. (Huang et al., 2017a) have used electromagnetic conductivity imaging and ensemble Kalman filter to monitor and model soil water dynamics. The results of EnFK modeling (Lin's concordance = 0.89) were superior to the physical model, and superior or equivalent to the empirical model (Artificial Neural Networks) on loamy, clayey, and biphasic soil profiles. This shows the feasibility of combining soil apparent conductivity data with the ensemble Kalman algorithm. While the study area of this paper is arid saline soil and the soil salinity prediction model is a linear model, the Kalman filter algorithm is preferred to calibrate and validate the model to verify the feasibility of the scheme. The results showed that the Kalman filter method significantly improved the accuracy of the model. Although the corrected R^2values for June and October were slightly higher than those for July (0.59 and 0.60, respectively), the model showed the greatest improvement in July, with an 81.3% increase in R^2. However, the intense heat and high temperatures in July may have interfered with the sensitivity of the instrument, resulting in the worst model performance during this period, with an R^2 of only 0.58 and a maximum RMSE of 337.22 μS cm^{-1}. The model performance for the month of March was also higher than that for the month of October (0.59 and 0.60, respectively). In March, on the other hand, its coefficient of determination (R^2) reached 0.79, which is higher than the values of the other months, but the $R^2$of its linear model is 0.78, and the linear model calibrated by the Kalman filter algorithm is equivalent to the linear model. It can be seen that compared with the traditional model, the Kalman filter algorithm is able to dynamically adjust the model predictions so as to continuously improve the predicted values according to the observed data.

Although, the results of these studies show that the Kalman filter method algorithm can be an important tool to improve the accuracy of soil salinity prediction, especially when applied to the linear relationship between soil depth and salinity content. However, there are still some limitations of the research results. First, the study area of this method is extremely arid salinized soil. The feasibility of this idea needs to be further verified if the salinization degree of the soil is studied in a non-arid area. Secondly, the prerequisite salt prediction model using Kalman filter algorithm is a linear model, and once the model is nonlinear, the algorithm has to be re-selected for validation (Huang et al., 2017b). Therefore, future research will have to explore the selection of suitable data assimilation algorithms according to the actual situation.

4.3 Implications for cotton cultivation

Cotton production areas in South Xinjiang are deeply affected by soil salinization, which on the one hand will lead to soil loss of fertility, crop yield reduction; on the other hand, it will also have a great impact on the ecological environment (Li et al., 2014). Therefore, the monitoring, management and improvement of salinized farmland is particularly important, and how to obtain soil salinity information quickly, accurately and in real time is the basic premise of management and improvement of farmland. In this paper, Kalman filter algorithm is applied to the multivariate linear model of conductivity and soil depth, which greatly improves the

accuracy of the model. The relationship between soil profile salinity and depth was clearly established, and the model predictions were validated by comparing the measured and predicted conductivity values. For the spatial and temporal scale changes of salinity in the 0-100 cm soil layer of drip-irrigated cotton under the membrane at different times, the salinity in the 0-20 cm soil layer in the vertical direction at the inter-membrane site was less, and the salinity in the 60-100 cm layer was accumulated to a greater extent. Drip irrigation water washed the salts in all directions, and in the vertical direction, as the soil layer deepened, the washing effect gradually weakened, and the salts eventually accumulated in the deep soil layer. Although drip irrigation water has a significant leaching effect on salts, but this is a certain number of times and short-lived (Zhao et al., 2016). In contrast, under strong transpiration, the root system of cotton absorbs water continuously, and the root system absorbs water so that salts are transported to the root system. Although cotton can absorb and utilize some of the trace elements in brackish water, most of the salts remain in the root zone.

The numerical simulation results of this study showed that the 80-100 cm soil layer was relatively stable with a small range of salinity dynamics due to less water infiltration. In the 40-60 cm soil layer, the salinity is significantly higher than that in the 0-20 cm layer, and the deep soil water and salt have a tendency to migrate to the bare surface at the edge of the membrane due to evaporation and root absorption. In this study, we found that soil salts accumulated during the reproductive period of cotton fluctuated in a sawtooth pattern, and showed a gradual accumulation trend. This is due to the increase in soil water content after irrigation led to a transient decrease in the salinity of each soil layer, and then soil water gradually evaporated and dissipated as well as being absorbed and utilized by cotton, resulting in a gradual rebound of soil salinity (Liu et al., 2013); Drip irrigation is effective in reducing salinity in the upper soil layer (0-60 cm), but salts tend to accumulate at greater depths over time, suggesting the need for additional salt management practices such as deep plowing, winter irrigation, and drainage. Generally, no irrigation is required during the seedling stage, but when the relative soil moisture content is lower than 50%, small irrigation will be carried out; when the relative soil moisture content is lower than 60% during the bud stage, timely irrigation and topdressing of water-soluble fertilizers are required. Generally, irrigation is carried out 1-2 times during this period, and the irrigation cycle is 10-12 days; when the relative soil moisture content is lower than 70% during the flower and boll stage, timely irrigation and topdressing of flower and boll fertilizers are required. Generally, irrigation is carried out 8-10 times during this period, and the irrigation cycle is 7-10 days. After the full boll stage, the amount of irrigation gradually decreases, and irrigation will be stopped before defoliation. However, during the growth period, specific measures must be taken according to the specific situation. For example: if the temperature is high in autumn, the water stop time must be appropriately postponed (Datta et al., 2019). These measures can help prevent salt accumulation at deeper depths and support the sustainable cultivation of cotton in arid regions. And this study can better simulate the spatial and temporal distribution of soil salinity in drip-irrigated cotton fields and predict the yield and production of cotton fields, but in the future simulation process need to take into account the factors of soil evaporation, plant transpiration and root water uptake, so that the more reasonable optimization of the model parameters, the closer the simulation results are to the actual measurement values (Li et al., 2019).

Data availability statement

The datasets presented in this article are not readily available because Our data has taken a lot of time and effort, and there is still research going on, and the data is confidential. Requests to access the datasets should be directed to qingsongjiang219@126.com.

Author contributions

QJ: Writing – review & editing, Conceptualization, Methodology, Project administration, Resources, Supervision. YG: Writing – original draft. LC: Formal analysis, Methodology, Software, Writing – review & editing. MZ: Data curation, Investigation, Methodology, Writing – review & editing. QH: Validation, Visualization, Writing – review & editing. JH: Investigation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research was partially funded by [Chinese Universities Scientific Fund (Joint fund project of Nanjing Agricultural University and Tarim University)] grant number [NNLH202306]; [Fundamental Research Funds for the Central Universities] grant number [KYLH2023005] and Tarim University first-class course construction project (Mathematical Modeling) grant number [TDYLKC202430].

Conflict of interest

The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmed, R., Sreeram, V., Mishra, Y., and Arif, M.D. (2020). A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renewable Sustain. Energy Rev.* 124, 109792. doi: 10.1016/j.rser.2020.109792

Akram, M. A., Liu, P., Tahir, M. O., Ali, W., and Wang, Y. (2019). A state optimization model based on Kalman filtering and robust estimation theory for fusion of multi-source information in highly non-linear systems. *Sensors* 19, 1687. doi: 10.3390/s19071687

Buizza, C., Casas, C. Q., Nadler, P., Mack, J., Marrone, S., Titus, Z., et al. (2022). Data learning: Integrating data assimilation and machine learning. *J. Comput. Sci.* 58, 101525. doi: 10.1016/j.jocs.2021.101525

Corwin, D. L. (2021). Climate change impacts on soil salinity in agricultural areas. *Eur. J. Soil Sci.* 72, 842–862. doi: 10.1111/ejss.13010

Datta, A., Ullah, H., Ferdous, Z., Santiago-Arenas, R., and Attia, A. (2019). Water management in cotton. *Cotton production*, 47–59. doi: 10.1002/9781119385523.ch3

Debella-Gilo, M., and Gjertsen, A. K. (2021). Mapping seasonal agricultural land use types using deep learning on Sentinel-2 image time series. *Remote Sens.* 13, 289. doi: 10.3390/rs13020289

Desamsetti, S., Dasari, H. P., Langodan, S., Titi, E. S., Knio, O., and Hoteit, I. (2019). Efficient dynamical downscaling of general circulation models using continuous data assimilation. *Q. J. R. Meteorological Soc.* 145, 3175–3194. doi: 10.1002/qj.v145.724

El-Naggar, A. G., Hedley, C. B., Roudier, P., Horne, D., and Clothier, B. E. (2021). Imaging the electrical conductivity of the soil profile and its relationships to soil water patterns and drainage characteristics. *Precis. Agric.* 22, 1045–1066. doi: 10.1007/s11119-020-09763-x

Feng, J., Liu, H., Wang, G., Tian, R., Cao, M., Bai, Z., et al. (2021). Effect of periodic winter irrigation on salt distribution characteristics and cotton yield in drip irrigation under plastic film in Xinjiang. *Water* 13, 2545. doi: 10.3390/w13182545

Geer, A. J. (2021). Learning earth system models from observations: machine learning or data assimilation? *Philos. Trans. R. Soc. A.* 379, 20200089. doi: 10.1098/rsta.2020.0089

Gits-Muselli, M., White, P. L., Mengoli, C., Chen, S., Crowley, B., Dingemans, G., et al. (2020). The fungal PCR initiative's evaluation of in-house and commercial Pneumocystis jirovecii qPCR assays: toward a standard for a diagnostics assay. *Med. mycology* 58, 779–788. doi: 10.1093/mmy/myz115

Gorji, T., Yıldırım, A., Sertel, E., and Tanık, A. (2019). Remote sensing approaches and mapping methods for monitoring soil salinity under different climate regimes. *Int. J. Environ. Geoinformatics* 6, 33–49. doi: 10.30897/ijegeo.500452

Gruber, A., De Lannoy, G., and Crow, W. (2019). A Monte Carlo based adaptive Kalman filtering framework for soil moisture data assimilation. *Remote Sens. Environ.* 228, 105–114. doi: 10.1016/j.rse.2019.04.003

Huang, J., McBratney, A. B., Minasny, B., and Triantafilis, J. (2017a). Monitoring and modelling soil water dynamics using electromagnetic conductivity imaging and the ensemble Kalman filter. *Geoderma* 285, 76–93. doi: 10.1016/j.geoderma.2016.09.027

Huang, J., McBratney, A. B., Minasny, B., and Triantafilis, J. I. (2017b). 3D soil water nowcasting using electromagnetic conductivity imaging and the ensemble Kalman filter. *J. hydrology* 549, 62–78. doi: 10.1016/j.jhydrol.2017.03.067

Huang, Y., Zhang, Y., Zhao, Y., and Chambers, J. A. (2019). A novel robust Gaussian-Student's t mixture distribution based Kalman filter. *IEEE Trans. Signal Process.* 67, 3606–3620. doi: 10.1109/TSP.78

Huerta-Bátiz, H. E., Constantino-Recillas, D. E., Monsiváis-Huertero, A., Hernández-Sánchez, J. C., Judge, J., and Aparicio-García, R. S. (2022). Understanding root-zone soil moisture in agricultural regions of Central Mexico using the ensemble Kalman filter, satellite-derived information, and the THEXMEX-18 dataset. *Int. J. Digital Earth* 15, 52–78. doi: 10.1080/17538947.2021.2012534

Kalinitchenko, V. P., Glinushkin, A. P., Swidsinski, A. V., Minkina, T. M., Andreev, A. G., Mandzhieva, S. S., et al. (2021). Thermodynamic mathematical model of the Kastanozem complex and new principles of sustainable semiarid protective silviculture management. *Environ. Res.* 194, 110605. doi: 10.1016/j.envres.2020.110605

Kasim, N., Maihemuti, B., Sawut, R., Abliz, A., Dong, C., and Abdumutallip, M. (2020). Quantitative estimation of soil salinization in an arid region of the Keriya Oasis based on multidimensional modeling. *Water* 12, 880. doi: 10.3390/w12030880

Khodarahmi, M., and Maihami, V. (2023). A review on Kalman filter models. *Arch. Comput. Methods Eng.* 30, 727–747. doi: 10.1007/s11831-022-09815-7

Kivi, M. S., Blakely, B., Masters, M., Bernacchi, C. J., Miguez, F. E., and Dokoohaki, H. (2022). Development of a data-assimilation system to forecast agricultural systems: A case study of constraining soil water and soil nitrogen dynamics in the APSIM model. *Sci. Total Environ.* 820, 153192. doi: 10.1016/j.scitotenv.2022.153192

Lei, F., Crow, W. T., Kustas, W. P., Dong, J., Yang, Y., Knipper, K. R., et al. (2020). Data assimilation of high-resolution thermal and radar remote sensing retrievals for soil moisture monitoring in a drip-irrigated vineyard. *Remote Sens. Environ.* 239, 111622. doi: 10.1016/j.rse.2019.111622

Li, K., Liu, H., He, X., and Li, X. (2019). Simulation of water and salt transport in soil under pipe drainage and drip irrigation conditions in Xinjiang. *Water* 11, 2456. doi: 10.3390/w11122456

Li, H., Liu, X., Hu, B., Biswas, A., Jiang, Q., Liu, W., et al. (2020). Field-scale characterization of spatio-temporal variability of soil salinity in three dimensions. *Remote Sens.* 12, 4043. doi: 10.3390/rs12244043

Li, J., Pu, L., Han, M., Zhu, M., Zhang, R., and Xiang, Y. (2014). Soil salinization research in China: Advances and prospects. *J. Geographical Sci.* 24, 943–960. doi: 10.1007/s11442-014-1130-2

Liu, M., Yang, J., Li, X., Liu, G., Yu, M., and Wang, J. (2013). Distribution and dynamics of soil water and salt under different drip irrigation regimes in northwest China. *Irrigation Sci.* 31, 675–688. doi: 10.1007/s00271-012-0343-3

Lu, F. E. N. G., Chi, B. J., and Dong, H. Z. (2022) Cotton cultivation technology with Chinese characteristics has driven the 70-year development of cotton production in China. *J. Integr. Agric.* 21 (3), 597–609. doi: 10.1016/j.funbio.2018.11.007

Lu, Y., Chibarabada, T. P., Ziliani, M. G., Onema, J. M. K., McCabe, M. F., and Sheffield, J. (2021). Assimilation of soil moisture and canopy cover data improves maize simulation using an under-calibrated crop model. *Agric. Water Manage.* 252, 106884. doi: 10.1016/j.agwat.2021.106884

Mello, D. C. D., Safanelli, J. L., Poppiel, R. R., Veloso, G. V., Cabrero, D. R. O., Greschuk, L. T., et al. (2022). Soil apparent electrical conductivity survey in different pedoenvironments by geophysical sensor EM38: a potential tool in pedology and pedometry studies. *Geocarto Int.* 37, 13057–13078. doi: 10.1080/10106049.2022.2076913

Rahman, H., Carter, I., Basile, K., Donovan, L., Kumar, S., Tran, T., et al. (2020). Interpret with caution: an evaluation of the commercial AusDiagnostics versus in-house developed assays for the detection of SARS-CoV-2 virus. *J. Clin. Virol.* 127, 104374. doi: 10.1016/j.jcv.2020.104374

Reddy, T. A., and Henze, G. P. (2011). *Applied data analysis and modeling for energy engineers and scientists* (New York: Springer).

Ren, D., Wei, B., Xu, X., Engel, B., Li, G., Huang, Q., et al. (2019). Analyzing spatiotemporal characteristics of soil salinity in arid irrigated agro-ecosystems using integrated approaches. *Geoderma* 356, 113935. doi: 10.1016/j.geoderma.2019.113935

Rooij, R. V., and Schulz, K. (2020). Generics and typicality: a bounded rationality approach. *Linguistics Philosophy* 43, 83–117. doi: 10.1007/s10988-019-09265-8

Schmidt, J., and Bijmolt, T. H. (2020). Accurately measuring willingness to pay for consumer goods: a meta-analysis of the hypothetical bias. *J. Acad. Marketing Sci.* 48, 499–518. doi: 10.1007/s11747-019-00666-6

Song, J. H., Shi, X. Y., Cui, J., et al. (2019). Assessment of the accuracy of a soil salinity model for shallow groundwater areas in Xinjiang based on electromagnetic induction. *Appl. Ecol. Environ. Res.* 17 (4), 10037–10057. doi: 10.15666/aeer/1704_1003710057

Sun, Y., Ding, J., Qu, J., and Jiang, H. (2012). Saline soil identification based on electromagnetic induction and remote sensing in Xinjiang Oasis. *Trans. Chin. Soc. Agric. Eng.* 28, 180–187. doi: 10.3969/j.issn.1002-6819.2012.20.024

Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., et al. (2020). Fourier features let networks learn high frequency functions in low dimensional domains. *Adv. Neural Inf. Process. Syst.* 33, 7537–7547. doi: 10.48550/arXiv.2006.10739

Wang, F., Yang, S., Wei, Y., Shi, Q., and Ding, J. (2021). Characterizing soil salinity at multiple depth using electromagnetic induction and remote sensing data with random forests: A case study in Tarim River Basin of southern Xinjiang, China. *Sci. Total Environ.* 754, 142030. doi: 10.1016/j.scitotenv.2020.142030

Wang, Z., Zhang, X., Zhang, F., weng Chan, N., Liu, S., and Deng, L. (2020). Estimation of soil salt content using machine learning techniques based on remote-sensing fractional derivatives, a case study in the Ebinur Lake Wetland National Nature Reserve, Northwest China. *Ecol. Indic.* 119, 106869. doi: 10.1016/j.ecolind.2020.106869

Zare, E., Arshad, M., Zhao, D., Nachimuthu, G., and Triantafilis, J. (2020). Two-dimensional time-lapse imaging of soil wetting and drying cycle using EM38 data across a flood irrigation cotton field. *Agric. Water Manage.* 241, 106383. doi: 10.1016/j.agwat.2020.106383

Zhang, Y., Liao, Q. V., and Bellamy, R. K. (2020). "Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making," in *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 295–305.

Zhao, B., Wang, Z., and Li, W. (2016). Effects of winter drip irrigation mode and quota on water and salt distribution in cotton field soil and cotton growth next year in northern Xinjiang. *Trans. Chin. Soc. Agric. Eng.* 32, 139–148. doi: 10.11975/j.issn.1002-6819.2016.06.019

Zhao, C. W., Xu, L., Wang, Z. C., Wang, Y. H., and Chi, C. M. (2013). Studying soil salinity and its relations with microtopography and vegetation at field scale. *Polish J. Environ. Stud.* 22, 1227–1237. doi: 10.5555/20133303815

Zhou, B., Yang, L., Chen, X., Ye, S., Peng, Y., and Liang, C. (2021). Effect of magnetic water irrigation on the improvement of salinized soil and cotton growth in Xinjiang. *Agric. Water Manage.* 248, 106784. doi: 10.1016/j.agwat.2021.106784

Zhu, Y., Zheng, B., Luo, Q., Jiao, W., and Yang, Y. (2023). Uncovering the drivers and regional variability of cotton yield in China. *Agriculture* 13, 2132. doi: 10.3390/agriculture13112132

Ziliani, M. G., Altaf, M. U., Aragon, B., Houborg, R., Franz, T. E., and Lu, Y. (2022). Early season prediction of within-field crop yield variability by assimilating CubeSat data into a crop mode. *Agric. For. meteorology* 313, 108736. doi: 10.1016/j.agrformet.2021.108736

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 January 2025
DOI 10.3389/fpls.2024.1512660

Check for updates

OPEN ACCESS

EDITED BY
Liujun Xiao,
Nanjing Agricultural University, China

REVIEWED BY
Navod Neranjan Thilakarathne,
University of Colombo, Sri Lanka
Jiaxing Xie,
South China Agricultural University, China

*CORRESPONDENCE
Naisen Liu
boomzip@163.com

†These authors contributed equally to this work and share first authorship

RECEIVED 17 October 2024
ACCEPTED 19 December 2024
PUBLISHED 09 January 2025

CITATION
Liu N, Guo J, Liu F, Zha X, Cao J, Chen Y, Yan H, Du C, Wang X, Li J and Zhao Y (2025) Development of a vegetation canopy reflectance sensor and its diurnal applicability under clear sky conditions.
Front. Plant Sci. 15:1512660.
doi: 10.3389/fpls.2024.1512660

COPYRIGHT
© 2025 Liu, Guo, Liu, Zha, Cao, Chen, Yan, Du, Wang, Li and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Development of a vegetation canopy reflectance sensor and its diurnal applicability under clear sky conditions

Naisen Liu[1,2,3*†], Jingyu Guo[1†], Fuxia Liu[1,2,3], Xuedong Zha[4], Jing Cao[5], Yuezhen Chen[6], Haixia Yan[7], Chenggong Du[1,2,3], Xuqi Wang[1], Jiping Li[1,2,3] and Yongzhen Zhao[1]

[1]Jiangsu Collaborative Innovation Center of Regional Modern Agriculture & Environmental Protection, Huaiyin Normal University, Huai'an, China, [2]Jiangsu Key Laboratory for Eco-Agricultural Biotechnology around Hongze Lake, Huaiyin Normal University, Huai'an, China, [3]Jiangsu Engineering Research Center for Cyanophytes Forecast and Ecological Restoration of Hongze Lake, Huaiyin Normal University, Huai'an, China, [4]Huai'an Agricultural Information Center, Huai'an, China, [5]Jiangsu Academy of Agricultural Sciences, Wuxi, China, [6]Huai'an Institute of Vegetable Sciences, Huai'an, China, [7]Huai'an Agricultural Technology Extension Center, Huai'an, China

The spectral reflectance provides valuable information regarding vegetation growth and plays an important role in agriculture, forestry, and grassland management. In this study, a small, portable vegetation canopy reflectance (VCR) sensor that can operate throughout the day was developed. The sensor includes two optical bands at 710 nm and 870 nm, with the light separated by filters, and has a field of view of 28°. It is powered by two 14500 rechargeable batteries and uses Wi-Fi for data transmission. The calibration of the sensor was performed using an integrating sphere, and a solar altitude correction model was constructed. The sensor's accuracy was validated using a standard reflectance gray scale board. The results indicate that the root mean square error (RMSE) and mean absolute error (MAE) at 710 nm were 1.07% and 0.63%, respectively, while those at 870 nm were 0.94% and 0.50%, respectively. Vegetation at 14 sites was measured using both the VCR sensor and an Analytical Spectral Devices (ASD) spectroradiometer at nearly the same time for each site. The results show that the reflectance values measured by both devices were closely aligned. Measurements of Bermuda grass vegetation on clear days revealed that the intra-day reflectance range at 710 nm narrowed from 12.3–19.2% before solar altitude correction to 11.1–13.4% after correction, and the coefficient of variation (CV) decreased from 10.86% to 2.93%. Similarly, at 870 nm, the intra-day reflectance range decreased from 41.6–60.3% to 39.0–42.0%, and the CV decreased from 9.69% to 1.53%. In summary, this study offers a fundamental tool for monitoring vegetation canopy reflectance in the field, which is crucial for advancing high-quality agricultural, grassland, and forest management practices.

KEYWORDS

vegetation canopy reflectance, spectral reflectance, adaptability to varying light intensities, solar altitude correction, stability of full-daytime measurements

1 Introduction

The vegetation canopy reflectance can be used to obtain vegetation growth information such as the chlorophyll content, biomass, and leaf area index. Hence, it plays an important role in agriculture, forestry, and grassland management (Wang et al., 2012; Zulfa et al., 2020; Zhao et al., 2021; Galal et al., 2022). The reflectance spectrum of the vegetation canopy is influenced by factors such as the leaf area index, biomass, chlorophyll content, and nitrogen content. Therefore, by measuring the canopy's spectral reflectance in specific bands, valuable information about vegetation growth can be obtained. Instruments capable of collecting data over a wide spectral band have promise in advanced research on vegetation spectral monitoring, such as the FieldSpec spectroradiometers (Analytical Spectral Devices Inc., Boulder, CO, USA). The FieldSpec FR has a spectral range of 350–2500 nm, while the FieldSpec HandHeld has a spectral range of 325–1075 nm, and both devices have a spectral resolution of 1 nm. Due to its ability to capture detailed vegetation spectral characteristics, it has been widely used to monitor vegetation growth, including parameters such as the nitrogen content (Liu et al., 2020; Smith et al., 2020), biomass (Pradhan et al., 2014; Zhang et al., 2018), chlorophyll content (Li et al., 2020; Morley et al., 2020), and sugar content (Zhang et al., 2015). It has also been used to monitor the aging process (Anderegg et al., 2020), disease (Zhao et al., 2012; Martínez-Martínez et al., 2018), and water stress (Sönmez et al., 2008). The FieldSpec spectroradiometer uses sunlight as its light source and requires calibration with a white target board before measurement. To ensure accurate reflectance measurements, frequent calibration is necessary, especially when the light intensity changes rapidly, which can greatly reduce the working efficiency. Multispectral radiometers (MSR, Cropscan, Rochester, MN, USA) also use sunlight as the light source, but they can simultaneously measure the incident sunlight and canopy-reflected light to obtain the reflectance (Zhu et al., 2007; Gianquinto et al., 2011). The spectral range includes the visible and near-infrared bands. The MSR16R has 16 bands, which can be selected from 116 bands within the range of 460–1700 nm. The MSR87 has eight bands with central wavelengths of 460, 510, 560, 610, 660, 710, 760, and 810 nm. The MSR5 has five bands with central wavelengths of 485, 560, 660, 830, and 1650 nm. Although the FieldSpec spectroradiometer and MSR offer rich spectral bands, their high costs and large sizes make them unsuitable for use in agricultural production.

Based on previous research on spectral technologies (e.g., sensitive bands and inversion models) for monitoring vegetation growth, developing low-cost specialized sensors with only the necessary sensitive bands is of great significance for widespread application. Wang et al. (2004) developed a portable normalized difference vegetation index (NDVI) instrument for estimating the growth of winter wheat, utilizing spectral bands in the red and near-infrared regions. Measurements of winter wheat demonstrated that the NDVI values exhibited minimal variation around midday, making midday the optimal time for data collection. Cui et al. (2009) developed an optical sensor for measuring the crop leaf chlorophyll content. This sensor obtains the reflectance by measuring the light intensities of the canopy-reflected light and incident sunlight at 610 and 1220 nm, and it then calculates the NDVI based on the reflectance. Subsequently, the chlorophyll content is derived from the NDVI. The NDVI values of the white reference panel were measured from 10:00 to 15:40, and the results indicated that the NDVI values were relatively stable between 10:00 and 14:40, but they exhibited a sharp decline after 14:40. Ni et al. (2018) designed a portable apparatus for crop-growth monitoring and diagnosis (CGMD). It measures the spectral reflectance of rice and wheat canopies at 720 and 810 nm, and the leaf nitrogen content (LNC), leaf nitrogen accumulation (LNA), leaf area index (LAI), and leaf dry weight (LDW) can be calculated. The intensity of skylight varies significantly across different seasons and throughout the day, potentially impacting the accuracy of sensor measurements. However, for most existing sensors, there is a lack of studies on their adaptability to both strong and weak light conditions. Newly developed sensors require thorough evaluation of their all-day measurement accuracies to ensure their suitability for practical applications in production.

Differences in ambient light can affect the accuracy of spectral sensors that use sunlight as a light source (Shibayama and Wiegand, 1985). The position of the sun, primarily the solar altitude or zenith angle, affects ambient light, especially the angle of the sunlight. Throughout the day, the solar altitude first increases, reaches a peak at noon, and then slowly decreases. The solar zenith angle, which is complementary to the solar altitude, exhibits the opposite trend. When the sensor's optical window is facing downward (i.e., view zenith = 0°) to receive reflected light from the ground, the measured reflectance is called the nadir reflectance, and the reflectance measured at other view zeniths is called the off-nadir reflectance. In most practical applications, the sensor measures the nadir reflectance. Rijks (1967) measured the all-day nadir reflectance of cotton on multiple dates and found that the nadir reflectance decreased as the solar altitude increased, that is, a U-shaped diurnal variation pattern with the smallest values observed at noon. Davies and Buttimor (1969) reported similar diurnal variations in the reflectance of peppers, tobacco, perennial ryegrass, tomatoes, winter wheat, and corn. Proctor et al. (1972) showed that the diurnal variations in the reflectance of apple trees depended on the solar zenith angle, but the seasonal variations were small. Similar findings have been reported in other studies (Ranson et al., 1985; Shikada and Miyakita, 1992), indicating that the solar altitude impacts the vegetation reflectance. Kimes et al. (1980) determined the relationship between the reflectance and the vegetation canopy characteristics and solar zenith angle. Ranson et al. (1985) attributed such reflectance variations to specular reflection caused by direct sunlight. The influence of solar altitude on the measurement accuracy of sensors poses a significant challenge for all-day vegetation spectral reflectance monitoring.

To reduce the influence of the solar altitude on the spectral reflectance, the optical window of the sensor is often equipped with a cosine corrector. The light passing through the cosine corrector is diffuse and non-directional. The cosine corrector is often constructed from opal glass or polytetrafluoroethylene (Ni et al., 2018; Liu et al., 2022; Bai et al., 2023; Heusinkveld et al., 2023). However, assessments of the extent to which a cosine corrector can reduce the impact of the solar altitude angle on reflectance

measurements, especially throughout the day, are extremely scarce. To minimize the impact of the solar altitude on vegetation canopy reflectance measurements, data collection is typically performed around midday (e.g., 10:00–14:00 or 11:00–13:00) when changes in the solar altitude are minimal (Tian et al., 2005; Huang et al., 2007; Zhang et al., 2018; Anderegg et al., 2020; Chandel et al., 2021). However, this approach reduces the data collection time each day, and the high temperatures around noon in summer are not suitable for fieldwork.

To address the aforementioned typical issues, in this study, we developed a novel portable sensor for measuring the vegetation canopy reflectance (VCR). The sensor features a photodetection circuit with 250 gain levels, providing a wide dynamic range that effectively adapts to variations in the ambient light intensity. Additionally, the sensor is equipped with a resistance calibration circuit, ensuring a high reflectance measurement accuracy. A solar elevation correction model was constructed based on the characteristics of optical transmission, enabling the sensor to operate reliably throughout the day under clear weather conditions.

2 Materials and methods

In this section, we detail the design process of the VCR sensor, as well as the experiments conducted to evaluate its accuracy and stability in measuring vegetation spectral reflectance.

2.1 Design of the VCR sensor

The VCR sensor consists of mechanical components, optical systems, hardware circuits, power supplies, and software systems; the overall structure of the sensor is illustrated in Figure 1A. The sensor uses Wi-Fi for data transmission.

2.1.1 Optical sensing unit

The spectral reflectance of the vegetation canopy is defined as the ratio of the light energy reflected by the canopy in a specific band to the incident light energy in that band. The up-facing optical sensing unit of the VCR sensor captures incident light, while the down-facing optical sensing unit detects light reflected from the vegetation canopy. Given that the vegetation canopy receives light from the entire sky, the up-facing optical sensing unit has a 180° FOV. On sunny days, the light from the sky predominantly consists of direct sunlight (Liu and Jordan, 1960; Sen, 2004). The solar altitude influences the transmittance of the direct sunlight entering the sensor, thereby affecting the accuracy of the reflectance measurements. The optical window of the up-facing optical sensing unit is equipped with a cosine corrector made of polytetrafluoroethylene (PTFE), which could reduce the impact of the solar altitude on the measurements. Moreover, the cosine corrector ensures that the up-facing optical sensing unit maintains a 180°FOV. Additionally, an attenuator, a filter, and a photodetector are installed below the cosine corrector (Figure 1B). The down-facing optical sensing unit is equipped with filters and photodetectors from bottom to top. The FOV significantly influences the area of vegetation measured by the sensor when capturing spectral information, thereby affecting the measurement accuracy. A larger FOV covers a wider vegetation area, enabling the spectral information to better represent the overall vegetation conditions. However, in areas with incomplete vegetation coverage, particularly for row-planted crops, an excessively large FOV may capture spectral information about both the vegetation and soil, leading to inaccurate measurements. Conversely, a smaller FOV can effectively reduce or eliminate soil interference, but it may result in measurements that are not representative of the vegetation. Therefore, the FOV must be carefully balanced. Currently, an FOV of approximately 30° is commonly used. The VCR sensor is designed with an FOV of 28°. A schematic diagram of the FOV is shown in Figure 1C.

The filters used are FB710-10 and FB870-10 (Thorlabs Inc., Newton, NJ, USA), which have wavelengths of 710 nm and 870 nm, respectively, and a full width at half-maximum (FWHM) of 10 nm. The photodetector converts the optical signal into an electrical signal, and its spectral responsivity determines the performance of the sensor. The VCR sensor uses OPT101 photodiodes (Texas Instruments, Dallas, TX, USA), which have a spectral responsivity of 0.5–0.6 A/W at 710 nm and 870 nm. The photosensitive surface

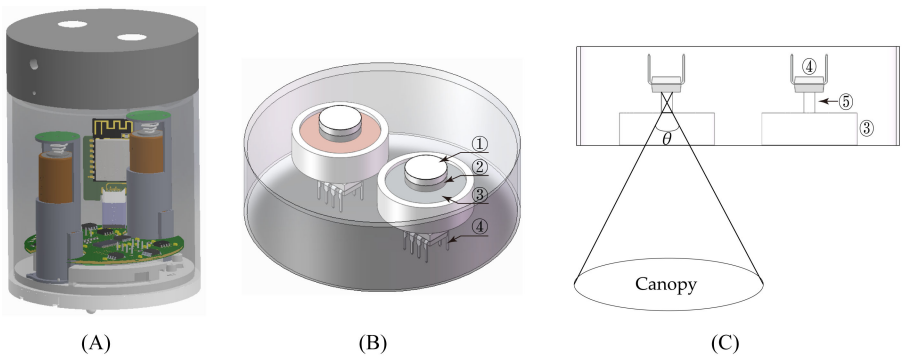


FIGURE 1
Schematic diagrams of **(A)** the VCR sensor, **(B)** the up-facing optical sensing unit, and **(C)** the down-facing optical sensing unit with the field of view. ① Cosine corrector, ② Attenuator, ③ Filter, ④ Photodetector, and ⑤ Optical channel.

of the OPT101 measures 2.29 mm × 2.29 mm. The OPT101 photodiode operates in photoconductive mode, resutling in excellent linearity and a low dark current.

2.1.2 Design of the hardware circuit

The sensor's circuitry primarily consists of a power module, optical signal acquisition module, calibration module for the amplifier circuit resistance, and wireless communication module. The power module, wireless communication module, and up-facing optical signal acquisition module each has its own printed circuit board (PCB), while the down-facing optical signal acquisition module and microcontroller unit (MCU) are on the same PCB. This split design facilitates sensor installation and maintenance.

The power module utilizes two 3.7 V 14500 batteries connected in parallel to supply 3.3 V and −3.3 V power to the system. The OPT101 photodetector has an integrated operational amplifier that enables current-to-voltage (I–V) conversion, with outputs from pin 5. Different gains can be achieved by connecting an external feedback resistor between pins 2 and 5. The VCR sensor is intended for year-round operation. Given that the ambient light intensity varies significantly across the seasons, it is necessary to dynamically adjust the gain so as to obtain an appropriate voltage. The VCR sensor employs a T-type feedback network to achieve a high amplification factor using three small resistors (Figure 2A). R1 and R2 are precision resistors with a resistance of 68 kΩ and an accuracy of 0.01%. R3 is replaced by a digital potentiometer TPL0102 (Texas Instruments, Dallas, TX, USA), which has 256 taps and a resistance range of 0–100 kΩ. A larger resistor number indicates a higher resistance value. A smaller resistance value of R3 means a greater gain of the OPT101 (Equation 1). The accuracy of R3 directly affects the gain of the OPT101 and consequently the accuracy of the sensor's reflectance measurement. Therefore, it is essential to calibrate the different resistance values of the TPL0102. Specifically, the TPL0102 is connected to the feedback network and a constant current source through an analog multiplexer, namely, ADG774 (Analog Devices, MA, USA), to achieve resistance calibration (Figure 2A). Due to the limited FOV and weak light reflected from the canopy, the optical signal received by the down-facing optical sensing unit is very weak and requires a higher gain. Based on preliminary testing under both strong and low-light conditions, the required gain ranges approximately from 1 M to 6 M. Hence, only resistors 1–25 of the TPL0102 are calibrated. The up-facing optical sensing unit receives strong light from the sky, so it requires only a small amount of amplification. The minimum required gain is approximately 0.2 M, as determined through initial testing. Hence, resistors 1–250 are calibrated. In the down-facing optical signal acquisition circuit, the TPL0102 is connected to a 100 μA constant current source, while the up-facing TPL0102 is connected to a 10 μA constant current source. The 100 μA current is provided by REF200 (Texas Instruments, Dallas, TX, USA), and the 10 μA current is provided by REF200 and OPA128 (Texas Instruments, Dallas, TX, USA) (Figure 2B). The on-resistance of ADG774 is 2.2 Ω, which is minimal and has a negligible impact on the other circuits during operation. The MCU uses the 16-bit MSP430I2041 (Texas Instruments, Dallas, TX, USA) for signal acquisition control, data calculation, and communication. The MSP430I2041 features a low supply voltage (2.2–3.6 V), ultra-low power consumption, five power-saving modes, and four 24-bit analog-to-digital converters (ADCs), making it suitable for portable devices and equipment used in field operations. The VCR sensor utilizes the MSP430I2041's four built-in ADCs to convert analog electrical signals, which represent the light intensity, into digital signals. The sensor uses the ESP-12S (Ai-thinker Technology Co., Shenzhen, China) for wireless communication.

$$G = 68 + 68 + \frac{68 \times 68}{R} \tag{1}$$

where G is the amplification (K); and R is the resistance value after calibration of the TPL0102 (kΩ). The calibration method is described in Section 2.2.1.

2.1.3 Design of the VCR sensor control system

The embedded software system of the VCR sensor handles the measurement and communication tasks. The software was developed using the Contiki embedded operating system, which is event-driven and supports multithreading. The Contiki requires minimal resources during the run period, making it highly suitable for sensor applications (Dunkels et al., 2004). Upon powering on, the VCR sensor initializes, and then, it enters a sleep state and waits for events to occur. The events and processes of this system are shown in Figure 3. The events include *event_received_information,*

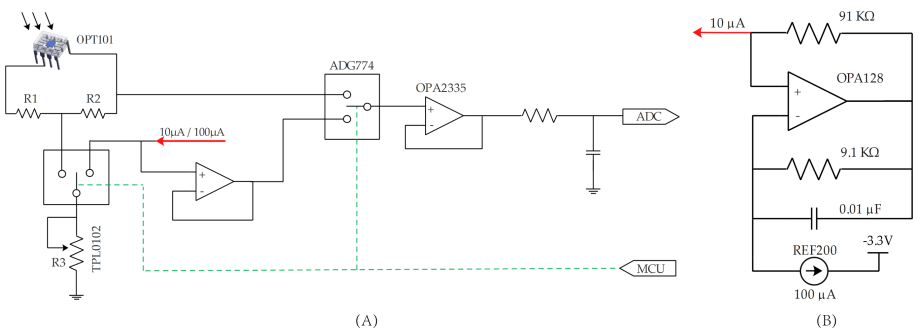


FIGURE 2
(A) Schematic diagram of the optical signal acquisition and TPL0102 resistor calibration circuit, and **(B)** schematic diagram of the 10 μA circuit.

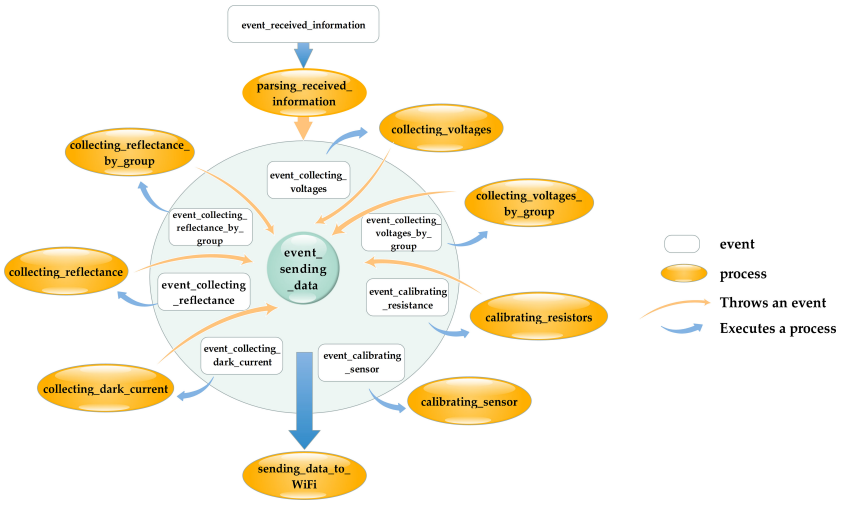


FIGURE 3
Schematic diagram of the system events and processes.

event_collecting_reflectance, event_collecting_reflectance_by_group, event_collecting_voltages, event_collecting_voltages_by_group, event_calibrating_resistance, event_calibrating_sensor, event_collecting_dark_current, and *event_sending_data*. The system processes include *parsing_received_information, collecting_reflectance, collecting_reflectance_by_group, collecting_voltages, collecting_voltages_by_group, calibrating_resistors, calibrating_sensor, collecting_dark_current*, and *sending_data_to_WiFi*.

The user group for this sensor includes staff who measure the spectral reflectance of the vegetation canopies and scientific researchers. To facilitate their work, a command set was designed, which includes the following commands: "COLLECT_Ref" for collecting the reflectance, "COLLECT_Ref Repeat[]" for collecting the reflectance by group, "COLLECT_Voltages" for collecting the voltages of the optical channels, "COLLECT_Voltages Repeat[]" for collecting the voltages of optical channels by group, "Calibrate_Resistors" for calibrating the resistors of TPL0102, "Calibrate_Sensor" for calibrating the reflectance of the sensor, and "Dark_Current" for collecting each channel's dark current. The sensor's temperature is also recorded during each measurement. The commands "COLLECT_Ref Repeat[]" and "COLLECT_Voltages Repeat[]" enable multiple measurements to be obtained using a single command. The brackets can be filled with multiple numbers separated by commas, each of which represents the number of sampling times during the measurement. For example, [1,5,10] means three measurements are required: the first is a single measurement, the second measurement is taken five times consecutively and the average is taken as the result, and the third measurement is taken ten times consecutively with the average as the result. Similarly, [5,5,5,5] indicates four measurements, each taken five times consecutively, and the averages are taken as the results. This type of command enables repeated measurements within a short time frame, which is particularly useful for studying the impact of the wind speed on the vegetation canopy reflectance.

Devices such as mobile phones and computers can connect to the sensor via Wi-Fi and control it by sending commands, and the sensor will send back the measured data. Upon receiving a command, the *parsing_received_information* process parses the content and triggers the corresponding event based on the command. Once the event is triggered, the corresponding process will run. For measurement commands, the *event_sending_data* event is triggered after the data are measured, and the *sending_data_to_WiFi* process sends the data back.

Figure 4 illustrates the processes for measuring the reflectance or voltages, as well as for measuring the reflectance or voltages by group. When measuring the reflectance or voltages (Figure 4A), the sensor first measures the voltage (V_{710U}) of the 710 nm up-facing optical sensing unit. If the voltage is less than 300 mV, the resistance of the TPL0102 is reduced to increase the gain of the OPT101. If the voltage exceeds 900 mV, the gain of the OPT101 is reduced. After multiple measurements, the final voltage (V_{710U}) is $300mV \leq V_{710U} \leq 900mV$. Subsequently, the gains of the 870 nm up-facing optical sensing unit and the 710 nm and 870 nm down-facing optical units are adjusted. The voltages of all four optical units are then re-measured. If reflectance measurements are required, the reflectance is calculated. Finally the *event_sending_data* event is triggered.

When measuring by group (Figure 4B), the gains of all of the OPT101 optical units are first adjusted using the method illustrated in Figure 4A. After adjustment, the first number (n_1) from the sequence [n_1, n_2, …, n_i, …] is selected and V_{710U}, V_{870U}, V_{710D}, and V_{870D} are measured successively n_1 times. The average voltages or reflectance values are calculated and recorded as the result. Next, the second number (n_2) is taken from the sequence, and V_{710U}, V_{870U}, V_{710D}, and V_{870D} are measured n_2 times. The average voltages or reflectance values are calculated and recorded again. This process is repeated until all of the numbers in the sequence are utilized. Finally, the *event_sending_data* event is triggered.

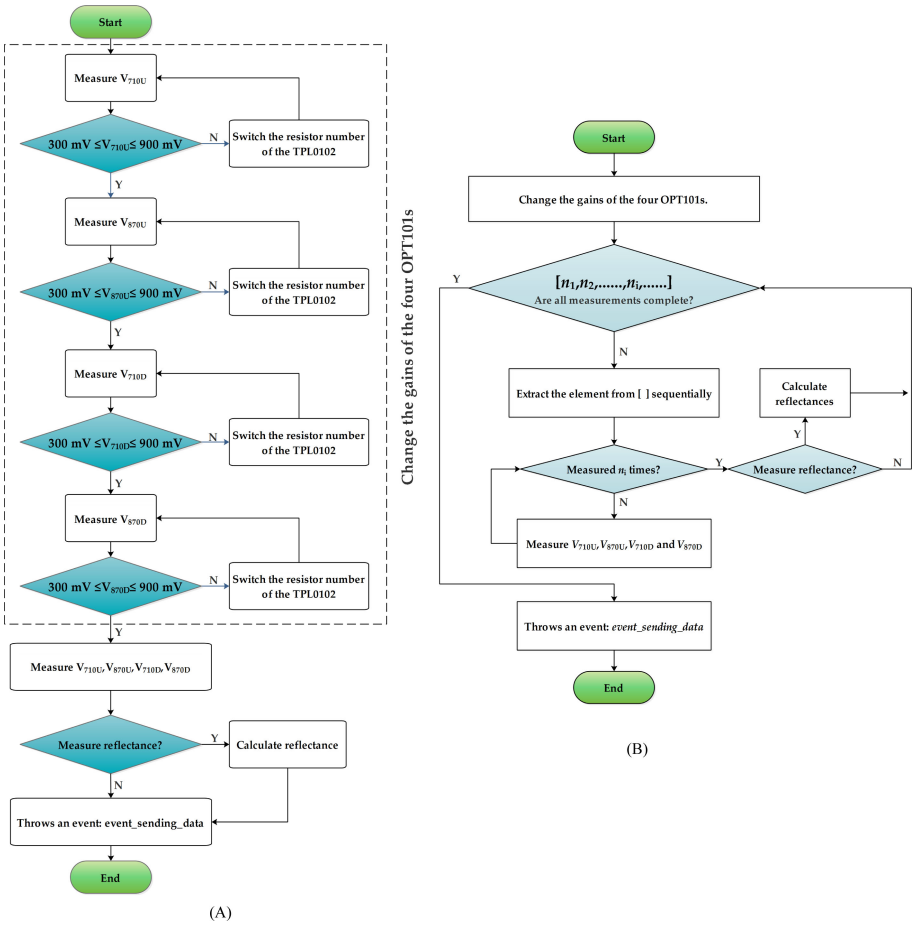


FIGURE 4
Procedure **(A)** for reflectance value and voltage measurement, and **(B)** that for reflectance value and voltage measurement by group.

2.2 Research on calibration methods for the VCR sensor

The gain and dark current of each optical unit of the sensor affect the accuracy of the measurement, so the resistance of the TPL0102 potentiometer, the dark current of the OPT101, and the reflectivity of the sensor are calibrated.

2.2.1 Calibration of the resistance values of the TPL0102

As can be seen from Equation 1, the accuracy of the TPL0102 resistance directly influences the gain of the OPT101. During calibration, the "Calibrate_Resistors" command was sent to the VCR sensor. The MCU then controlled the ADG774 (Figure 2A) to connect the TPL0102 chips within the 710 nm and 870 nm down-facing optical unit to a 100 μA current for calibration of resistors 1–25. The calibrated resistance value is denoted as R_D. Similarly, the ADG774 was controlled to connected the TPL0102 chips within the 710 nm and 870 nm up-facing optical unit to a 10 μA current for calibration of resistors 1–250. The calibrated resistance value is denoted as R_U. The calculation method for the resistance values is shown in Equations 2, 3.

$$R_U = \frac{V}{10\mu A} = 10V \tag{2}$$

$$R_D = \frac{V}{10\mu A} = V \tag{3}$$

where V is the output voltage of the sensor (0.1 mV).

2.2.2 Calibration of dark currents of the OPT101 in each optical unit

The VCR sensor was placed in a black, light-proof box. The command "Dark_Current" was sent to the sensor. The MCU controlled the ADG774 to connect the TPL0102 to the OPT101 circuit, and the resistor number of TPL0102 switched to adjust the gain of the OPT101. The dark current, expressed as the measured voltage, of each optical unit of the sensor at different gains was then measured.

2.2.3 Calibration of the sensor reflectance

The sensor's down-facing optical unit detects the light reflected from vegetation and converts it into voltage signal U_D, while the up-facing optical unit detects incident light from the sky and converts it

into voltage signal U_U. The reflectance is calculated as the ratio $\rho = \frac{U_D}{U_U}$. Calibration can be performed on the raw reflectance values, as $\rho = f\left(\frac{U_D}{U_U}\right)$, where $f()$ is calibration function. Alternatively, the voltages U_D and U_U can be calibrated separately, yielding a reflectance relationship of $\rho = \frac{f_1(U_D)}{f_2(U_U)}$, where $f_1()$ and $f_2()$ are calibration functions. It is also possible to calibrate only one of the voltage signals, resulting in reflectance relationships such as $\rho = \frac{f_1(U_D)}{U_U}$ or $\rho = \frac{U_D}{f_2(U_U)}$.

An integrating sphere was used to calibrate the VCR sensor's output voltages, thereby calibrating the reflectance values. By adjusting the voltage of the integrating sphere's light source, different light intensities were achieved. Under consistent light intensity conditions, the up-facing and down-facing optical windows of the sensor were aligned with the light outlet of the integrating sphere. The output voltage of the sensor was then measured, and the dark current was subtracted. Since the light intensity is uniform, the voltage of the down-facing optical unit in the same band should match that of the up-facing optical unit. Based on this principle, a voltage calibration equation $f()$ was constructed to calibrate the reflectance. The reflectance was calibrated using Equation 4.

$$\rho_\lambda = \frac{f(U_D/G_D)}{U_U/G_U} \times 100\,\% \tag{4}$$

where ρ_λ is the reflectance at wavelength λ, U_D and G_D are respectively the measured voltage and gain of the down-facing optical unit at wavelength λ, U_U and G_U are the measured voltage and gain, respectively, of the up-facing optical unit at the same wavelength.

The accuracy of the reflectance measurement of the sensor was verified using a gray scale board with a reflectance of 30%. The experiment was conducted on the roof of the School of Life Sciences building at Huaiyin Normal University on August 4, 2024. The sensor was positioned approximately 17 cm above the 20 cm × 20 cm gray scale board. If the distance was greater than 17 cm, the sensor's FOV would have extended beyond the gray scale board.

2.3 Experimental design for performance testing of the VCR sensor

2.3.1 Accuracy testing experiment using standard reflectance gray boards

To verify the accuracy of the VCR sensor's reflectance measurements, a total of seven gray scale boards were used as the measurement targets. The nominal reflectance values of the gray calibration panels were 10%, 20%, 25%, 40%, 60%, 80%, and 99%. However, the reflectance of the same gray panel varies across different spectral bands. According to the datasheets provided by the manufacturer, the reflectance values of the seven gray scale boards at the 710 nm wavelength are 11.5%, 18.8%, 24.3%, 40.2%, 61.5%, 79.2%, and 98.9%. At the 870 nm wavelength, the reflectance values are 11.8%, 18.0%, 23.0%, 38.3%, 61.2%, 79.1%, and 98.7%. The experiment was conducted from 14:57 to 15:06 on August 5, 2024, under clear sky conditions, during which the solar altitude decreased from 47.9° to 46.3°. The instrument was located on the roof of the School of Life Sciences building at Huaiyin Normal University. Each measurement was repeated ten times.

2.3.2 Accuracy testing experiment using vegetation as the measurement subject

On a clear day, the VCR sensor and the ASD Fieldspec 4 spectroradiometer (ASD Inc., Alexandria, VA, USA) were used to simultaneously measure different vegetation canopies to further verify the accuracy of the VCR sensor's vegetation reflectance measurements. A total of 14 vegetation sites were selected. Sites 1, 2, 4, 8, 10, 11, 12, and 14 contained *Imperata cylindrica* in varying growth stages or densities. Sites 3 and 9 consisted of *Goosegrass*, site 5 contained *Setaria viridis*, site 6 contained *Tall Fescue*, and sites 7 and 13 contained *Clover*. The FOV of the VCR sensor and the ASD spectroradiometer were 28° and 25°, respectively. To ensure consistent canopy coverage in both measurements, the VCR sensor and the ASD spectroradiometer's fiber optic probe were located 40 cm and 48 cm above the measured vegetation, respectively. Given that the FWHM of the VCR sensor filter is 10 nm, the reflectance of the 705–715 nm band measured by the ASD spectroradiometer was averaged to represent the reflectance at 710 nm, and the average of the reflectance in the 865–875 nm band was taken as the reflectance at 870 nm.

2.3.3 Accuracy testing for all-day measurement

To verify the accuracy of the sensor for full-day measurements, the spectral reflectance of Bermuda grass was measured using the VCR sensor from 8:00 to 17:00 under clear sky conditions (with occasional clouds after 15:00) (Figure 5). The solar altitude ranged from 23.0° to 74.5°. The VCR sensor was positioned 50 cm above the vegetation, and measurements were taken once per minute while recording the solar altitude.



FIGURE 5
Photograph of the reflectance measurements conducted using the VCR sensor.

3 Results and discussion

3.1 Analysis of calibration results for the VCR sensor

3.1.1 Calibration analysis of TPL0102 resistance values

The calibration results of the resistance values of the TPL0102s are shown in Figure 6. As can be seen, the resistance values of the TPL0102 in all of the optical units have a strong linear relationship with the resistor number, with R^2 values reaching 0.99999.

3.1.2 Calibration analysis of OPT101 dark currents in each optical unit

The calibration results of the dark currents of the OPT101 in each optical unit are shown in Figure 7. As illustrated, there is a strong linear relationship between the dark current and the gain of the circuit, with an R^2 value of 0.9999.

3.1.3 Calibration analysis of the sensor reflectances

We used an integrating sphere to calibrate the sensor's output voltages, thereby achieving reflectance calibration for the VCR sensor. The results of the voltage calibration are shown in Figure 8. As can be seen, the down-facing output voltage and the up-facing output voltage at the same wavelength exhibit a strong linear relationship, with R^2 values of 0.9998 at 710 nm and 0.9999 at 870 nm.

3.1.4 Influence of solar elevation on reflectance measurements by the VCR sensor and the construction of its correction model

A 30% reflectance gray panel was used as the monitoring target to examine the variations in the reflectance measured by the calibrated VCR sensor throughout the day. Due to limitations imposed by the distance between the sensor and the gray panel, the sensor's shadow entered the FOV of the 710 nm optical unit between 11:00 and 13:40, while its shadow entered the FOV of the 870 nm optical unit between 10:30 and 13:00. Consequently, these data points were removed during the preprocessing. The results are shown in Figure 9. As shown, the reflectance measurements show a distinct trend throughout the day, with lower values occurring closer to noon (Figures 9A, B). Specifically, the reflectance decreases nonlinearly as the solar elevation angle increases (Figures 9C, D). The experimental results indicate that the cosine corrector alone cannot fully eliminate the nonlinear influence of the solar elevation on reflectance measurements made by spectral sensors. Therefore, to improve measurement accuracy, it is necessary to develop a solar elevation correction model.

Sky light can be categorized into two main types: sky-scattered light and direct sunlight. Sky-scattered light is diffuse and not affected by the solar elevation angle. As the solar elevation angle increases, the angle of the direct sunlight entering the sensor's optical window decreases, allowing more light to pass into the sensor. Under clear weather conditions, the proportion of the direct sunlight in the total sky light can reach 80–90%. However, when the solar elevation angle is zero (the Sun at the horizon), all of the light in the sky is diffuse, with no direct sunlight, making the proportion

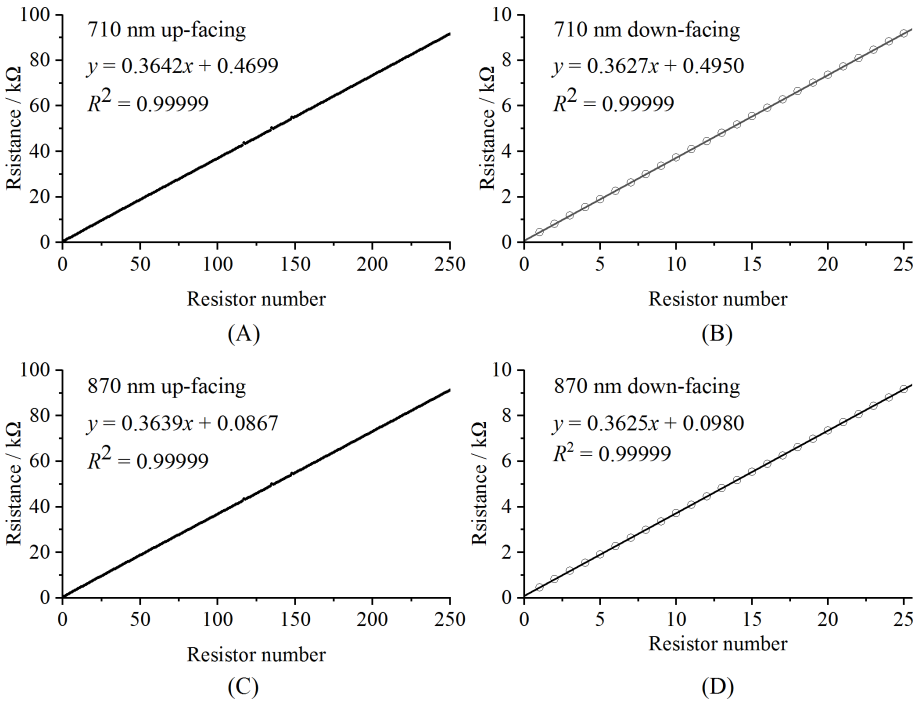


FIGURE 6
Calibration curves of the TPL0102 for the different optical units: **(A)** 710 nm up-facing optical unit, **(B)** 710 nm down-facing optical unit, **(C)** 870 nm up-facing optical unit, and **(D)** 870 nm down-facing optical unit.

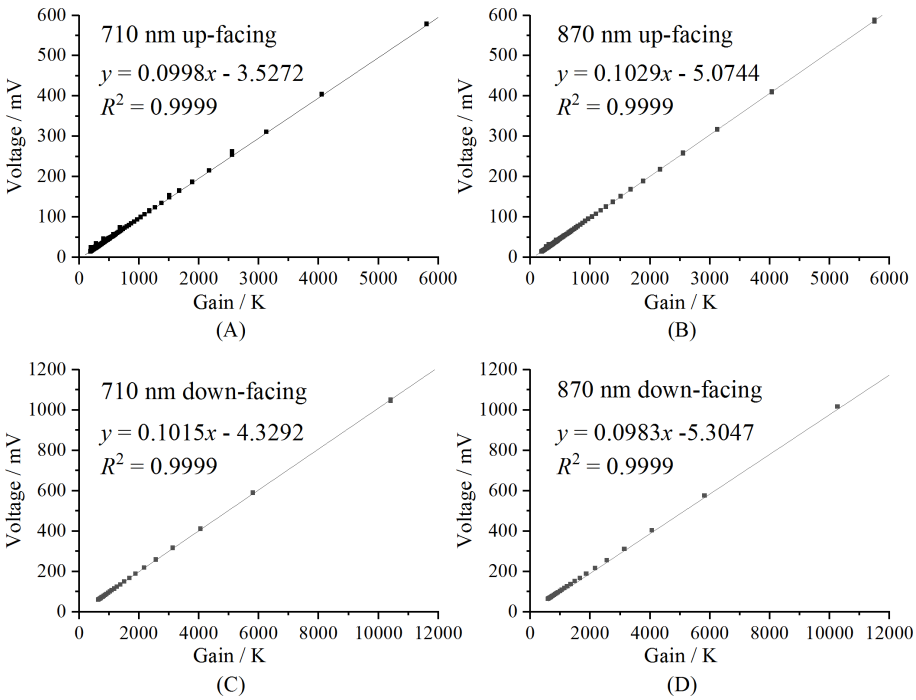


FIGURE 7
Calibration curves of the voltage generated by the dark current for each optical unit: **(A)** 710 nm up-facing optical unit, **(B)** 870 nm up-facing optical unit, **(C)** 710 nm down-facing optical unit, and **(D)** 870 nm down-facing optical unit.

of direct light 0%. To model the variations in the proportions of direct and diffuse light with changes in the solar elevation, a sine function can be used. To ensure that the proportion of direct sunlight is 90% at a solar elevation angle of 90° and 0% at an elevation of 0°, the proportion of direct sunlight (PD) can be expressed as $PD = \sin(0.7128\theta)$, where θ is the solar elevation angle. The corresponding proportion of diffuse light (PS) is expressed as $PS = 1 - \sin(0.7128\theta)$. This model effectively captures the gradual increase in direct sunlight as the sun rises and the predominance of diffuse light when the sun is low. The voltage value measured by the sensor's up-facing optical unit (U_U), which represents the light intensity, can be separated into two components: the voltage converted from the diffuse light, expressed as $U_U \cdot PS$ and the voltage converted from the direct sunlight, expressed as $U_U \cdot PD$. The direct sunlight can be decomposed into two vectors. One portion follows a vertically downward path, perpendicular to the optical window of the sensor. This portion of the light enters the sensor, and the measured voltage $U_U \cdot PD$ is generated from this light. The other part of the sunlight travels horizontally, parallel to the sensor's optical window, and does not enter the sensor. Based on the vector decomposition, the intensity of the direct sunlight before decomposition can be determined using the formula $\frac{U_U \cdot PD}{\sin\theta}$. Thus, by taking Equation 4 into consideration, the following solar elevation correction model for reflectance can be derived (Equation 5).

$$\rho_{\lambda,\theta} = \frac{f(U_D/G_D)}{\frac{U_U \cdot PS}{G_U} + \frac{U_U \cdot PD}{G_U} \cdot \frac{1}{\sin\theta}} = \rho_\lambda \cdot \frac{1}{1 - \sin(0.7128\theta) + \frac{\sin(0.7128\theta)}{\sin\theta}} \quad (5)$$

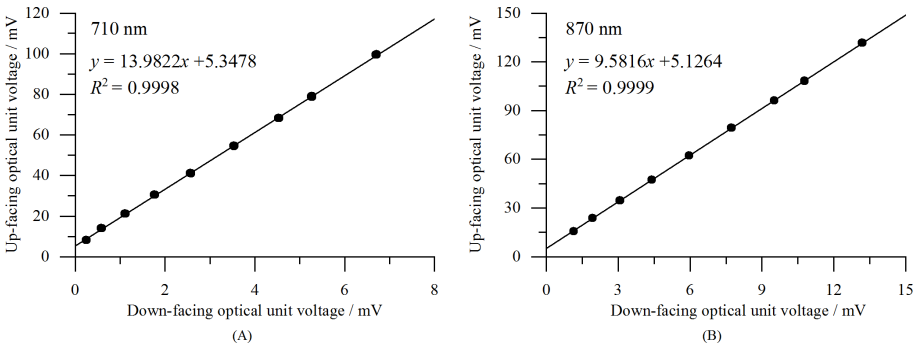


FIGURE 8
Calibration curves of the output voltages for the down-facing optical units with a gain of 1 K: **(A)** 710 nm, and **(B)** 870 nm.

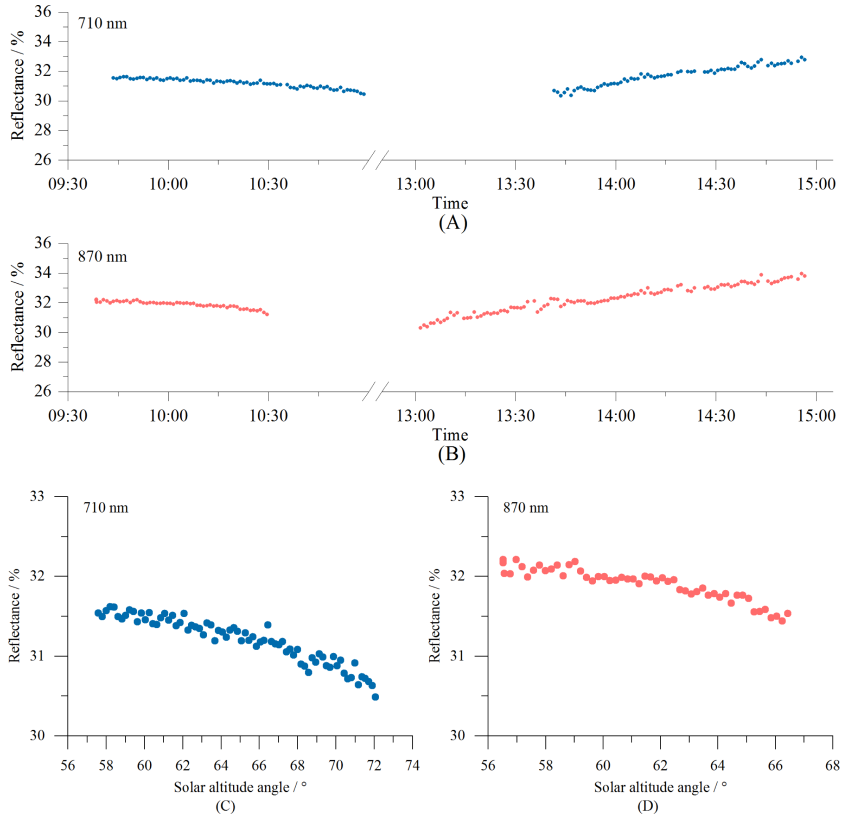


FIGURE 9
Diurnal variations in the reflectance at 710 nm **(A)** and 870 nm **(B)** measured by the VCR sensor using a 30% reflectance gray scale board. Reflectance changes at 710 nm **(C)** and 870 nm **(D)** as a function of the solar elevation angle.

where $\rho_{\lambda,\theta}$ is the reflectance in wavelength λ when the solar elevation is θ (°). The meanings of the other symbols are the same as in Equation 4.

The reflectance measured using the VCR sensor can be corrected using Equation 5 to reduce the impact of the solar altitude on the measurements, enabling all-day operation of the sensor. The corrected results for the data presented in Figure 9 are shown in Figure 10. It is evident that after correction, the diurnal variations in the reflectance at 710 nm and 870 nm are mostly stable, indicating an excellent correction effect.

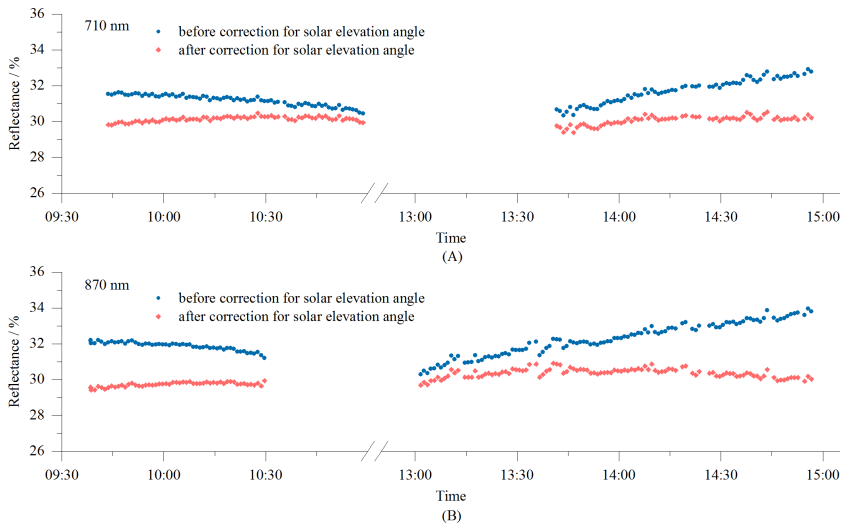


FIGURE 10
Diurnal variations in reflectance measured by the VCR sensor before and after solar altitude correction: **(A)** 710 nm, and **(B)** 870 nm.

3.2 Analysis of accuracy and stability

3.2.1 Accuracy of standard reflectance gray board measurements

The accuracy of the VCR sensor measurements was tested using seven gray boards with known reflectance values. The root mean square error (RMSE), mean absolute error (MAE), and 1:1 plot of the measured versus standard reflectance were used to assess the accuracy of the measurements. The results shown in Figure 11 indicate that the measured reflectance values closely matched those of the gray scale boards. The MAE and RMSE for the 710 nm measurements were 1.07% and 0.63%, respectively, and those for the 870 nm measurements were 0.94% and 0.50%, respectively, demonstrating a high measurement accuracy.

3.2.2 Accuracy of vegetation reflectance measurements

The VCR sensor and ASD spectrometer were used to simultaneously measure the spectral reflectance of vegetation canopies at 14 sites. The results are shown in Figure 12. The reflectance values at 710 nm and 870 nm measured by the VCR sensor closely matched those measured by the ASD spectroradiometer. The slight deviations observed in some measurements may be due to inconsistency in the vegetation coverage within the fields of view of the VCR sensor and the ASD spectroradiometer.

3.2.3 Accuracy and stability of all-day measurements of vegetation reflectance

The VCR sensor was used to measure the reflectance of Bermuda grass throughout the day. The measured reflectance values were corrected for the solar elevation using Equation 5. The coefficients of variation (CVs) of the reflectances at 710 nm and 870 nm were calculated both before and after the solar altitude correction. The results are shown in Figure 13. As can be seen, after correcting for the solar altitude the reflectance measurements remained relatively stable, effectively eliminating the influence of the solar altitude on the results. The all-day reflectance at 710 nm changed from 12.3–19.2% to 11.1–13.4%, and the CV decreased from 10.86% before correction to 2.93% after correction. For the 870 nm all-day reflectance, the values changed from 41.6–60.3% to 39.0–42.0%, and the CV decreased from 9.69% before correction to 1.53% after correction. It can also be seen from the graph that after 16:00, when the solar elevation angle dropped below 23°, the effectiveness of the solar elevation correction model in mitigating the impact of the solar elevation on the measurements began to decline.

3.3 Discussion

The spectral reflectance is crucial for monitoring vegetation growth and has been extensively utilized in various fields (Pfitzner et al., 2022; Yu et al., 2022). However, the light intensity in different environments varies significantly. For instance, sun-loving plants thrive in stronger light, while shade-loving plants prefer weaker light. Even for the same plant, the light intensity can vary greatly throughout the year and throughout the day. The variability of the light intensity requires the sensor to have a wide light signal gain range. In this study, precision resistors and TPL0102 potentiometers were used to construct the gain circuit of the photodetector, achieving a gain range of approximately 0.2 M to 10 M. Due to the insufficient accuracy of the potentiometer resistance values, calibration was necessary to enhance the sensor's accuracy. In this study, a 10 μA constant current source was used to calibrate the TPL0102 for each up-facing optical unit, while a 100 μA constant current source was used for the down-facing optical units. The calibration exhibited excellent linearity, with R^2 values reaching 0.99999. Based on the gain of each optical unit, the dark current model achieved excellent linearity, with an R^2 value of 0.9999. During the measurements, for the same wavelength, the up-facing and down-facing optical sensing units had different gain values. After subtracting the dark current, the calculated reflectance was more accurate and reasonable (Ni et al., 2018). Otherwise, nonlinear changes would occur.

Reflectance measurements of vegetation using spectrometers or spectral sensors are mostly conducted around noon (Tian et al., 2005; Huang et al., 2007; Chandel et al., 2021), such as from 11:00 to 13:00 or from 10:00 to 14:00. During this period, due to the small changes in the solar altitude, the measured reflectance is relatively stable. However, this limits the operation hours and reduces the efficiency. Additionally, the high temperatures at noon in summer

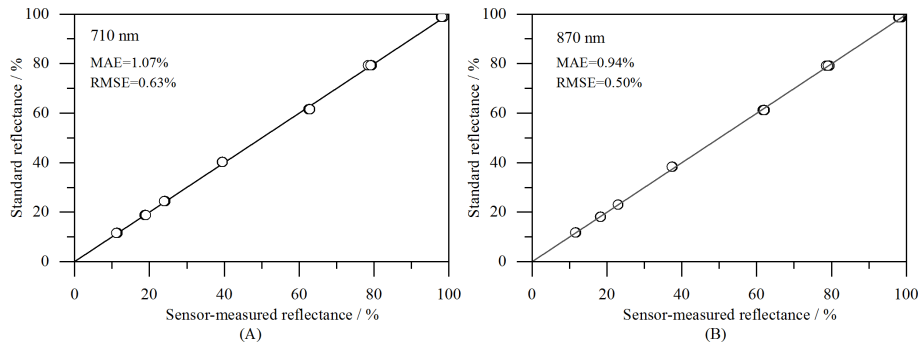


FIGURE 11
Comparison of measured reflectance and standard reflectance: **(A)** 710 nm, and **(B)** 870 nm.

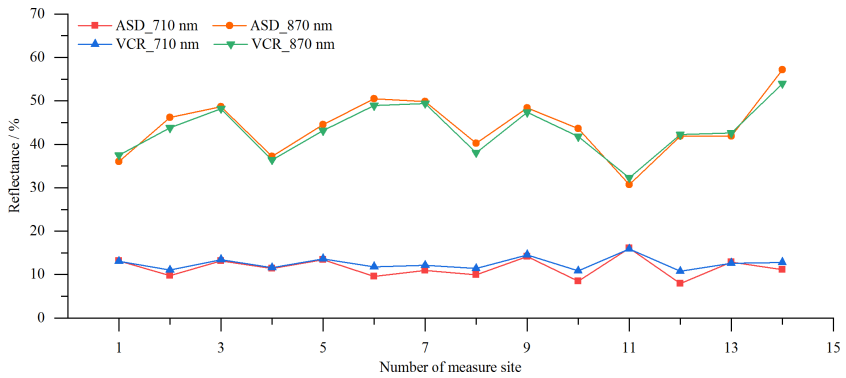


FIGURE 12
Reflectances of different vegetation canopies measured by the VCR sensor and ASD spectrometer.

make fieldwork challenging. Therefore, developing reflectance measurement technology that enables sensors to operate throughout the day is crucial. Like other similar sensors, the VCR sensor is equipped with cosine correctors. However, experimental data reveal that the solar altitude still affects the accuracy of the reflectance measurements (Figure 9). Specifically, the sensor's measured reflectance is a function of both the vegetation canopy and solar altitude, with the lowest reflectance occurring at noon and higher reflectance occurring in the morning and afternoon, resulting in a U-shaped pattern throughout the day. This observation is consistent with previous research findings (Rijks, 1967; Davies and Buttimor, 1969). A plausible explanation for this is that variations in the solar elevation alter the angle of incidence of the direct sunlight on the sensor's optical window, leading to differences in the cosine corrector's light transmission efficiency across different solar angles. In this study, we developed a solar altitude correction model and verified the model using Bermuda grass as the test subject. The results demonstrate that after correction, the all-day reflectance measured became much more stable. The CV of the reflectance at 710 nm decreased from 10.86% to 2.93%, while the CV at 870 nm decreased from 9.69% to 1.53%, indicating that the VCR sensor accurately measured the reflectance throughout the day under clear sky conditions. In future research, we will test the sensor on more vegetation to further verify the sensor's performance.

The characteristics of the VCR sensor developed in this study, along with those of other comparable sensors, are summarized in Table 1.

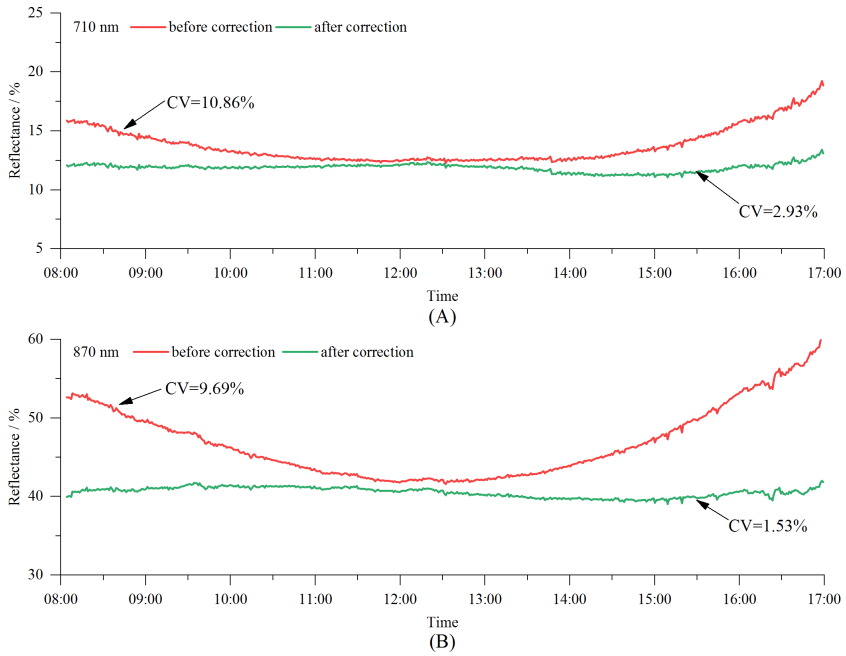


FIGURE 13
Diurnal variations in spectral reflectance of Bermuda grass measured by the VCR sensor before and after solar altitude correction: **(A)** 710 nm, and **(B)** 870 nm.

TABLE 1 The characteristics of the selected sensors.

Sensor/Spectrometer	Bands	Size	Portable	Capable of all-day measurements	Capable of batch measurements	Application Field
Sensor developed in this study	710 and 870 nm	71 × 71 × 99 mm	Yes	Yes	Yes	Scientific research/ industrial application
CGMD (Ni et al., 2018)	720 and 810 nm	44 × 44 × 52 mm	Yes	Not reported	No	Scientific research/ industrial application
Optical sensor for monitoring chlorophyll content (Cui et al., 2009)	610 and 1220 nm	105 × 54 × 124 mm	Yes	No	No	Scientific research/ industrial application
FieldSpec FR spectroradiometer	350–2500 nm	127 × 368 × 292 mm	No	Yes Frequent calibration using a white reference panel is required	No	Scientific research

4 Conclusions

In this study, we successfully developed a vegetation canopy reflectance sensor that operates at spectral bands of 710 nm and 870 nm, has a 28°FOV and is powered by two 14500 rechargeable batteries, making it highly suitable for field applications. The sensor features Wi-Fi data transmission and a photoelectric conversion circuit with a gain range of 0.2 M to 10 M, ensuring stable and accurate measurements under varying light conditions. A solar altitude correction model was constructed to further enhance the measurement accuracy throughout the day. The sensor's performance was validated using seven standard reflectance gray scale boards, and MAEs of 1.07% and 0.94%, and RMSEs of 0.63% and 0.50% were achieved for 710 nm and 870 nm, respectively. Field experiments conducted at 14 vegetation sites demonstrated a close agreement between the reflectance values obtained by the VCR sensor and an ASD spectroradiometer in both spectral bands. Additionally, the VCR sensor's ability to measure the reflectance throughout the day was assessed using Bermuda grass on a clear day (08:00–17:00; solar altitudes of 23.0–74.5°). After applying the solar altitude correction, the reflectance range at 710 nm narrowed from 12.3–19.2% to 11.1–13.4%, and the CV decreased from 10.86% to 2.93%. At 870 nm, the reflectance range narrowed from 41.6–60.3% to 39.0–42.0%, and the CV decreased from 9.69% to 1.53%. In summary, this research provides a foundational tool for obtaining vegetation growth information, with potential applications in improving the efficiency of agricultural and grassland production.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

NL: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. JG: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. FL: Formal analysis, Methodology, Validation, Writing – original draft, Writing – review & editing. XZ: Formal analysis, Methodology, Visualization, Writing – original draft. JC: Formal analysis, Validation, Visualization, Writing – original draft. YC: Investigation, Validation, Writing – original draft. HY: Investigation, Validation, Writing – original draft. CD: Formal analysis, Software, Writing – original draft. XW: Investigation, Writing – original draft. JL: Visualization, Writing – original draft. YZ: Investigation, Software, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by Jiangsu Provincial Social Development - Major Scientific and Technological Demonstration Project (no. BE2018678) and the Jiangsu Provincial Graduate Research and Practice Innovation Plan Project (no. SJCX24_2119 and no. SJCX24_2125).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Anderegg, J., Yu, K., Aasen, H., Walter, A., Liebisch, F., and Hund, A. (2020). Spectral vegetation indices to track senescence dynamics in diverse wheat germplasm. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01749

Bai, G., Ge, Y., Leavitt, B., Gamon, J. A., and Scoby, D. (2023). Goniometer in the air: enabling BRDF measurement of crop canopies using a cable-suspended plant phenotyping platform. *Biosyst. Eng.* 230, 344–360. doi: 10.1016/j.biosystemseng.2023.04.017

Chandel, N. S., Rajwade, Y. A., Golhani, K., Tiwari, P. S., Dubey, K., and Jat, D. (2021). Canopy spectral reflectance for crop water stress assessment in wheat (*triticum aestivum*, l.)*. *Irrig. Drain.* 70, 321–331. doi: 10.1002/ird.2546

Cui, D., Li, M. Z., and Zhang, Q. (2009). Development of an optical sensor for crop leaf chlorophyll content detection. *Comput. Electron. Agric.* 69, 171–176. doi: 10.1016/j.compag.2009.08.001

Davies, J. A., and Buttimor, P. H. (1969). Reflection coefficients, heating coefficients and net radiation at simcoe, southern ontario. *Agric. Met.* 6, 373–386. doi: 10.1016/0002-1571(69)90028-4

Dunkels, A., Gronvall, B., and Voigt, T. (2004). "Contiki-a lightweight and flexible operating system for tiny networked sensors," in *29th Annual IEEE Conference on Local Computer Networks* (Tampa, FL: IEEE).

Galal, H., Elsayed, S., Allam, A., and Farouk, M. (2022). Indirect quantitative analysis of biochemical parameters in banana using spectral reflectance indices combined with machine learning modeling. *Horticulturae* 8, 438. doi: 10.3390/horticulturae8050438

Gianquinto, G., Orsini, F., Fecondini, M., Mezzetti, M., Sambo, P., and Bona, S. (2011). A methodological approach for defining spectral indices for assessing tomato nitrogen status and yield. *Eur. J. Agron.* 35, 135–143. doi: 10.1016/j.eja.2011.05.005

Heusinkveld, B. G., Mol, W. B., and van Heerwaarden, C. C. (2023). A new accurate low-cost instrument for fast synchronized spatial measurements of light spectra. *Atmos. Meas. Tech.* 16, 3767–3785. doi: 10.5194/amt-16-3767-2023

Huang, W., Lamb, D. W., Niu, Z., Zhang, Y., Liu, L., and Wang, J. (2007). Identification of yellow rust in wheat using *in-situ* spectral reflectance measurements and airborne hyperspectral imaging. *Precis. Agric.* 8, 187–197. doi: 10.1007/s11119-007-9038-9

Kimes, D. S., Smith, J. A., and Ranson, K. J. (1980). Vegetation reflectance measurements as a function of solar zenith angle. *Photogramm. Eng. Remote Sens.* 46, 1563–1573.

Li, D., Chen, J. M., Zhang, X., Yan, Y., Zhu, J., Zheng, H., et al. (2020). Improved estimation of leaf chlorophyll content of row crops from canopy reflectance spectra through minimizing canopy structural effects and optimizing off-noon observation time. *Remote Sens. Environ.* 248, 111985. doi: 10.1016/j.rse.2020.111985

Liu, H., Bruning, B., Garnett, T., and Berger, B. (2020). The performances of hyperspectral sensors for proximal sensing of nitrogen levels in wheat. *Sensors* 20, 4550. doi: 10.3390/s20164550

Liu, B. Y., and Jordan, R. C. (1960). The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Sol. Energy* 4, 1–19. doi: 10.1016/0038-092X(60)90062-1

Liu, N., Zhang, W., Liu, F., Zhang, M., Du, C., Sun, C., et al. (2022). Development of a crop spectral reflectance sensor. *Agronomy* 12, 2139. doi: 10.3390/agronomy12092139

Martínez-Martínez, V., Gomez-Gil, J., MaChado, M. L., and Pinto, F. A. (2018). Leaf and canopy reflectance spectrometry applied to the estimation of angular leaf spot disease severity of common bean crops. *PloS One* 13, e0196072. doi: 10.1371/journal.pone.0196072

Morley, P. J., Jump, A. S., West, M. D., and Donoghue, D. N. (2020). Spectral response of chlorophyll content during leaf senescence in european beech trees. *Environ. Res. Commun.* 2, 71002. doi: 10.1088/2515-7620/aba7a0

Ni, J., Zhang, J. C., Wu, R. S., Pang, F. R., and Zhu, Y. (2018). Development of an apparatus for crop-growth monitoring and diagnosis. *Sensors* 18, 3129. doi: 10.3390/s18093129

Pfitzner, K., Bartolo, R., Whiteside, T., Loewensteiner, D., and Esparon, A. (2022). Multi-temporal spectral reflectance of tropical savanna understorey species and implications for hyperspectral remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* 112, 102870. doi: 10.1016/j.jag.2022.102870

Pradhan, S., Bandyopadhyay, K. K., Sahoo, R. N., Sehgal, V. K., Singh, R., Gupta, V. K., et al. (2014). Predicting wheat grain and biomass yield using canopy reflectance of booting stage. *J. Indian Soc Remote Sens.* 42, 711–718. doi: 10.1007/s12524-014-0372-x

Proctor, J. T. A., Kyle, W. J., and Davies, J. A. (1972). The radiation balance of an apple tree. *Can. J. Bot.* 50, 1731–1740. doi: 10.1139/b72-214

Ranson, K. J., Daughtry, C., Biehl, L. L., and Bauer, M. E. (1985). Sun-view angle effects on reflectance factors of corn canopies. *Remote Sens. Environ.* 18, 147–161. doi: 10.1016/0034-4257(85)90045-8

Rijks, D. A. (1967). Water use by irrigated cotton in Sudan. I. Reflection of short-wave radiation. *J. Appl. Ecol.* 4, 561–568. doi: 10.2307/2401357

Sen, Z. (2004). Solar energy in progress and future research trends. *Prog. Energy Combust. Sci.* 30, 367–416. doi: 10.1016/j.pecs.2004.02.004

Shibayama, M., and Wiegand, C. L. (1985). View azimuth and zenith, and solar angle effects on wheat canopy reflectance. *Remote Sens. Environ.* 18, 91–103. doi: 10.1016/0034-4257(85)90040-9

Shikada, M., and Miyakita, K. (1992). Effects of solar and view angles on reflectance for paddy field canopies. *Geocarto. Int.* 7, 9–17. doi: 10.1080/10106049209354375

Smith, C., Karunaratne, S., Badenhorst, P., Cogan, N., Spangenberg, G., and Smith, K. (2020). Machine learning algorithms to predict forage nutritive value of in *situ* perennial ryegrass plants using hyperspectral canopy reflectance. *Remote Sens.* 12, 928. doi: 10.3390/rs12060928

Sönmez, N. K., Emekli, Y., Sari, M., and Bastug, R. (2008). Relationship between spectral reflectance and water stress conditions of Bermuda grass (*cynodon dactylon* l.). *N. Z. J. Agric. Res.* 51, 223–233. doi: 10.1080/00288230809510451

Tian, Y., Zhu, Y., and Cao, W. (2005). Monitoring leaf photosynthesis with canopy spectral reflectance in rice. *Photosynthetica* 43, 481–489. doi: 10.1007/s11099-005-0078-y

Wang, W., Yao, X., Yao, X. F., Tian, Y. C., Liu, X. J., Ni, J., et al. (2012). Estimating leaf nitrogen concentration with three-band vegetation indices in rice and wheat. *Field Crops Res.* 129, 90–98. doi: 10.1016/j.fcr.2012.01.014

Wang, X., Zhao, C. J., Zhou, H. C., Liu, L. Y., Wang, J. H., Xue, X. Z., et al. (2004). Development and experiment of portable NDVI instrument for estimating growth condition of winter wheat. *Trans. Chin. Soc Agric. Eng.* 20, 95–98.

Yu, H., Kong, B., Hou, Y. T., Xu, X. Y., Chen, T., and Liu, X. M. (2022). A critical review on applications of hyperspectral remote sensing in crop monitoring. *Exp. Agr.* 58, e26. doi: 10.1017/S0014479722000278

Zhang, Y., Qin, Q. M., Ren, H. Z., Sun, Y. H., Li, M. Z., Zhang, T. Y., et al. (2018). Optimal hyperspectral characteristics determination for winter wheat yield prediction. *Remote Sens.* 10, 2015. doi: 10.3390/rs10122015

Zhang, Y., Zheng, L. H., Li, M. Z., Deng, X. L., and Ji, R. H. (2015). Predicting apple sugar content based on spectral characteristics of apple tree leaf in different phenological phases. *Comput. Electron. Agric.* 112, 20–27. doi: 10.1016/j.compag.2015.01.006

Zhao, Y. J., Sun, Y. H., Chen, W. H., Zhao, Y. P., Liu, X. L., and Bai, Y. F. (2021). The potential of mapping grassland plant diversity with the links among spectral diversity, functional trait diversity, and species diversity. *Remote Sens.* 13, 3034. doi: 10.3390/rs13153034

Zhao, J., Zhang, D., Luo, J., Dong, Y., Yang, H., and Huang, W. (2012). Characterization of the rice canopy infested with brown spot disease using field hyperspectral data. *Wuhan. Univ. J. Nat. Sci.* 17, 86–92. doi: 10.1007/s11859-012-0809-4

Zhu, Y., Zhou, D. Q., Yao, X., Tian, Y. C., and Cao, W. X. (2007). Quantitative relationships of leaf nitrogen status to canopy spectral reflectance in rice. *Aust. J. Agr. Res.* 58, 1077–1085. doi: 10.1071/AR06413

Zulfa, A. W., Norizah, K., Hamdan, O., Zulkifly, S., Faridah-Hanum, I., and Rhyma, P. P. (2020). Discriminating trees species from the relationship between spectral reflectance and chlorophyll contents of mangrove forest in Malaysia. *Ecol. Indic.* 111, 106024. doi: 10.1016/j.ecolind.2019.106024

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 January 2025
DOI 10.3389/fpls.2024.1492059

Check for updates

OPEN ACCESS

EDITED BY
Liujun Xiao,
Nanjing Agricultural University, China

REVIEWED BY
Rui Jiang,
South China Agricultural University, China
Xiaohe Gu,
Beijing Research Center for Information Technology in Agriculture, China

*CORRESPONDENCE
Meichen Feng
fmc101@sxau.edu.cn

RECEIVED 06 September 2024
ACCEPTED 19 December 2024
PUBLISHED 14 January 2025

CITATION
Yang C, Feng M, Bai J, Sun H, Bi R, Song L, Wang C, Zhao Y, Yang W, Xiao L, Zhang M and Song X (2025) Hyperspectral estimation of chlorophyll density in winter wheat using fractional-order derivative combined with machine learning.
Front. Plant Sci. 15:1492059.
doi: 10.3389/fpls.2024.1492059

COPYRIGHT
© 2025 Yang, Feng, Bai, Sun, Bi, Song, Wang, Zhao, Yang, Xiao, Zhang and Song. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Hyperspectral estimation of chlorophyll density in winter wheat using fractional-order derivative combined with machine learning

Chenbo Yang[1,2], Meichen Feng[1]*, Juan Bai[1], Hui Sun[3], Rutian Bi[2], Lifang Song[1], Chao Wang[1], Yu Zhao[1], Wude Yang[1], Lujie Xiao[1], Meijun Zhang[1] and Xiaoyan Song[1]

[1]College of Agriculture, Shanxi Agricultural University, Taigu, Shanxi, China, [2]College of Resources and Environment, Shanxi Agricultural University, Taigu, Shanxi, China, [3]Life Sciences Department, Yuncheng University, Yuncheng, Shanxi, China

Chlorophyll density (ChD) can reflect the photosynthetic capacity of the winter wheat population, therefore achieving real-time non-destructive monitoring of ChD in winter wheat is of great significance for evaluating the growth status of winter wheat. Derivative preprocessing has a wide range of applications in the hyperspectral monitoring of winter wheat chlorophyll. In order to research the role of fractional-order derivative (FOD) in the hyperspectral monitoring model of ChD, this study based on an irrigation experiment of winter wheat to obtain ChD and canopy hyperspectral reflectance. The original spectral reflectance curves were preprocessed using 3 FOD methods: Grünwald-Letnikov (GL), Riemann-Liouville (RL), and Caputo. Hyperspectral monitoring models for winter wheat ChD were constructed using 8 machine learning algorithms, including partial least squares regression, support vector regression, multi-layer perceptron regression, random forest regression, extra-trees regression (ETsR), decision tree regression, K-nearest neighbors regression, and gaussian process regression, based on the full spectrum band and the band selected by competitive adaptive reweighted sampling (CARS). The main results were as follows: For the 3 types of FOD, GL-FOD was suitable for analyzing the change process of the original spectral curve towards the integer-order derivative spectral curve. RL-FOD was suitable for constructing the hyperspectral monitoring model of winter wheat ChD. Caputo-FOD was not suitable for hyperspectral research due to its insensitivity to changes in order. The 3 FOD calculation methods could all improve the correlation between the original spectral curve and Log(ChD) to varying degrees, but only the GL method and RL method could observe the change process of correlation with order changes, and the shorter the wavelength, the smaller the order, and the higher the correlation. The bands screened by CARS were distributed throughout the entire spectral range, but there was a relatively concentrated distribution in the visible light region. Among all models, CARS was used to screen bands based on the 0.3-order RL-FOD spectrum, and the model constructed using ETsR reached the best accuracy and stability. Its R^{2c}, $RMSE_c$, R^{2v}, $RMSE_v$, and RPD were 1.0000, 0.0000, 0.8667, 0.1732, and 2.6660, respectively. In conclusion, based on the winter wheat ChD data set and the corresponding canopy hyperspectral

data set, combined with 3 FOD calculation methods, 1 band screening method, and 8 modeling algorithms, this study constructed hyperspectral monitoring models for winter wheat ChD. The results showed that based on the 0.3-order RL-FOD, combined with the CARS screening band, ETsR modeling has the highest accuracy, and hyperspectral estimation of winter wheat ChD can be realized. The results of this study can provide some reference for the rapid and nondestructive estimation of ChD in winter wheat.

KEYWORDS

hyperspectral, chlorophyll density, fractional-order derivative, competitive adaptive reweighted sampling, machine learning

1 Introduction

Chlorophyll is an important pigment for photosynthesis in most plants, so it plays a vital role in the growth and development of plants (Xu et al., 2001; Chatterjee and Kundu, 2015). Wheat is an important food crop, and about one-third of the world's population eats wheat as a staple food (Jiang et al., 2021). However, during its growth process, it is easily affected by factors such as water, fertilizer, and diseases, which hinder the synthesis of chlorophyll, thereby affecting photosynthesis and leading to reduced yield (He et al., 2018; Yang et al., 2021). Chlorophyll density (ChD) is an indicator that can evaluate the chlorophyll content of crop populations. During the growth process of wheat, obtaining ChD can be used to evaluate the overall growth status of wheat and provide a reference for adjusting agricultural production measures (Liu et al., 2017; Xie et al., 2020). However, traditional methods for obtaining ChD commonly suffer from issues such as destructive sampling, complex measurement processes, and unable to obtain data in real-time. Therefore, there is an urgent need for a fast, real-time, and non-destructive method to obtain wheat ChD to meet the current needs of precision agriculture development.

Remote sensing technology has the advantage of quick and non-destructive acquisition of target object features, providing technical support for real-time acquisition of crop growth status. Especially hyperspectral remote sensing technology, with its advantages of high resolution and large spectral information, has been used by many scholars to quantitatively monitor wheat chlorophyll. For example, Li et al. (2022); Zhang et al. (2022), and Jiang et al. (2010) all performed 1.0-order derivative preprocessing on the original spectral curve, and used multiple vegetation indices combined with linear, exponential, and power regression models to monitor wheat chlorophyll content and ChD, respectively, all achieved good monitoring results. Huang et al. (2010) performed multiple scattering correction, 1.0-order derivative, and 2.0-order derivative preprocessing on spectral curve, and constructed monitoring models of wheat chlorophyll content using partial least squares regression (PLSR) and stepwise multiple linear regression. The results showed that the stepwise multiple linear regression model combining multiple scattering correction with 2.0-order derivative reached the best performance. It can be seen that previous scholars have conducted relatively mature research on quantitative monitoring of wheat chlorophyll using hyperspectral technology.

Through analysis of previous research, it can be seen that in the process of constructing hyperspectral monitoring models of winter wheat chlorophyll content or ChD, previous scholars preprocessed the original spectral curve to varying degrees, such as derivatives, to reduce noise interference, and to reduce model complexity by constructing vegetation indices. Finally, by comparing the accuracy of various modeling methods, to achieve rapid estimation of winter wheat chlorophyll. It can be seen that appropriate preprocessing, bands, and modeling algorithms are all important means to improve the accuracy of hyperspectral models.

For preprocessing, derivative preprocessing was widely used in hyperspectral studies of other crops and growth physiological parameters, and was considered to have good effects (Ji et al., 2022; Liu et al., 2022; Yang et al., 2023b). However, these studies all used integer-order derivatives, but there is a significant difference between the original spectral curve and the integer-order derivative spectral curve. This means that there may be curves with higher signal-to-noise ratios during the transition from the original spectral curve to the integer-order derivative curve. In order to find this curve, some studies used FOD to further analyze the role of derivative preprocessing in hyperspectral monitoring models of crop growth physiological parameters. For example, Li et al (2021a; 2021b; 2023). applied FOD to hyperspectral monitoring of wheat leaf moisture content, leaf area index, and chlorophyll content, Bhadra et al. (2020) applied FOD to hyperspectral monitoring of sorghum ChD, all achieved good monitoring results. FOD is a mathematical concept that can be defined in many ways, with Grünwald-Letnikov (GL), Riemann Liouville (RL), and Caputo being the most widely used definitions (Benkhettou et al., 2014). Among them, GL method is defined based on discrete points. Both RL method and Caputo method calculate derivatives

by curve fitting discrete points first. But when calculating the traditional integer-order derivative, the curve fitting is performed first and then the derivative is calculated. However, in the preprocessing of spectral curves with FOD, most of the predecessors used GL method. The reason is that the data constituting the spectral curve itself is discrete point data, which is more consistent with the definition of GL method. This leads to a certain difference between its calculation of integer-order derivative and traditional methods. Therefore, in order to more accurately observe the role of different orders of FOD preprocessing in the construction of winter wheat ChD, it is necessary to compare the three calculation methods at the same time.

For band selection, many studies chosen to reduce model complexity by constructing vegetation index when constructing hyperspectral monitoring models of winter wheat ChD (Zhao et al., 2011; Xing et al., 2022). However, the vegetation index often only retains a few bands in the full spectrum band, which may result in the loss of much useful spectral information in the spectral curve. Therefore, it is necessary to use appropriate band selection algorithms to screen an appropriate number of band to improve the utilization rate of spectral information (Wang et al., 2022). Competitive adaptive reweighted sampling (CARS) is a screening algorithm based on the regression coefficients of the PLSR model, which has a wide range of applications in hyperspectral model research (Chen et al., 2023; Li and Yang, 2023). This study will use CARS to screen band that are important to ChD.

For modeling algorithms, a large number of current studies have shown that using machine learning algorithms to construct hyperspectral monitoring models has become one of the hot directions in the field of hyperspectral research, and there have been many reports on wheat chlorophyll (Sun et al., 2007; Wang et al., 2022; Feng et al., 2023; Huang et al., 2023). Based on previous research, it can be seen that PLSR, support vector regression (SVR), multi-layer perceptron regression (MLPR), random forest regression (RFR), extra-trees regression (ETsR), decision tree regression (DTR), K-nearest neighbors regression (KNR), and gaussian process regression (GPR) are commonly used machine learning algorithms. This study will use these algorithms to construct hyperspectral monitoring models.

Based on the above analysis, this study will use winter wheat as the research material, preprocess the original spectral curve using GL, RL, and Caputo FOD calculation methods, screen the sensitive band of winter wheat ChD using CARS method, and construct hyperspectral monitoring models of winter wheat ChD using various modeling methods such as PLSR, SVR, MLPR, RFR, ETsR, DTR, KNR, and GPR. Hope to achieve hyperspectral monitoring of ChD in winter wheat. The aims of this study are: (1) Clarify the preprocessing effects of GL, RL, and Caputo FOD calculation methods on the original spectral curve and their relationship with winter wheat ChD; (2) Analyze the effects of CARS method in screening ChD sensitive band in winter wheat; (3) Under 3 types of FOD preprocessing, hyperspectral monitoring models for winter wheat ChD were constructed using various machine learning algorithms based on the full spectrum band and the sensitive band screened by CARS, achieving rapid monitoring of winter wheat ChD.

2 Materials and methods

2.1 Experimental design

This study conducted a two-year winter wheat irrigation experiment from October 2020 to July 2022 at the Agricultural Station of Shanxi Agricultural University in Taigu District, Jinzhong City, Shanxi Province. The tested variety was 'Jintai 182', which was sown in October every year and harvested in July the next year. The average organic matter content of the tested soil was 13.92 $g{\cdot}kg^{-1}$, total nitrogen content was 1.19 $g{\cdot}kg^{-1}$, available phosphorus content was 17.43 $g{\cdot}kg^{-1}$, and available potassium content was 189.21 $g{\cdot}kg^{-1}$. Treatment started from the jointing stage of winter wheat, chose the jointing, flowering, and filling stages for irrigation. A total of 5 treatments were set up: T1 (No irrigation), T2(Irrigation once during the jointing stage), T3 (Irrigation once each during the jointing and flowering stage), T4(Irrigation once each during the jointing and filling stage), and T5(Irrigation once each during the jointing, flowering, and filling stage). Each irrigation amount was 60 mm, and each treatment was repeated 3 times. The experimental site constructed a water tank according to FAO standards, with a total of 15 plots, the area of each plot was 6 m^2(2 m×3 m). Nitrogen, phosphorus, and potassium fertilizers were used as base fertilizers and applied uniformly before sowing. The fertilization standards were N: 150 $kg{\cdot}hm^{-2}$, P_2O_5: 120 $kg{\cdot}hm^{-2}$, and K_2O: 120 $kg{\cdot}hm^{-2}$, respectively. The spacing between sowing rows was 20 cm. The remaining field operations were consistent with local farmers.

2.2 Index measurement

Canopy hyperspectral reflectance data and ChD of winter wheat were obtained at regreening, jointing, booting, heading, flowering, early grain filling, later grain filling, and maturity stages, a total of 8 stages of data were obtained.

The canopy hyperspectral data was measured using a Field-Spec 3.0 spectrometer(ASD Company, Boulder, CO, USA), with a collection wavelength range of 350-2500 nm. The 350~1000 nm spectral sampling interval is 1.4 nm, and the spectral resolution is 3 nm; the 1000~2500 nm spectral sampling interval is 2 nm, and the spectral resolution is 10 nm. In order to reduce the error caused by light changes, the spectral collection should be conducted in sunny and cloudless weather, with no wind or wind speeds below level 3, the collection time was arranged at 10:00-14:00. When measuring, place the probe 1 meter above the canopy and measure 10 times in each plot. A whiteboard was used to correct before each measurement. After removing the abnormal curve, the average spectrum was calculated as the final spectrum of that plot.

Several representative functional leaves(Before the jointing stage was first unfolded leaf, after the jointing stage was flag leaf) were collected at the location where canopy hyperspectral data was collected. The leaves were transferred to the room. After the veins were removed, the leaves were cut into pieces. 0.0500 g was weighed and put into a 25 ml volumetric flask. 96% ethanol was used to fix the volume to the scale line. The volumetric flask was stored in the

dark and away from light. The volumetric flask was shaken for 3 to 4 times in the middle. After 24 hours, the absorbance values at 649 nm and 665 nm were measured by spectrophotometer. The concentration formula of chlorophyll a, b, and total chlorophyll was as follows:

$$C_a = 13.95 \times A_{665} - 6.88 \times A_{649}$$

$$C_b = 24.96 \times A_{649} - 7.32 \times A_{665}$$

$$C = C_a + C_b$$

Where, C_a, C_b, and C were the concentration of chlorophyll a, chlorophyll b, and total chlorophyll in the soaking solution ($mg{\cdot}L^{-1}$). A_{649} and A_{665} were the absorbance values measured at the wavelength of 649 nm and 665 nm respectively.

The calculation formula of chlorophyll content and ChD was:

$$\text{Chlorophyll content}(mg \cdot g^{-1}\ \ FW) = \frac{C \times V}{FW \times 1000} \times n$$

$$\text{CHD}(g \cdot m^{-2}) = \text{Chlorophyll content} \times \text{Aboveground fresh leaf biomass} \times 1000$$

Where, V was the volume of soaking solution, which was 25 ml in this study; FW was the blade weight; N was the dilution ratio; 1000 was the unit conversion factor; The aboveground fresh leaf biomass refers to the fresh weight of all leaves per unit surface area ($kg{\cdot}m^{-2}$).

2.3 Data analysis methods

Before data analysis, the data set was divided into calibration set and validation set according to the 2:1 ratio using the concentration gradient method. Specifically, arranged the dependent variables from small to large, selected the first and third samples from every three samples to enter the calibration set, and the second sample to enter the validation set. At the same time, ensure that the maximum and minimum values of the data set were put into the calibration set.

2.3.1 Fractional-order derivative

FOD was first proposed by Italian mathematician Samuelson in 1695 and had a history of over 300 years. However, systematic research on FOD has mainly focused on the past few decades (MaChado et al., 2011; Li et al., 2017). After years of development, many forms of FOD definitions have emerged, among which GL, RL, and Caputo are the most widely used definitions (Benkhettou et al., 2014; Yang et al., 2022b). The as follows:

$$\mathbf{GL}: \frac{d^{\alpha} f(\lambda)}{d\lambda^{\alpha}} = \lim_{h \to 0} \frac{1}{h^{\alpha}} \sum_{m=0}^{\frac{p-q}{h}} (-1)^m \frac{\Gamma(\alpha+1)}{m!\,\Gamma(\alpha-m+1)} f(\lambda - mh)$$

$$\mathbf{RL}: \frac{d^{\alpha} f(\lambda)}{d\lambda^{\alpha}} = \frac{1}{\Gamma(k-\alpha)} \frac{d^{k}}{dx^{k}} \int_{q}^{\lambda} \frac{f(\tau)}{(\lambda-\tau)^{\alpha-k+1}} d\tau, 0 \le k-1 < \alpha < k$$

$$\mathbf{Caputo}: \frac{d^{\alpha} f(\lambda)}{d\lambda^{\alpha}} = \frac{1}{\Gamma(k-\alpha)} \int_{q}^{\lambda} \frac{f^{(k)}(\tau)}{(\lambda-\tau)^{\alpha-k+1}} d\tau, 0 \le k-1 < \alpha < k$$

Among them, α was a real number, representing any order, in this study, its value range was [0,2], with a step size of 0.1; λ and τ were wavelengths; For the GL method, f(λ) was the spectral reflectance value at λ wavelength, for the RL method and Caputo method, f(λ) was the spectral curve formula at λ; h was the differential step size, which was 1 in this paper; p was the upper limit of differentiation; q was the lower limit of differentiation; m represented the number of bands before the λ wavelength, and m=p-q; K was the smallest integer greater than α; Σ was the summation symbol; Γ(x) was the Gamma function; ∫ was the integral symbol; $f^{(k)}(\tau)$ represented taking the k-th derivative of function f(τ).

From the above equation, it can be seen that the GL method is based on a discrete definition. The FOD at point λ is related to the spectral reflectance of all wavelengths before that wavelength. For the convenience of calculation, combined with previous research, we uniformly selected the 40 points before λ for calculation (Yang et al., 2022b). Both the RL method and Caputo method need to first fit the formula f(λ) at λ. In this paper, the spectral reflectance at three consecutive wavelengths (i.e. λ-1, λ, λ+1) was selected for binomial fitting calculation of f (λ).

2.3.2 Competitive adaptive reweighted sampling

CARS is an algorithm based on the regression coefficients of the PLSR model. The main idea of this algorithm is to mimic the "survival of the fittest" principle in Darwin's evolutionary theory. Each time, a certain number of samples (usually 80% of all sample sizes) were selected through Monte Carlo sampling to construct the PLSR model, and the bands with the higher absolute weight of regression coefficients were selected as the new subset. Then selected samples through Monte Carlo sampling to construct the model. After multiple calculations, only two bands were retained to enter the model, and the cycle ended. From multiple PLSR models obtained, selected the band in the subset with the smallest root mean square error of cross-validation as the final selected feature band (Tang et al., 2023; Zhang et al., 2024).

2.3.3 Modeling algorithm

PLSR is a modeling method that combines principal component analysis, canonical correlation analysis, and multiple linear regression analysis. It is currently one of the most effective linear regression methods in constructing hyperspectral models (Yang et al., 2022b). The other seven algorithms are all nonlinear regression methods. SVR mainly maps raw data to a high-dimensional feature space, and achieves regression tasks by finding a suitable hyperplane (Yang et al., 2022a). MLPR is a type of artificial neural network, which is a relatively simple neural network (Yang et al., 2022a). RFR is an ensemble algorithm based on DTR, which completes regression tasks by integrating the prediction results of multiple decision trees (Han et al., 2022). ETsR is an improved algorithm of RFR, which is an ensemble learning algorithm that reduces fitting errors by combining the prediction results of multiple extreme random trees (Han et al., 2022). DTR is a method of regression analysis of data by constructing a decision tree model (Arjmandi et al., 2023). KNR is an instance-based learning method that finds K nearest neighbor samples in a sample and performs regression prediction based on

their labels (Maraden et al., 2023). GPR is a non-parametric model that uses Gaussian processes as priors and is a probability-based machine learning algorithm (Sahoo et al., 2023).

In this paper, Microsoft Excel 2021 was used to organize data. MATLAB 2021 was used to perform the FOD algorithm. Python 3.11 was used to perform CARS and modeling algorithms. Origin 2021 was used for mapping. Evaluate the accuracy of the model using the coefficient of determination (R^2), root mean square error (RMSE), and relative predictive deviation (RPD).

3 Result and analysis

3.1 Descriptive statistical analysis

The descriptive statistical analysis results were shown in Figure 1. From the figure, it can be seen that the maximum and minimum values of the total data set were 11.5426 g·m^{-2} and 0.0129 g·m^{-2}, respectively. When dividing the dataset, both the maximum and minimum values were assigned to the calibration set. The average value of the total data set was 3.5552 g·m^{-2}, and the average values of the calibration and validation sets were also close to this. For the standard deviation, it showed that the total data set and calibration set were slightly higher than the validation set. However, the sample distribution of the three data sets does not conform to a normal distribution to a certain extent. Based on this, this study performed logarithmic processing on ChD data, and the results were shown in Figure 2. From the figure, it can be seen that, similar to Figure 1, the maximum and minimum values were also assigned to the calibration set when dividing the data set. The average and standard deviation of the three datasets were relatively close. However, after logarithmic processing, the distribution of the data sets was more in line with a normal distribution.

3.2 Changes in spectral reflectance

3.2.1 Changes in original spectral reflectance

From Figure 2, it can be seen that the Log(ChD) values were mainly distributed in the range of 0 to 0.8. Therefore, this paper mainly analyzed the changes in the original spectral reflectance of samples with four gradients of Log(ChD)<0, 0<Log(ChD)<0.4, 0.4<Log(ChD)<0.8, and 0.8<Log(ChD), as shown in Figure 3. At the same time, in order to reduce the effect of factors such as moisture in the air, the bands within the range of 350-399 nm, 1351-1399 nm, 1801-1950 nm, and 2451-2500 nm in the original spectrum were removed. From the figure, it can be seen that the hyperspectral reflectance of winter wheat canopy mainly showed a trend of first increasing and then decreasing with the increasing of wavelength. In terms of details, a small reflection peak was formed near 550 nm, a near-infrared reflection platform was formed in the range of 780-1100 nm, and two obvious absorption valleys were formed near 1000 nm and 1450 nm, which was consistent with the basic characteristics of spectral reflectance of green plant canopies. In addition, in the range of 400-700 nm, samples with different Log (ChD) exhibited an overall trend of decreasing spectral reflectance with increasing Log(ChD), while in the range of 740-1800 nm, it showed an increasing trend with increasing Log(ChD). This indicated that there may be a certain quantitative relationship between Log(ChD) and the spectral reflectance at some specific wavelengths.

3.2.2 The effect of FOD on spectral reflectance

Figure 4 showed the spectral reflectance curve based on GL-FOD preprocessing. Combined with Figure 3, it can be seen that as the order increased, there was a gradual change process in the GL spectral reflectance. In terms of details, there was a clear trend of gradual decreasing in the range of 0 to 1.0-order, and the smaller the order, the greater the magnitude of the decrease. In the range of 1.0 to 2.0-order, there were various trends of variation, such as a gradual decreasing in the visible light band range (such as 700-750 nm). In the partial band range of near-infrared (such as 1000-1050 nm), it generally showed a trend of first decreasing and then increasing, and reached its lowest point in the 1.4-order. In addition, compared to the change in spectral reflectance in the 0 to 1.0-order range, the change amplitude in spectral reflectance in the 1.0 to 2.0-order range was relatively small.

Figure 5 showed the spectral reflectance curve based on RL-FOD preprocessing. Unlike the spectral curve changing trend after GL preprocessing, the 0.1-order RL-FOD spectral curve lost the

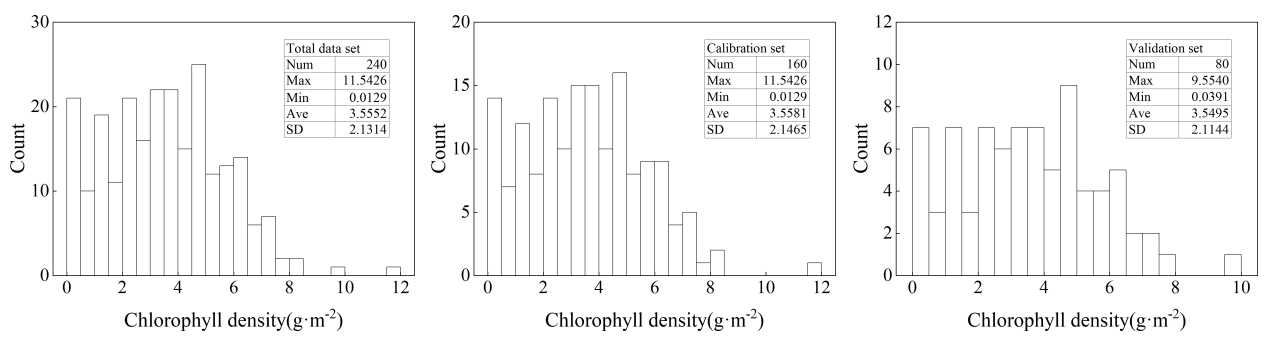


FIGURE 1
Descriptive statistical analysis of ChD. Num, Max, Min, Ave, and SD represent the number of samples, maximum, minimum, average, and standard deviation of the dataset, respectively. The same below.

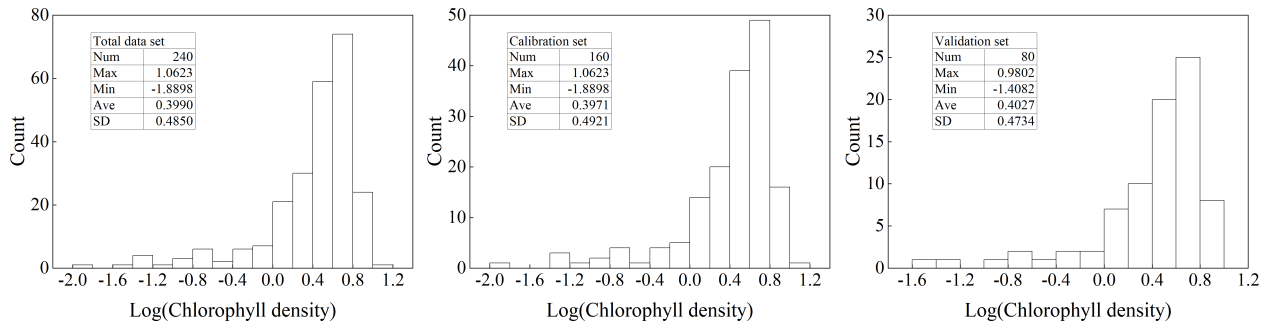


FIGURE 2
Descriptive statistical analysis of Log(ChD).

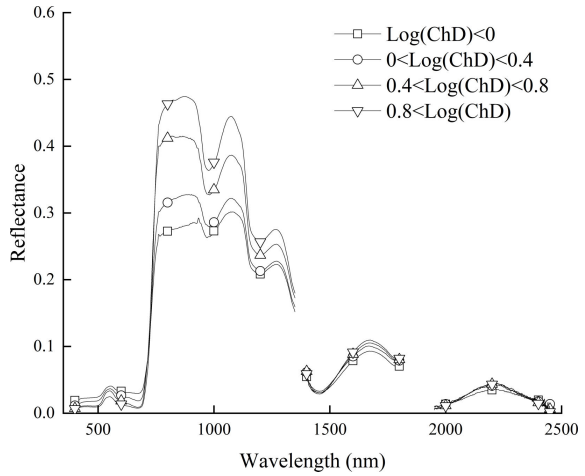


FIGURE 3
Changes in original spectral reflectance.

basic characteristics of the original spectral curve and had a basic changing trend of the 1.0-order derivative. As the order increased, the overall change in RL spectral reflectance was also different from that of GL spectral reflectance. The main performance was that as the order increased, the RL spectral reflectance gradually increased and gradually decreased alternately. For example, in the range of 700-750 nm, it showed a gradually decreased trend, while in the range of 1100-1150 nm, it showed a gradually increased trend. And during the continuous 0.1-order variation process, the amplitude of the change in RL spectral reflectance was relatively small.

Figure 6 showed the spectral reflectance curve based on Caputo-FOD preprocessing. Similar to the preprocessing results of the RL method, began from the 0.1-order, the Caputo-FOD spectral curve lost the basic characteristics of the original spectral curve. However, in the range of 0.1-1.0 order and 1.1-2.0 order, Caputo spectral reflectance exhibited similar changes with increasing wavelength. As the order changed, the Caputo spectral reflectance at the same wavelength alternated between increasing and decreasing. For example, in the range of 0.1 to 0.5-order, the spectral reflectance gradually increased in the range of 700 to 750 nm, while it gradually decreased in the range of 1140 to 1150 nm. The change in the range of 0.6 to 1.0-order was opposite to this, showing a gradual decreasing in the range of 700-750 nm and an increasing in the range of 1140-1150 nm. The performance in the range of 1.1 to 2.0-order was similar to that of 0.1 to 1.0-order, with 1.5-order and 1.6-order being the critical values for spectral reflectance changing at the same wavelength. However, within the range of 1.1 to 2.0-order, the changing amplitude of the Caputo spectral reflectance curve was smaller. It was worth noting that, unlike the GL and RL methods, the Caputo spectrum exhibited a significant changing amplitude from 1.0 to 1.1-order, and a gradual change cannot be observed, which may be related to different algorithmic properties.

3.3 Correlation analysis

Figure 7 showed the correlation coefficients between three FOD spectra and Log(ChD). From the figure, it can be seen that most of the bands in the original spectrum had a high correlation with Log (ChD), with a maximum correlation coefficient of -0.8625. The three types of FOD all increased the correlation between the original spectrum and Log(ChD) to varying degrees, but the correlation showed different laws of change. For GL-FOD, as the order increased, it can be clearly observed gradually changing that the correlation coefficient between the original spectrum to 2.0-order spectrum and Log(ChD), and followed the law of higher correlation with smaller order and shorter wavelength. For RL-FOD, only a gradual change in the correlation coefficient between the 0.1 to 2.0-order spectrum and Log(ChD) could be observed, and it also showed a law of higher correlation with smaller order and shorter wavelength. For Caputo-FOD, the correlation coefficients between the 0.1 to 1.0-order spectrum with the 1.1 to 2.0-order spectrum and Log(ChD) were the same basically. However, in the range of 0.1 to 1.0-order, the correlation between Caputo-FOD spectrum and Log (ChD) was higher, and the trend of the correlation coefficient between RL spectrum and Log(ChD) was similar.

3.4 CARS screened band results

Figure 8 showed the positions of bands screened by CARS based on three FOD spectra, and Table 1 showed the number of bands screened by CARS. It can be seen that under different FOD spectra

10.3389/fpls.2024.1492059

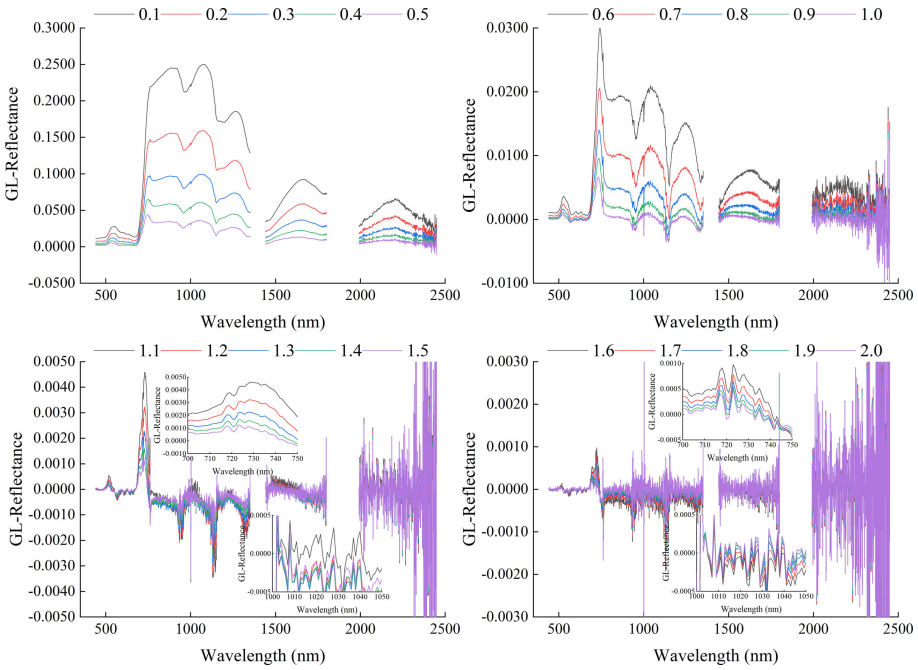


FIGURE 4
Spectral reflectance curve based on GL method preprocessing.

and order conditions, the minimum number of bands screened by CARS was 11 bands screened under 1.5-order RL-FOD spectral conditions, and the maximum was 1060 bands screened based on the original spectrum. Most of the number of screened bands were less than 200. In terms of the position of the selected bands, they were distributed throughout the entire spectral range. For the GL-FOD spectrum, the bands selected at lower orders were mainly concentrated in the visible light band range, with less distribution in the near-infrared band. The bands selected at higher orders were distributed throughout the entire spectral range. For RL-FOD and Caputo FOD spectra, the selected band positions were evenly distributed throughout the entire spectral range, but there was also a relatively concentrated distribution in the visible light band range.

3.5 Hyperspectral monitoring model for ChD in winter wheat

3.5.1 Hyperspectral monitoring model based on the full spectrum band

Figure 9 showed the results of using 8 machine learning algorithms to construct models based on the full spectrum band of GL-FOD spectra. From the figure, it can be seen that in terms of modeling accuracy, all ETsR, DTR, and GPR models had high calibration accuracy, with R^{2c} of 1.0000 and $RMSE_c$ of 0.0000. Next, was the model constructed based on RFR, whose R^{2c} was generally stable around 0.9600 and $RMSE_c$ was stable around 0.0900. The model calibration accuracy based on PLSR and SVR was also high, with R^{2c} generally stable around 0.8000 and $RMSE_c$ stable around 0.2000. The model constructed based on MLPR and KNR only had high calibration accuracy when the order was low, and as the order increased, its calibration accuracy decreased to varying degrees. Unlike calibration accuracy, in models with ETsR, DTR, and GPR with R^{2c} of 1.0000, only ETsR had relatively stable validation accuracy, with R^{2v} stable around 0.8000, $RMSE_v$ stable in the range of 0.2000 to 0.2500, and RPD stable in the range of 1.8000 to 2.2000. The validation accuracy of the DTR model varied greatly with the changing of order, with R^{2v} mainly distributed in the range of 0.6000 to 0.7000. However, the change amplitude of RPD was large, with a maximum of 1.7476 and a minimum of only 0.9091. The validation accuracy of the GPR model was relatively low, and it gradually decreased with increasing order, with lower R^{2v} and RPD. The model constructed based on RFR and PLSR still had relatively stable validation accuracy, with R^{2v} stable around 0.8000 and $RMSE_v$ stable around 0.2000. For the same order, the RPD of the RFR model was higher when the order was small, and the RPD of the PLSR model was higher when the order was high. For the SVR model, when the order was between 0.0-1.3, the model still had high validation accuracy, but as the order increased, the validation accuracy gradually decreased. Similar to calibration accuracy, models constructed based on MLPR and KNR also had high validation accuracy at lower order. Through comprehensive comparison, the model constructed by RFR based on the 0.3-order GL-FOD spectrum was the best among all models, with R^{2c}, $RMSE_c$, R^{2v}, $RMSE_v$, and RPD of 0.9640, 0.0931, 0.8442, 0.1865, and 2.4762, respectively. It had high accuracy and stability.

Figure 10 showed the results of using 8 machine learning algorithms to construct models based on the full spectrum band of RL-FOD spectra. From the figure, it can be seen that the changes in calibration accuracy and validation accuracy of the models constructed based on RL-FOD spectra were similar to those of

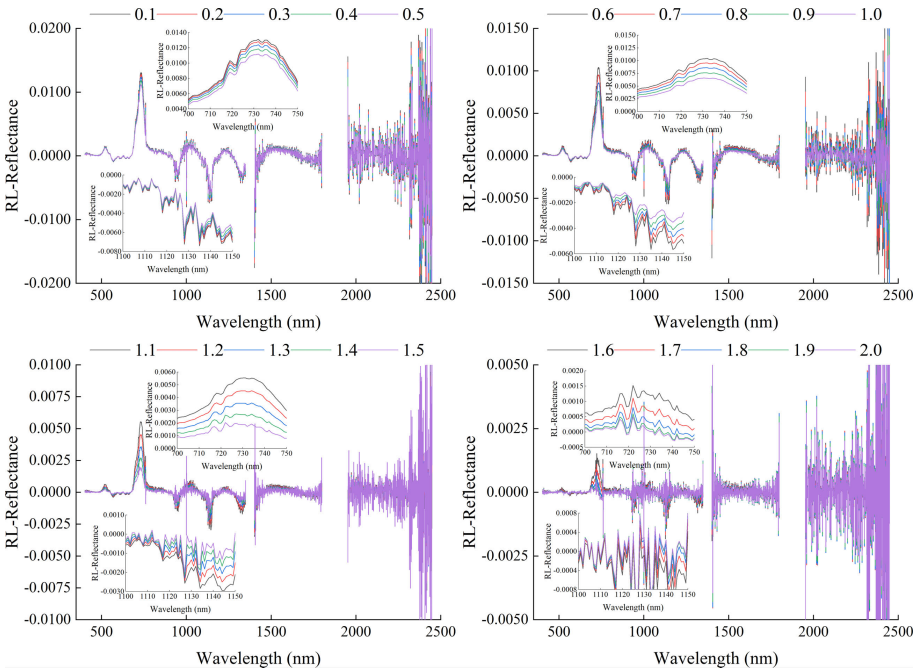


FIGURE 5
Spectral reflectance curve based on RL method preprocessing.

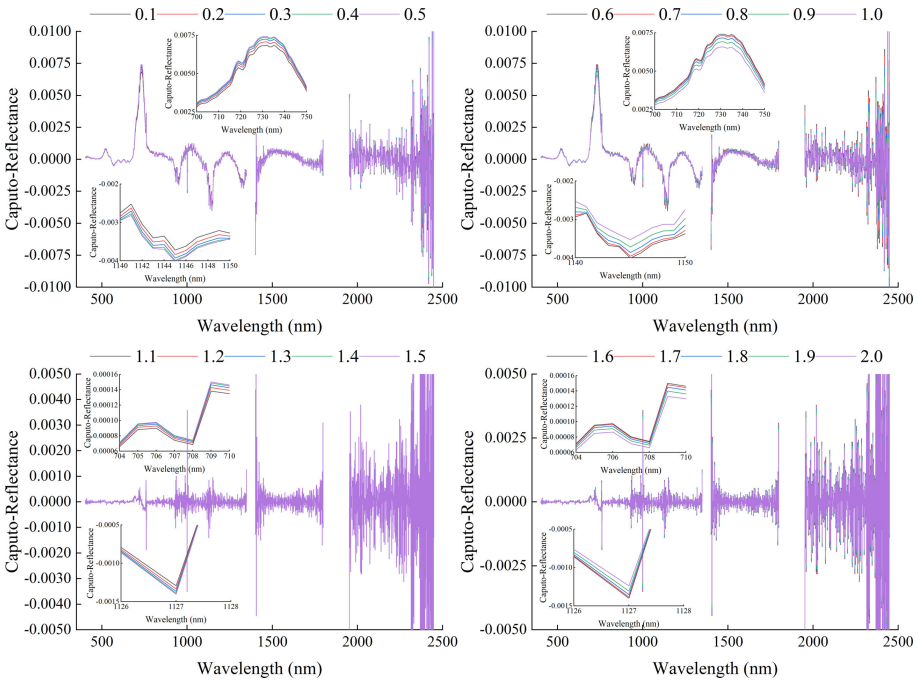


FIGURE 6
Spectral reflectance curve based on Caputo method preprocessing.

the models constructed based on GL-FOD spectra. Among all models, the model constructed by ETsR based on the 0.3-order RL-FOD spectrum had the highest accuracy, with R^{2c}, $RMSE_c$, R^{2v}, $RMSE_v$, and RPD of 1.0000, 0.0000, 0.8470, 0.1860, and 2.4823, respectively.

Figure 11 showed the results of using 8 machine learning algorithms to construct models based on the full spectrum band of Caputo-FOD spectra. From the figure, it can be seen that similar to the GL and RL methods, the models constructed based on Caputo-FOD spectra by ETsR, DTR, GPR, and RFR still had high

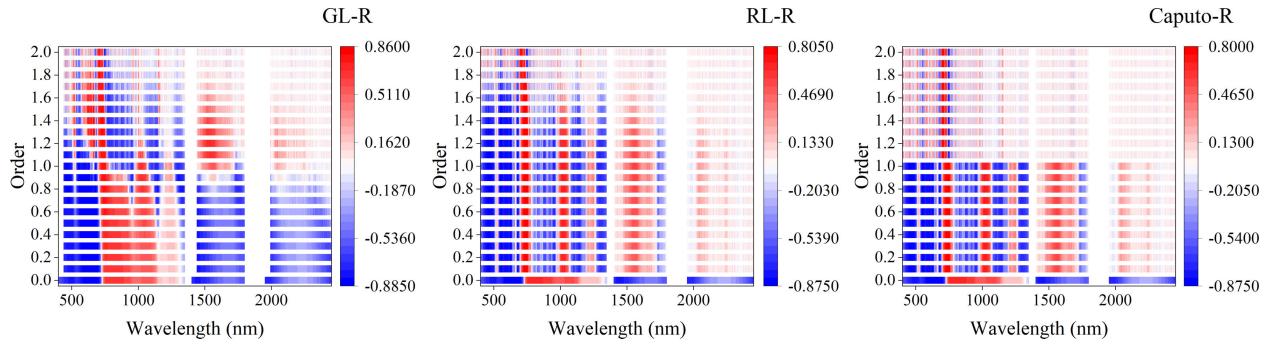


FIGURE 7
The correlation between different FOD spectra and Log(ChD). 0.0-order represents the original spectrum.

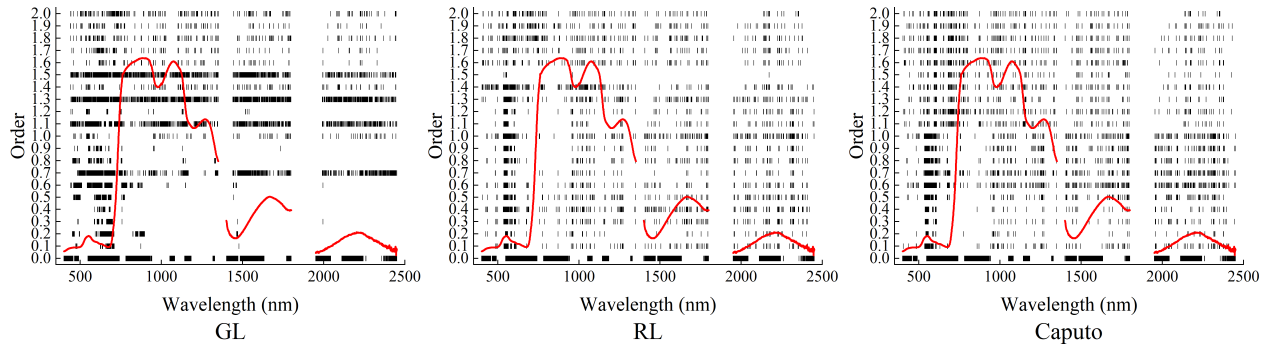


FIGURE 8
Position of bands screened by CARS based on three types of FOD spectra.

calibration accuracy, but only the ETsR and RFR models had relatively stable validation accuracy. The R^{2c} of the model constructed based on PLSR and SVR was stable around 0.8000, but the validation accuracy of the SVR model constructed in the range of 1.1 to 2.0-order was significantly reduced compared to the model constructed in the range of 0.1 to 1.0-order. The model constructed based on MLPR and KNR only had relatively high calibration accuracy and validation accuracy in the range of 0.0 to 1.0-order. Through comprehensive comparison, among all models constructed based on the full spectrum band of Caputo-FOD spectra, the model constructed by RFR based on the 0.8-order Caputo-FOD spectrum had the best accuracy, with R^{2c}, $RMSE_c$, R^{2v}, $RMSE_v$, and RPD of 0.9594, 0.0989, 0.8408, 0.1892, and 2.4398, respectively.

By comparison, it can be seen that among the models constructed based on the full spectrum band of the three FOD spectra, the model constructed by ETsR based on the full spectrum band of the 0.3-order RL-FOD spectrum had the highest accuracy.

3.5.2 Hyperspectral monitoring model based on screening band

Figure 12 showed the results of using 8 machine learning algorithms to construct models based on the CARS band of GL-FOD spectra. From the figure, it can be seen that the models constructed based on ETsR, DTR, GPR, and RFR all had high calibration accuracy, but only the models constructed based on ETsR and RFR had relatively high validation accuracy. The R^{2c} of the model constructed based on PLSR and SVR was stable around 0.8000, and the $RMSE_c$ was distributed around 0.2000. For validation accuracy, the validation accuracy of PLSR model increased with increasing order, while the SVR model decreased with increasing order. The MLPR and KNR models only had high calibration and validation accuracy when the order was low, and the KNR model generally had higher accuracy than the MLPR model. Among all models, the model constructed by SVR based on the 1.0-order GL-FOD spectrum had the best accuracy, with R^{2c}, $RMSE_c$, R^{2v}, $RMSE_v$, and RPD of 0.8716, 0.1758, 0.8595, 0.1832, and 2.5205, respectively.

Figure 13 showed the results of using 8 machine learning algorithms to construct models based on the CARS band of RL-FOD spectra. From the figure, it can be seen that the calibration accuracy of the models constructed based on ETsR, DTR, GPR, RFR, PLSR, and SVR was relatively high, with R^{2c} distributed in the range of 0.8000 to 1.0000 and $RMSE_c$ distributed in the range of 0.0000 to 0.2000, respectively. The calibration accuracy of the model constructed based on KNR was relatively low, with R^{2c} mainly distributed between 0.6000 and 0.8500, and $RMSE_c$ mainly

10.3389/fpls.2024.1492059

TABLE 1 Number of bands screened by CARS based on three types of FOD spectra.

Order	GL	RL	Caputo	Order	GL	RL	Caputo
0.0	1060						
0.1	55	131	150	1.1	659	99	131
0.2	166	65	57	1.2	48	99	173
0.3	63	150	75	1.3	869	199	114
0.4	32	228	75	1.4	126	228	150
0.5	126	173	173	1.5	757	11	86
0.6	251	131	347	1.6	83	114	150
0.7	659	75	302	1.7	110	114	150
0.8	95	131	199	1.8	144	173	150
0.9	48	99	199	1.9	95	131	131
1.0	95	262	262	2.0	218	114	114

distributed between 0.1900 and 0.3200. The calibration accuracy of models constructed based on MLPR was all relatively low, with R^{2c} approaching 0. Only models constructed based on ETsR and RFR had relatively high and stable validation accuracy in the validation model. Among all models, the model constructed by ETsR based on the 0.3-order RL-FOD spectrum had the best accuracy, with R^{2c}, $RMSE_c$, R^{2v}, $RMSE_v$, and RPD of 1.0000, 0.0000, 0.8667, 0.1732, and 2.6660, respectively.

Figure 14 showed the results of using 8 machine learning algorithms to construct models based on the CARS band of Caputo-FOD spectra. Similar to RL-FOD spectra, models constructed based on ETsR, DTR, GPR, RFR, PLSR, and SVR all had high calibration accuracy. Models constructed based on KNR only had high calibration accuracy when the order was small, while models constructed based on MLPR had low calibration accuracy. For validation accuracy, models constructed based on ETsR and RFR all had high and stable validation accuracy. Excepted for the MLPR model, the validation accuracy of all other models varied significantly with the increasing in order. Among all models, the model constructed by KNR based on the 0.7-order Caputo-FOD spectrum had the best accuracy, with R^{2c}, $RMSE_c$, R^{2v}, $RMSE_v$, and RPD of 0.8234, 0.2062, 0.8345, 0.1954, and 2.3630, respectively. Comparing all models constructed based on the CARS band, the model constructed by ETsR based on the 0.3-order RL-FOD spectrum had the best accuracy. At the same time, this was also the model with the highest accuracy obtained in this study. Supplementary Figure 1 showed the 1:1 fitting figure of measured values and predicted values for this model. It can be seen from the figure that the model had a good prediction effect and can realize the hyperspectral estimation of Log(ChD).

4 Discussion

Proper preprocessing of the original hyperspectral reflectance data to reduce noise interference and improve the final model accuracy is an important step in the construction process of hyperspectral monitoring models. By comparing some previous studies, it can be seen that 1-order and 2-order derivative preprocessing in derivative preprocessing are common and effective spectral preprocessing methods (Ma et al., 2022; Cozzolino et al., 2023). However, there are significant differences in the original spectrum, 1-order derivative spectrum, and 2-order derivative spectrum curves, and there are also certain differences in the modeling results (Yang et al., 2023a). The emergence of FOD provided the possibility for studying the process of changes from original spectrum to integer-order derivative spectrum. There are

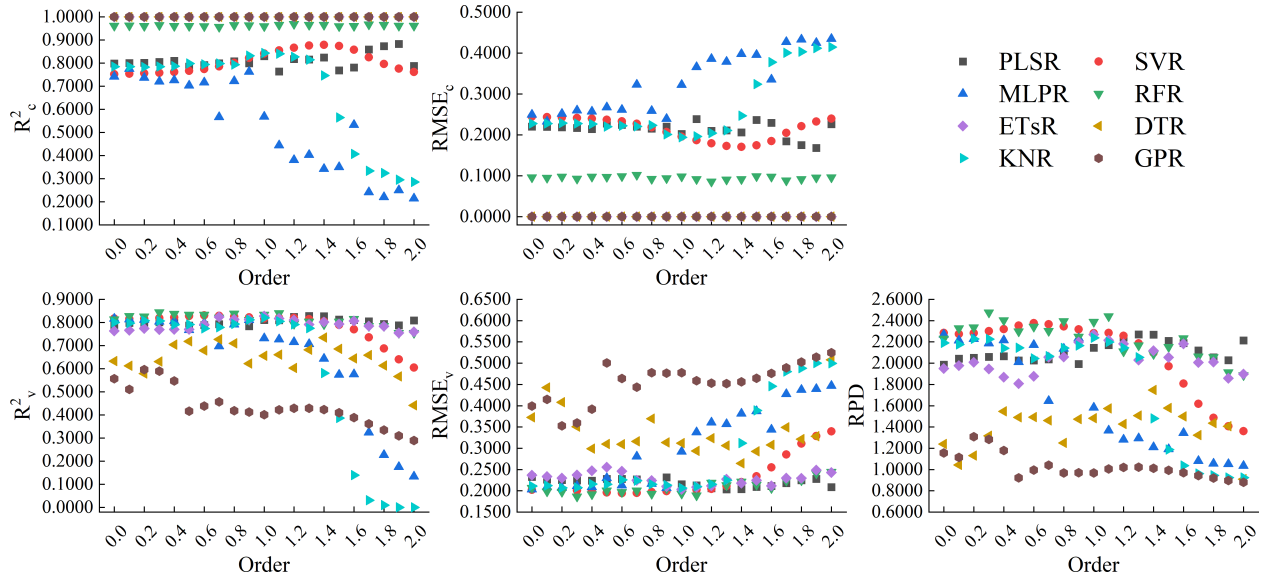


FIGURE 9
Modeling results based on the full spectrum band of GL-FOD spectra.

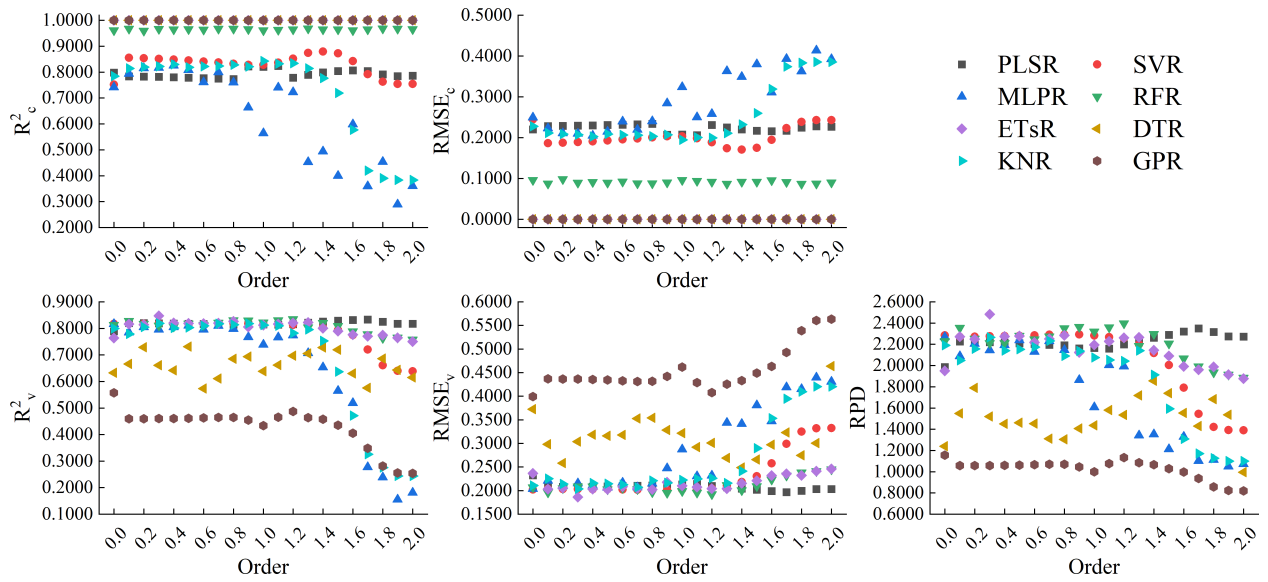


FIGURE 10
Modeling results based on the full spectrum band of RL-FOD spectra.

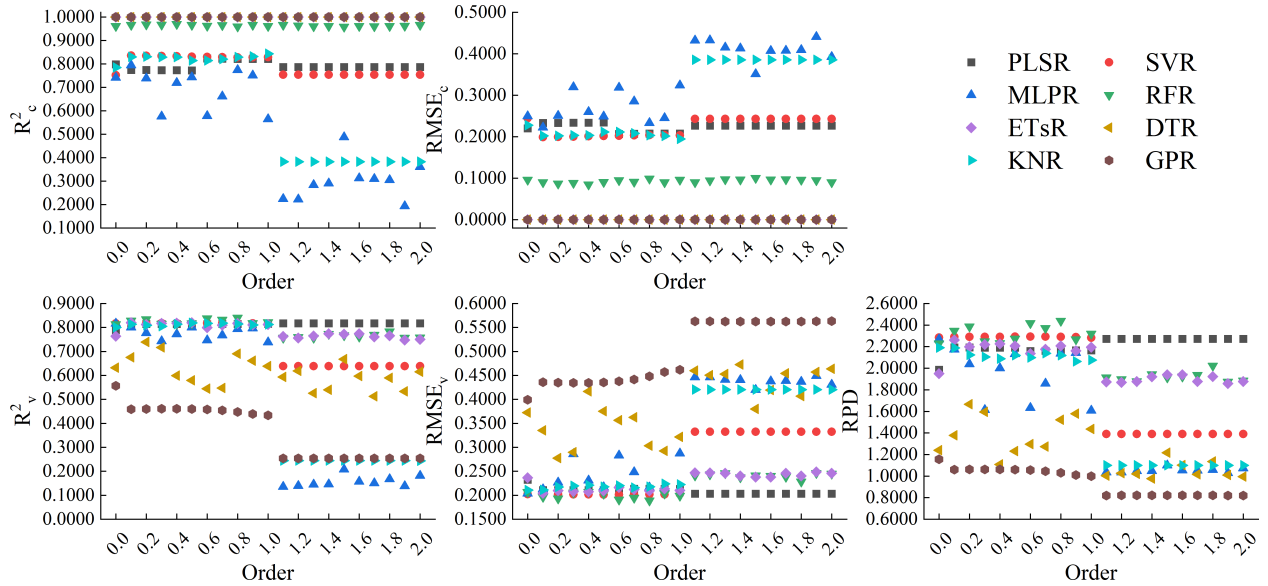


FIGURE 11
Modeling results based on the full spectrum band of Caputo-FOD spectra.

many calculation methods for FOD, among which GL, RL, and Caputo are the three most widely used calculation methods. From its calculation method, it can be seen that the GL method is defined based on discrete points, making it very suitable for processing hyperspectral reflectance data. It is also the most commonly used method for applying FOD to hyperspectral research at present (Geng et al., 2024; Liu et al., 2023). The RL method and Caputo method have a similar principle, both of which first perform function fitting on discrete points before calculating FOD. The difference is that the RL method first integrates the function and then takes the derivative, while the Caputo method takes the derivative first and then the integral. It can be seen that the three methods have different calculation principles. This paper compared the roles of three FOD calculation methods in constructing hyperspectral monitoring models for winter wheat ChD. Because the ChD data obtained in this study does not follow a normal distribution, logarithmic processing was performed on the ChD data to make it conform to a normal distribution before further analysis (Qi et al., 2024). The results indicated that there may be a certain quantitative relationship between Log(ChD) and spectral reflectance at specific wavelengths, which was consistent with the research results of Feng et al. (2023). From Figures 4-6, it can be

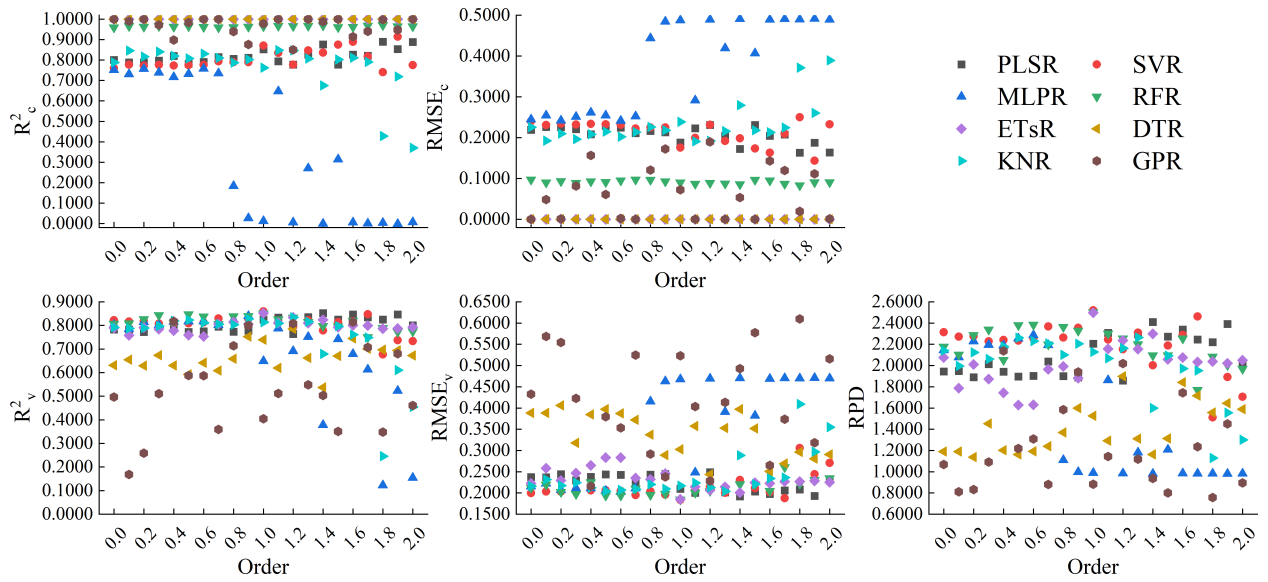


FIGURE 12
Modeling results based on the CARS band of GL-FOD spectra.

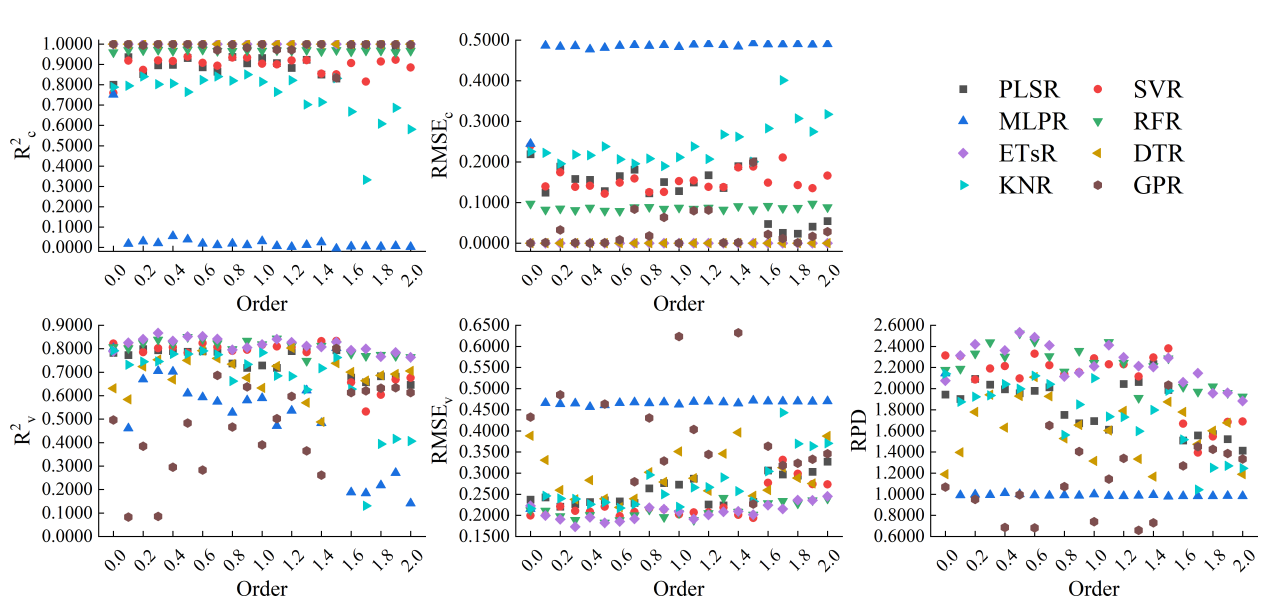


FIGURE 13
Modeling results based on the CARS band of RL-FOD spectra.

seen that the spectral curves preprocessed by the three FOD calculation methods had different variation characteristics. The GL-FOD spectra gradually lost the basic characteristics of the original spectral curve as the order increased, and the trend of the original spectral curve towards integer-order derivative can be observed clearly, especially the trend towards 1.0-order derivative curve. Previous studies have also shown that GL-FOD can display the process of changes from the original spectra to integer-order spectra, and it was believed that these subtle changes may provide more spectral features, thereby improving the accuracy of the final model (Cheng et al., 2022; Yang et al., 2022b). Both RL-FOD and Caputo FOD spectra lost the basic characteristics of the original spectra at the 0.1-order, which may be due to the calculation methods of these two methods. Because integer-order derivatives occurred in the calculation process of these two methods, this may directly lead to spectral curves in the range of 0.1-0.9 and 1.1-1.9 having varying characteristics of 1.0 and 2.0-order spectral curves, respectively (Yu and Liang, 2023; Alshammari et al., 2024). However, the trend of gradually changing from 0.1-order spectra to 2.0-order spectra can still be observed. The difference was that the

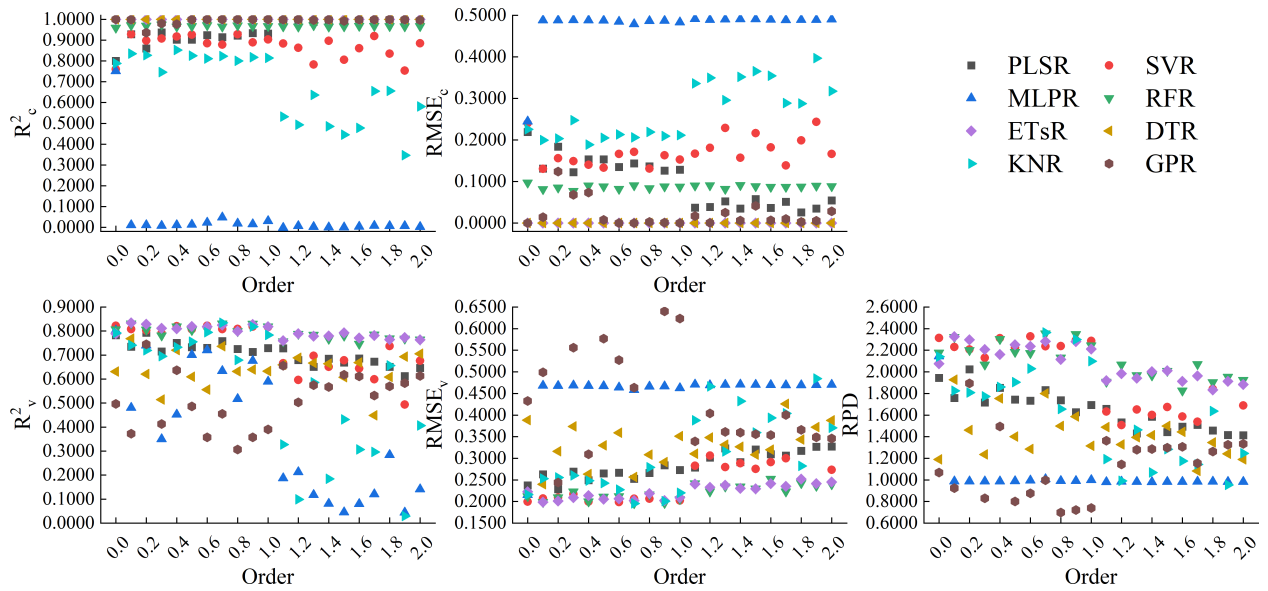


FIGURE 14
Modeling results based on the CARS band of Caputo-FOD spectra.

difference between Caputo-FOD spectra of adjacent orders was smaller. Meanwhile, compared to the other two algorithms, the spectral curves of Caputo-FOD were extremely close in the range of 0.1-1.0 orders and 1.1-2.0 orders, respectively. This indicated that the Caputo method may not be sensitive to explaining the changes between integer-order derivative spectra, and correlation analysis also confirmed this. The correlation between Caputo-FOD spectra and Log(ChD) at the same wavelength was consistent in the range of 0.1-1.0 and 1.1-2.0 orders, respectively. Both GL and RL methods can observe a gradual change in correlation with the order changed at the same wavelength, and the correlation between GL-FOD spectra and Log(ChD) was more sensitive to changes in order. Moreover, the GL method can observe a trend in the correlation between the original spectra towards FOD spectra and the log (ChD), while the RL method and Caputo method cannot. This was also related to the fact that the GL method can reflect the gradual change of the original spectra towards the integer-order derivative spectra.

With the development of remote sensing technology in recent years, hyperspectral remote sensing technology not only has more spectral information but also increases the redundancy of spectral information. Meanwhile, according to the Hughes phenomenon, the accuracy of hyperspectral monitoring models is closely related to the number of bands entering the model (Hughes, 1968). Therefore, selecting a certain number of spectral bands from the full spectrum band of hyperspectral data to construct models has become one of the important contents of current hyperspectral-related research (Yang et al., 2023b). This paper used CARS to screen the bands in the original spectrum and three FOD spectra. From the screening results, it can be seen that in addition to 1060 bands screened out from the original spectrum, GL-FOD screened out 32-869 bands to varying degrees, RL-FOD screened out 11-262 bands, and Caputo-FOD screened out 57-262 bands. Excluding most of the spectral bands provided the possibility to reduce model complexity and improve model accuracy. The screening effect was similar to the research results of Sudu et al. (2022) on maize SPAD. In addition, although the bands screened by CARS were distributed throughout the entire spectrum, there was a more concentrated distribution in the visible light section when the order was low. Correlation analysis also indicated that the visible light bands of low-order FOD spectra had a high correlation with Log(ChD). This may be due to the more important relationship between some bands of visible light and the winter wheat chlorophyll. Feng et al. (2013) believed that the 650 nm and 670 nm wavelengths were important bands for monitoring winter wheat chlorophyll. Yang et al. (2023c) believed that the hyperspectral reflectance at 549 nm and the 1.0-order derivative spectrum at 735 nm had a strong correlation with wheat chlorophyll. The spectral reflectance at 536 nm, 596 nm, 674 nm, and the 1.0-order derivative spectra at 756 nm and 778 nm can be used to estimate winter wheat chlorophyll content. This study screened out the bands at these positions to varying degrees.

Different model construction algorithms directly determine the structure of the model, so choosing the appropriate modeling algorithm to construct the model is an important step in improving model accuracy. By analyzing previous research, it can be seen that PLSR is a common and widely recognized linear modeling algorithm with good performance (Yang et al., 2022b). Meanwhile, comparing the effectiveness of different machine learning algorithms in constructing models is also one of the current hot topics in hyperspectral research (Graham Ram et al.,

2023; Silva et al., 2024). Therefore, this study selected 8 modeling algorithms, including PLSR, SVR, MLPR, RFR, ETsR, DTR, KNR, and GPR, to construct hyperspectral monitoring models of winter wheat Log(ChD) based on the full spectrum band and CARS band of three FOD spectra. From the model constructed based on the full spectrum band, it can be seen that the calibration accuracy of the model was relatively high when using ETsR, DTR, GPR, and RFR to construct the model. This was because DTR can effectively handle outliers and noise interference, reducing their effect on model accuracy, ETsR and RFR were both ensemble learning algorithms based on multiple decision trees. While GPR can capture complex structures and nonlinear relationships in the data. Therefore, the accuracy of the models constructed by them was high, but it was still necessary to be cautious of overfitting phenomena (Kwon et al., 2022; Sahoo et al., 2023). Secondly, the calibration accuracy of PLSR and SVR models was also relatively stable. The calibration accuracy of models constructed by MLPR and KNR decreased with increasing order. In terms of validation accuracy, models constructed based on GL-FOD and RL-FOD spectra exhibited a similar law of validation accuracy. The model constructed based on ETsR, PLSR, SVR, and RFR had high and stable validation accuracy, while the validation accuracy of the other 4 algorithms fluctuated significantly with changes in order. The model constructed based on Caputo-FOD spectra mainly showed that a certain algorithm had a relatively close accuracy in the range of 0.1-1.0 orders and 1.1-2.0 orders, respectively. This may be related to the close reflectance of Caputo-FOD spectra at the same wavelength, further indicating that the Caputo method was not sensitive to changes in FOD spectral curves. In addition, some algorithms such as GPR and DTR showed varying degrees of overfitting among the 8 modeling algorithms. Previous studies have suggested that this may be related to the high complexity and excessive number of iterations of the model. In future research, overfitting can be reduced by adjusting the number of iterations or regularization methods (Nansen et al., 2013; Singh et al., 2022). Among all models constructed based on the full spectrum band, the model constructed by ETsR based on the 0.3-order RL-FOD spectrum had the highest accuracy, with R^{2c}, $RMSE_c$, R^{2v}, $RMSE_v$, and RPD of 1.0000, 0.0000, 0.8470, 0.1860, and 2.4823, respectively.

This paper also constructed hyperspectral monitoring models of winter wheat Log(ChD) based on the bands selected by CARS. The results showed that the model constructed by ETsR, PLSR, SVR, RFR, SVR, and PLSR all had varying degrees of high calibration accuracy, but the calibration accuracy of the KNR model fluctuated significantly with increasing order. When using MLRP to construct models based on RL-FOD and Caputo FOD spectra, the calibration accuracy was extremely low, indicating serious underfitting during model construction, which made the model unusable. This indicated that MLPR was not suitable for constructing Log(ChD) hyperspectral monitoring models under these conditions. Excepted for models constructed by ETsR and RFR, which had relatively stable validation accuracy, the validation accuracy of other models had significant fluctuations. Among all models constructed based on CARS band, the model constructed by ETsR based on the 0.3-order RL-FOD spectrum had the highest accuracy, with R^{2c}, $RMSE_c$, R^{2v}, $RMSE_v$, and RPD of 1.0000, 0.0000, 0.8667, 0.1732, and 2.6660, respectively.

Through comprehensive comparison, it can be seen that both models based on full spectrum band and models based on CARS band achieved the highest monitoring accuracy when calculating FOD spectra using the RL method. It can be seen that compared to GL and Caputo methods, the RL method was more suitable for preprocessing hyperspectral reflectance data to improve model accuracy. However, if it was necessary to observe the gradient process from the original spectrum to the integer-order derivative spectrum, the GL method needed to be used for calculation. Among all the models constructed in this study, the original spectral curve was preprocessed with 0.3-order RL-FOD, combined with CARS, the model constructed by ETsR was the most accurate and could achieve hyperspectral monitoring of Log(ChD). This model only used 150 bands, reducing the number of bands by over 90% compared to the full spectrum band, greatly reducing the complexity of the model. For CARS, its advantage was that the number of bands retained during each iteration can be adjusted by adjusting the number of iterations, so that an appropriate number of bands can be selected to construct a model as the number of bands gradually decreases. However, due to the randomness involved in its calculation process, it was still possible to screen out some bands that were not related to ChD. At the same time, the number of bands screened by this research was still large, which makes the model may still have a high cost in actual use, thus limiting the use of the model. Therefore, in subsequent research, it may be considered to use algorithms such as the random frog algorithm to perform secondary screening on the screening results, while ensuring the accuracy of the model and further simplifying it (Chen et al., 2023). In addition, in order to improve the reliability and portability of the model, the accuracy of the model should also be further validated through the combination of unmanned aerial vehicles and other platforms in the follow-up research, so as to realize the rapid estimation of winter wheat ChD more efficiently.

5 Conclusion

This study focused on winter wheat, measured its ChD and corresponding canopy hyperspectral reflectance. By calculating Log (ChD), three FOD calculation methods, GL, RL, and Caputo, were used to preprocess the original spectral data. Based on the full spectrum band and CARS band, 8 machine learning algorithms, PLSR, SVR, MLPR, RFR, ETsR, DTR, KNR, and GPR, were used to construct hyperspectral monitoring models of winter wheat Log (ChD). The main conclusions were as follows:

1. GL-FOD can be used to observe the gradual change from the original spectral curve to the integer-order derivative spectral curve, while RL-FOD and Caputo-FOD spectra both exhibited varying degrees of changes in integer-order derivative spectra, but Caputo-FOD spectra were less sensitive to changes in order compared to GL-FOD and RL-FOD spectra. All three types of FOD spectra can improve the correlation between the original spectral curve and Log(ChD) to varying degrees, but only GL-FOD and RL-FOD can observe the change process of the

correlation between FOD spectra and Log(ChD) with changes in order, and it showed that the shorter the wavelength, the lower the order, and the higher the correlation.

2. The CARS method can remove most of the spectral band, and the screened bands were generally distributed throughout the entire spectral range, but had a relatively concentrated distribution in the visible light range.
3. The models constructed based on the full spectrum band and the CARS band all had the best accuracy and stability at using ETsR based on 0.3-order RL-FOD spectrum. Among the three FOD calculation methods, the RL method was more suitable for constructing hyperspectral monitoring models for winter wheat ChD. The CARS method can simplify the model while improving its accuracy. Among all models, the model based on the 0.3-order RL-FOD spectrum, using CARS to screen bands and combining with ETsR, had the highest accuracy. Its R^{2c}, $RMSE_c$, R^{2v}, $RMSE_v$, and RPD were 1.0000, 0.0000, 0.8667, 0.1732, and 2.6660, respectively, which can achieve hyperspectral monitoring of ChD in winter wheat.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The data that has been used is confidential. Requests to access these datasets should be directed to Chenbo Yang, ycb232008@sxau.edu.cn.

Author contributions

CY: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – original draft. MF: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. JB: Data curation, Investigation, Methodology, Software, Writing – review & editing. HS: Data curation, Investigation, Methodology, Software, Writing – review & editing. RB: Conceptualization, Methodology, Software, Supervision, Writing – review & editing. LS: Data curation, Investigation, Methodology, Software, Writing – review & editing. CW: Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. YZ: Conceptualization, Formal analysis, Methodology, Resources, Supervision, Writing – review & editing. WY: Conceptualization, Methodology, Resources, Supervision, Writing – review & editing. LX: Investigation, Methodology, Resources, Software, Writing – review & editing. MZ: Methodology, Resources, Software, Writing – review & editing. XS: Methodology, Resources, Software, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by Basic Research Program of Shanxi Province (20210302123411, 202203021211275), and Earmarked Fund for Modern Agro-industry Technology Research System (2024CYJSTX02-23).

Acknowledgments

Thanks to the experimental platform provided by Shanxi Agricultural University, as well as all editors and reviewers.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1492059/full#supplementary-material

SUPPLEMENTARY FIGURE S1
1:1 fitting figure of Log (ChD)

References

Alshammari, S., Alshammari, M., Alabedalhadi, M., Al-Sawalha, M. M., and Al-Smadi, M. (2024). Numerical investigation of a fractional model of a tumor-immune surveillance via Caputo operator. *Alexandria Eng. J.* 86, 525–536. doi: 10.1016/j.aej.2023.11.026

Arjmandi, M., Fattahi, M., Motevassel, M., and Rezaveisi, H. (2023). Evaluating algorithms of decision tree, support vector machine and regression for anode side catalyst data in proton exchange membrane water electrolysis. *Sci. Rep.* 13, 20309. doi: 10.1038/s41598-023-47174-w

Benkhettou, N., Cruz, A., and Torres, D. F. M. (2014). A fractional calculus on arbitrary time scales: fractional differentiation and fractional integration. *Mathematics* 107, 230–237. doi: 10.1016/j.sigpro.2014.05.026

Bhadra, S., Sagan, V., Maimaitijiang, M., Maimaitiyiming, M., Newcomb, M., Shakoor, N., et al. (2020). Quantifying leaf chlorophyll concentration of sorghum from hyperspectral data using derivative calculus and machine learning. *Remote Sens.* 12, 2082. doi: 10.3390/rs12132082

Chatterjee, A., and Kundu, S. (2015). Revisiting the chlorophyll biosynthesis pathway using genome scale metabolic model of Oryza sativa japonica. *Sci. Rep.* 5, 14975. doi: 10.1038/srep14975

Chen, S., Lou, F., Tuo, Y., Tan, S., Peng, K., Zhang, S., et al. (2023). Prediction of soil water content based on hyperspectral reflectance combined with competitive adaptive reweighted sampling and random frog feature extraction and the back-propagation artificial neural network method. *Water* 15, 2726. doi: 10.3390/w15152726

Cheng, J., Yang, G., Xu, W., Feng, H., Han, S., Liu, M., et al. (2022). Improving the estimation of apple leaf photosynthetic pigment content using fractional derivatives and machine learning. *Agronomy-Basel* 12, 1497. doi: 10.3390/agronomy12071497

Cozzolino, D., Williams, P. J., and Hoffman, L. C. (2023). An overview of pre-processing methods available for hyperspectral imaging applications. *Microchem. J.* 193, 109129. doi: 10.1016/j.microc.2023.109129

Feng, M. C., Wang, J. J., Wang, C., Yang, W. D., Wang, H. Q., Xiao, L. J., et al. (2013). Response of winter wheat (Triticum aestivum L.) spectral reflectance character istics to chlorophyll content in planting densities application rate. *Bulgarian J. Agric. Sci.* 19, 1190–1197.

Feng, Z., Guan, H., Yang, T., He, L., Duan, J., Song, L., et al. (2023). Estimating the canopy chlorophyll content of winter wheat under nitrogen deficiency and powdery mildew stress using machine learning. *Comput. Electron. Agric.* 211, 107989. doi: 10.1016/j.compag.2023.107989

Geng, J., Lv, J., Pei, J., Liao, C., Tan, Q., Wang, T., et al. (2024). Prediction of soil organic carbon in black soil based on a synergistic scheme from hyperspectral data: Combining fractional-order derivatives and three-dimensional spectral indices. *Comput. Electron. Agric.* 220, 108905. doi: 10.1016/j.compag.2024.108905

Graham Ram, B., Zhang, Y., Costa, C., Raju Ahmed, M., Peters, T., Jhala, A., et al. (2023). Palmer amaranth identification using hyperspectral imaging and machine learning technologies in soybean field. *Comput. Electron. Agric.* 215, 108444. doi: 10.1016/j.compag.2023.108444

Han, C., Liu, K., and Liu, Y. (2022). Application of nonlinear regression algorithm to aircraft load calibration test. *Eng. Test* 62, 31–35. doi: 10.3969/j.issn.1674-3407.2022.04.009

He, R., Li, H., Qiao, X., and Jiang, J. (2018). Using wavelet analysis of hyperspectral remote-sensing data to estimate canopy chlorophyll content of winter wheat under stripe rust stress. *Int. J. Remote Sens.* 39, 4059–4076. doi: 10.1080/01431161.2018.1454620

Huang, X., Guan, H., Bo, L., Xu, Z., and Mao, X. (2023). Hyperspectral proximal sensing of leaf chlorophyll content of spring maize based on a hybrid of physically based modelling and ensemble stacking. *Comput. Electron. Agric.* 208, 107745. doi: 10.1016/j.compag.2023.107745

Huang, H., Wang, W., Peng, Y., Wu, J., Gao, X., Wang, X., et al. (2010). Measurement of chlorophyll content in wheat leaves using hyperspectral scanning. *Spectrosc. Spectral Anal.* 30, 1811–1814. doi: 10.3964/j.issn.1000-0593(2010)07-1811-04

Hughes, G. F. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* 14, 55–63. doi: 10.1109/TIT.1968.1054102

Ji, S., Gu, C., Xi, X., Zhang, Z., Hong, Q., Huo, Z., et al. (2022). Quantitative monitoring of leaf area index in rice based on hyperspectral feature bands and ridge regression algorithm. *Remote Sens.* 14, 2777. doi: 10.3390/rs14122777

Jiang, J., Chen, Y., and Huang, W. (2010). Using hyperspectral remote sensing to estimate canopy chlorophyll density of wheat under yellow rust stress. *Spectrosc. Spectral Anal.* 30, 2243–2247. doi: 10.3964/j.issn.1000-0593(2010)08-2243-05

Jiang, Y., Zhang, L., Xue, P., and Wang, X. (2021). Development status of wheat industry in China and international experience for reference. *J. Agric. Sci. Technol.* 23, 1–10. doi: 10.13304/j.nykjdb.2021.0412

Kwon, S., Seo, I. W., Noh, H., and Kim, B. (2022). Hyperspectral retrievals of suspended sediment using cluster-based machine learning regression in shallow waters. *Sci. Total Environ.* 833, 155168. doi: 10.1016/j.scitotenv.2022.155168

Li, Z., Liu, L., Dehghan, S., Chen, Y., and Xue, D. (2017). A review and evaluation of numerical tools for fractional calculus and fractional order controls. *Int. J. Control* 90, 1165–1181. doi: 10.1080/00207179.2015.1124290

Li, C., Shi, J., Ma, C., Cui, Y., Wang, Y., and Li, Y. (2021a). Estimation of chlorophyll content in winter wheat based on wavelet transform and fractional differential. *Trans. Chin. Soc. Agric. Eng.* 52, 172–182. doi: 10.6041/j.issn.1000-1298.2021.08.017

Li, W., Sun, M., Zeng, F., and Wang, F. (2022). Hyperspectral estimation of chlorophyll content in winter wheat leaves under low temperature stress. *Chin. J. Agrometeorol.* 43, 137–147. doi: 10.3969/j.issn.1000-6362.2022.02.005

Li, C., Wang, Y., Ma, C., Ding, F., Li, Y., Chen, W., et al. (2021b). Hyperspectral estimation of winter wheat leaf area index based on continuous wavelet transform and fractional order differentiation. *Sensors* 21, 8497. doi: 10.3390/s21248497

Li, C., Xiao, Z., Liu, Y., Meng, X., Li, X., Wang, X., et al. (2023). Hyperspectral estimation of winter wheat leaf water content based on fractional order differentiation and continuous wavelet transform. *Agronomy-Basel* 13, 56. doi: 10.3390/agronomy13010056

Li, Y., and Yang, X. (2023). Quantitative analysis of near infrared spectroscopic data based on dual-band transformation and competitive adaptive reweighted sampling. *Spectrochimica Acta Part A: Mol. Biomol. Spectrosc.* 285, 121924. doi: 10.1016/j.saa.2022.121924

Liu, Y., Feng, H., Yue, J., Fan, Y., Jin, X., Zhao, Y., et al. (2022). Estimation of potato above-ground biomass using UAV-based hyperspectral images and machine-learning regression. *Remote Sens.* 14, 5449. doi: 10.3390/rs14215449

Liu, J., Li, Y., Zhao, F., and Liu, Y. (2023). Hyperspectral remote sensing images feature extraction based on spectral fractional differentiation. *Remote Sens.* 15, 2879. doi: 10.3390/rs15112879

Liu, X., Zhang, K., Zhang, Z., Cao, Q., Lv, Z., Yuan, Z., et al. (2017). Canopy chlorophyll density based index for estimating nitrogen status and predicting grain yield in rice. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01829

Ma, J., Cheng, J., Wang, J., Pan, R., He, F., Yan, L., et al. (2022). Rapid detection of total nitrogen content in soil based on hyperspectral technology. *Inf. Process. Agric.* 9, 566–574. doi: 10.1016/j.inpa.2021.06.005

MaChado, J. T., Kiryakova, V., and Mainardi, F. (2011). Recent history of fractional calculus. *Commun. Nonlinear Sci. Numerical Simulation* 16, 1140–1153. doi: 10.1016/j.cnsns.2010.05.027

Maraden, Y., Wibisono, G., Nugraha, I. G., Sudiarto, B., Jufri, F. H., Kazutaka,, et al. (2023). Enhancing electricity theft detection through K-nearest neighbors and logistic regression algorithms with synthetic minority oversampling technique: A case study on state electricity company (PLN) customer data. *Energies* 16, 5405. doi: 10.3390/en16145405

Nansen, C., Geremias, L. D., Xue, Y., Huang, F., and Parra, J. R. (2013). Agricultural case studies of classification accuracy, spectral resolution, and model over-fitting. *Appl. Spectrosc.* 67, 1332–1338. doi: 10.1366/12-06933

Qi, Y., Jiang, B., Lei, W., Zhang, Y., and Yu, W. (2024). Reliability analysis of normal, lognormal, and weibull distributions on mechanical behavior of wood scrimber. *Forests* 15, 1674. doi: 10.3390/f15091674

Sahoo, R. N., Gakhar, S., Rejith, R. G., Verrelst, J., Ranjan, R., Kondraju, T., et al. (2023). Optimizing the retrieval of wheat crop traits from UAV-borne hyperspectral image with radiative transfer modelling using gaussian process regression. *Remote Sens.* 15, 5496. doi: 10.3390/rs15235496

Silva, B., Prado, R., Baio, F. H. R., Campos, C. N. S., Teodoro, L. P. R., Teodoro, P. E., et al. (2024). New approach for predicting nitrogen and pigments in maize from hyperspectral data and machine learning models. *Remote Sens. Applications: Soc. Environ.* 33, 101110. doi: 10.1016/j.rsase.2023.101110

Singh, R., Dubey, A. K., and Kapoor, R. (2022). Deep neural network regularization (DNNR) on denoised image. *Int. J. Intelligent Inf. Technol. (IJIIT)* 18, 1–16. doi: 10.4018/IJIIT.309584

Sudu, B., Rong, G., Guga, S., Li, K., Zhi, F., Guo, Y., et al. (2022). Retrieving SPAD values of summer maize using UAV hyperspectral data based on multiple machine learning algorithm. *Remote Sens.* 14, 5407. doi: 10.3390/rs14215407

Sun, Y., Wang, J., Li, B., Liu, L., and Huang, W. (2007). Contrastive analysis based on neural network of winter wheat's chlorophyll concentration inversion with hyperspectral data. *Remote Sens. Technol. Appl.* 22, 492–496. doi: 10.11873/j.issn.1004-0323.2007.4.492

Tang, T., Chen, C., Wu, W., Zhang, Y., Han, C., Li, J., et al. (2023). Hyperspectral inversion model of relative heavy metal content in pennisetum sinese roxb via EEMD-db3 algorithm. *Remote Sens.* 15, 251. doi: 10.3390/rs15010251

Wang, T., Gao, M., Cao, C., You, J., Zhang, X., and Shen, L. (2022). Winter wheat chlorophyll content retrieval based on machine learning using in *situ* hyperspectral data. *Comput. Electron. Agric.* 193, 106728. doi: 10.1016/j.compag.2022.106728

Xie, Y., Feng, M., Wang, C., Yang, W., Sun, H., Yang, C., et al. (2020). Hyperspectral monitor on chlorophyll density in winter wheat under water stress. *Agron. J.* 112, 3667–3676. doi: 10.1002/agj2.20306

Xing, N., Huang, W., Ye, H., Dong, Y., Kong, W., Ren, Y., et al. (2022). Remote sensing retrieval of winter wheat leaf area index and canopy chlorophyll density at different growth stages. *Big Earth Data* 6, 580–602. doi: 10.1080/20964471.2021.1918909

Xu, H., Vavilin, D., and Vermaas, W. (2001). Chlorophyll b can serve as the major pigment in functional photosystem II complexes of cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* 98, 14168–14173. doi: 10.1073/pnas.251530298

Yang, C., Feng, M., Song, L., Jing, B., Xie, Y., Wang, C., et al. (2022a). Hyperspectral monitoring of soil urease activity under different water regulation. *Plant Soil* 477, 779–792. doi: 10.1007/s11104-022-05476-4

Yang, C., Feng, M., Song, L., Jing, B., Xie, Y., Wang, C., et al. (2022b). Study on hyperspectral monitoring model of soil total nitrogen content based on fractional-order derivative. *Comput. Electron. Agric.* 201, 107307. doi: 10.1016/j.compag.2022.107307

Yang, F., Liu, T., Wang, Q., Du, M., Yang, T., Liu, D., et al. (2021). Rapid determination of leaf water content for monitoring waterlogging in winter wheat based on hyperspectral parameters. *J. Integr. Agric.* 20, 2613–2626. doi: 10.1016/S2095-3119(20)63306-8

Yang, Y., Nan, R., Mi, T., Song, Y., Shi, F., Liu, X., et al. (2023c). Rapid and nondestructive evaluation of wheat chlorophyll under drought stress using hyperspectral imaging. *Int. J. Mol. Sci.* 24, 5825. doi: 10.3390/ijms24065825

Yang, C., Song, L., Wei, K., Gao, C., Wang, D., Feng, M., et al. (2023a). Study on hyperspectral monitoring model of total flavonoids and total phenols in Tartary buckwheat grains. *Foods* 12, 1354. doi: 10.3390/foods12071354

Yang, C., Xu, J., Feng, M., Bai, J., Sun, H., Song, L., et al. (2023b). Evaluation of hyperspectral monitoring model for aboveground dry biomass of winter wheat by using multiple factors. *Agronomy-Basel* 13, 983. doi: 10.3390/agronomy13040983

Yu, B., and Liang, Y. (2023). Approximation with continuous functions preserving fractal dimensions of the Riemann-Liouville operators of fractional calculus. *Fractional Calculus Appl. Anal.* 26, 2805–2836. doi: 10.1007/s13540-023-00215-7

Zhang, X., Sun, H., Qiao, X., Yan, X., Feng, M., Xiao, L., et al. (2022). Hyperspectral estimation of canopy chlorophyll of winter wheat by using the optimized vegetation indices. *Comput. Electron. Agric.* 193, 106654. doi: 10.1016/j.compag.2021.106654

Zhang, H., Zhao, J., Huang, L., Huang, W., Dong, Y., Ma, H., et al. (2024). Development of new indices and use of CARS-Ridge algorithm for wheat fusarium head blight detection using *in-situ* hyperspectral data. *Biosyst. Eng.* 237, 13–25. doi: 10.1016/j.biosystemseng.2023.11.009

Zhao, C., Wang, Z., Wang, J., Huang, W., and Guo, T. (2011). Early detection of canopy nitrogen deficiency in winter wheat (Triticum aestivum L.) based on hyperspectral measurement of canopy chlorophyll status. *New Z. J. Crop Hortic. Sci.* 39, 251–262. doi: 10.1080/01140671.2011.588713

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 29 January 2025
DOI 10.3389/fpls.2025.1516255

Check for updates

OPEN ACCESS

EDITED BY
Wenyu Zhang,
Jiangsu Academy of Agricultural Sciences Wuxi Branch, China

REVIEWED BY
Liying Chang,
Shanghai Jiao Tong University, China
Jiayi Zhang,
Jiangsu Academy of Agricultural Sciences Wuxi Branch (JAASWB), China

*CORRESPONDENCE
Masaaki Takahashi
takahashim088@affrc.go.jp

RECEIVED 24 October 2024
ACCEPTED 13 January 2025
PUBLISHED 29 January 2025

CITATION
Takahashi M, Kawasaki Y, Naito H, Lee U and Yoshi K (2025) Fruit size prediction of tomato cultivars using machine learning algorithms. *Front. Plant Sci.* 16:1516255. doi: 10.3389/fpls.2025.1516255

COPYRIGHT
© 2025 Takahashi, Kawasaki, Naito, Lee and Yoshi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Fruit size prediction of tomato cultivars using machine learning algorithms

Masaaki Takahashi[1]*, Yasushi Kawasaki[1], Hiroki Naito[1,2], Unseok Lee[1] and Koichi Yoshi[1]

[1]Research Center for Agricultural Robotics, National Agricultural and Food Research Organization (NARO), Tsukuba, Ibaraki, Japan, [2]Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan

Early fruit size prediction in greenhouse tomato (*Solanum lycopersicum* L.) is crucial for growers managing cultivars to reduce the yield ratio of small-sized fruit and for stakeholders in the horticultural supply chain. We aimed to develop a method for early prediction of tomato fruit size at harvest with machine learning algorithm, and three machine learning models (Ridge Regression, Extra Tree Regrreion, CatBoost Regression) were compared using the PyCaret package for Python. For constructing the models, the fruit weight estimated from the fruit diameter obtained over time for each cumulative temperature after anthesis was used as explanatory variable and the fruit weight at harvest was used as objective variable. Datasets for two different prediction periods after anthesis of three tomato cultivars ("CF Momotaro York," "Zayda," and "Adventure.") were used to develop tomato size prediction models, and their performance was evaluated. We also aimed to improve the model adding the average temperature during the prediction period as an explanatory variable. When the estimated fruit size data at cumulative temperatures of 200°C d, 300°C d, and 500°C d after anthesis were used as explanatory variables, the mean absolute percentage error (MAPE) was lowest for "Zayda," a cultivar with stable fruit diameter, at 9.8% for Ridge Regression. When the estimated fruit size at cumulative temperatures of 300°C d, 500°C d, and 800°C d after anthesis were used as explanatory variables for Ridge Regression, the MAPE decreased for all cultivars: 10.1% for "CF Momotaro York," 8.8% for "Zayda," and 10.0% for "Adventure." In addition, incorporating the average temperature during the fruit size prediction period as an explanatory variable slightly increased model performance. These results indicate that this method could effectively predict tomato size at harvest in three cultivars. If fruit diameter data acquisition could be automated or simplified, it would assist in cultivation management, such as tomato thinning.

KEYWORDS

size prediction, fruit grade, machine learning, diameter, tomato

1 Introduction

Fruit size and yield are crucial crop management considerations for horticultural fruit growers. These factors can vary based on weather conditions (Lötze and Bergh, 2004), crop load (Heuvelink, 1997; Naor et al., 2001), and responses to water and salt stress (Nuruddin et al., 2003; Itoh et al., 2020). Since dimensions, geometry, and fruit size are key determinations of fruit grade, for both growers and stakeholders in the horticultural supply chain, more precise predictions of these factors can enhance market value (Khojastehnazhand et al., 2010). Together with the total number of fruits, fruit size has a strong impact on yield estimation. Although technologies for predicting tomato yield using information on the growth environment and plant growth have been developed (Berrueta et al., 2020; Saito et al., 2020; Higashide, 2022), no methods for predicting tomato fruit size are currently available.

There are two lines of research into fruit size prediction. One uses a mathematical model based on environmental and crop growth data. Using this model, the average weekly error for cucumber fruit size was 6.6%, but at the end of the growing season, it was underestimated (Marcelis and Gijzen, 1998). Fruit size prediction can also be performed using mathematical models for peaches and tomatoes (Fishman and Génard, 2002; Liu et al., 2007). These techniques use preset parameters, which can lead to significant deviations from predictions when unexpected situations occur. The second is based on direct measurement methods with calipers and computer vision. Fruit size prediction at harvest based on measured dimensions over time during growth has also been done, but it is unsuitable for early-ripening apple cultivars (Zadravec et al., 2013). In greenhouses, counting fruits based on ripeness using visible images and estimating volume and surface area is easier (Zhao et al., 2016; Ziaratban et al., 2017; Gongal et al., 2018; Afonso et al., 2020; Lee et al., 2020; Ge et al., 2022). UAVs (Unmanned Aerial Vehicles) and robots could further improve yield prediction accuracy if the number of fruits in the entire field could be accurately quantified (Apolo-Apolo et al., 2020; Seo et al., 2021; Egi et al., 2022). By contrast, these techniques can only be used to predict the size of fruit close to harvest time. Given the antagonistic relationship between the size, composition, and number of fruits, predicting the size of fruits at an early stage is challenging even with current computer vision technology. Additionally, early fruit size prediction techniques are needed for artificial control, such as reducing the yield percentage of small-sized fruit.

Tomato fruits size is determined by cell number and cell size. Depending on the tomato cultivar, the division of pericarp cell progresses in a short period 12 to 25 days after anthesis (Bertin et al., 2009), and cell elongation continues until the start of fruit ripening (Giovannoni, 2004). Although most fruit volume increase occurs during cell elongation, final fruit size is highly correlated with the number of cells determined during early cell division (Bertin et al., 2009). After the middle stage of fruit enlargement, temperature strongly influenced the volume growth rate, which was lower at 14°C (low temperature) and 26°C (high temperature) compared to 18°C and 22°C, respectively (Adams, 2001). From the above, it is possible that the final size of harvested fruit can be predicted by two factors: the rate of volume increase and the temperature conditions, which are especially critical during the initial cell division phase of fruit enlargement and the middle growth stage. Fruit diameter has been widely used across various crops to estimate size and calculate the growth rate (Warrington et al., 1999,; Minchin et al., 2003; Tran et al., 2017). Using actual measured diameter data of fruits, fruit size can be estimated nondestructively over time and the rate of volume increase can be evaluated more accurately than by setting parameters in advance. It is also highly compatible with future use of computer vision-based technology.

In this study, we aimed to develop a technique for predicting the size of harvested fruit using the fruit size at the beginning and middle of the growth period, which we considered important for predicting harvested fruit size. To predict fruit size with high accuracy, we analyzed the data using various machine learning algorithms. We analyzed three tomato cultivars to identify the morphological characteristics those most adaptable to this prediction method and sought to enhance its precision by incorporating additional explanatory variables related to temperature.

2 Materials and methods

2.1 Plant materials and growth conditions

Plants were cultivated in a greenhouse (5.4 m width, 10.8 m length) at Tsukuba (36°01'N 140°05'E), Japan. Three tomato cultivars were used for this research: "CF momotaro York" (Takii & Co. Ltd., Kyoto, Japan), "Zayda," and "Adventure" (Rijk Zwaan, De Lier, the Netherlands). The cultivation experiment was conducted between August 16, 2022, and April 27, 2023. Tomato plants were transplanted on coconut shell medium (Coco-bag; Toyotane Co., Ltd., Aichi, Japan) with a plant density of 3.75 stems m^{-2}. Nine plants of each cultivar were surveyed up to about 15 or 16 trusses, with each truss limited to four fruit sets.

2.2 Tomato fruit data set acquisition method

Overview of the proposal methodology is shown in Figure 1. The first anthesis date of three tomato cultivars during the growing season was September 14, 2022, and they were continuously surveyed 2-3 times a week until March 6, 2023, for "CF momotaro York" and until February 27, 2023, for "Zayda," and "Adventure." The method of tomato data collection and integration is detailed in Figure 2. After fruit set, the long and short diameters of the fruit were measured once or twice a week and the data on the cumulative temperature after anthesis of each flower was recorded together (Figure 2A). These measurements continued until the fruit was harvested until April 27 for three cultivars. Tomatoes were assumed to be ellipsoids, and the fruit size of each cultivar was calculated from the long and short diameters using the following equation (Li et al., 2015).

10.3389/fpls.2025.1516255

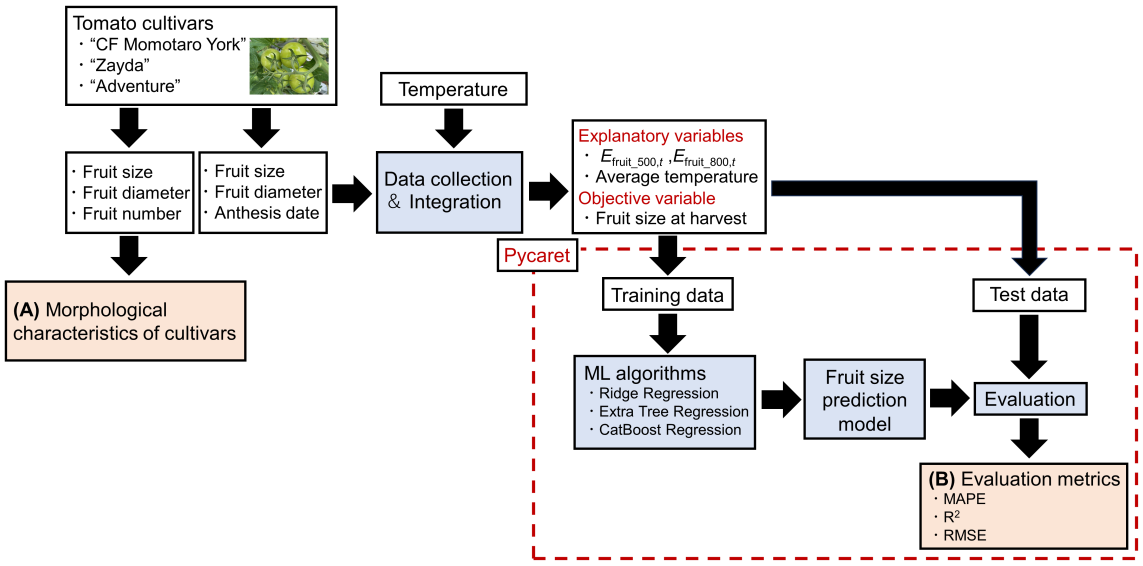


FIGURE 1
Overview of the proposal methodology. **(A)** Tomato size, fruit diameter and number of fruits were used in the analysis of morphological characteristics (Results Section 3.1). **(B)** The tomato fruit size prediction model was created using machine learning algorithms and evaluated (Results section 3.2, 3.3).

$$V_{\mathbf{fruit}} = \frac{4}{3}\ \pi\ \frac{l}{2}\ \left(\frac{s}{2}\right)^2 \quad (1)$$

$$C_{\mathbf{fruit}} = V_{\mathbf{fruit}}\ d \quad (2)$$

where V_{fruit} represents the fruit volume index (cm^3), l represents the long diameter (cm), and s represents the short diameter (cm), C_{fruit} represents the calculated fruit size (g), and d is the fruit density (g cm^{-3}). The fruit density for each tomato cultivar was determined through regression equations based on V_{fruit} and actual fruit sizes after harvesting or thinning at various growth stages, 435, 547 and 449 fruits for "CF Momotaro York," "Zayda," and "Adventure," respectively (Figure 2B). Using these equations, fruit size at each growth stage was calculated nondestructively from the diameters. As the date of anthesis differs for each tomato fruit and the temperature in the greenhouse is not constant, obtaining C_{fruit} at a specific cumulative temperature after anthesis is challenging. For each fruit, the C_{fruit} values at cumulative temperatures after anthesis were fitted to a third-order polynomial using Scientific Python (SciPy). A fitted curve ($r^2 \geq 0.94$) was created using C_{fruit} up to cumulative temperature after anthesis<625°C d, and the estimated

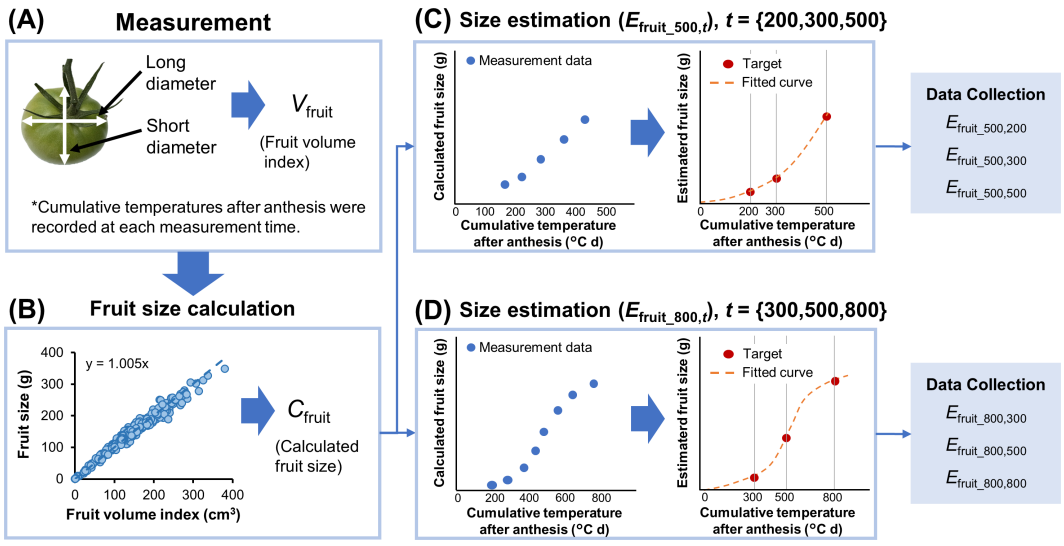


FIGURE 2
Proposed fruit size data collection and integration process. **(A)** Diameters of tomato fruit were measured at different cumulative temperature after anthesis of each flower, and V_{fruit} (Fruit volume index) was calculated. **(B)** The fruit density for each tomato cultivar was determined through regression equations based on V_{fruit} and actual fruit sizes, and C_{fruit} (the calculated fruit size) was calculated. **(C, D)** The estimated fruit size ($E_{fruit_500,t}$, $E_{fruit_800,t}$) for each cumulative temperature after anthesis was estimated from the fitted curve, and the data was collected.

fruit size at 200, 300, and 500°C d was obtained ($E_{fruit_500,200}$, $E_{fruit_500,300}$, $E_{fruit_500,500}$, Figure 2C). Similarly, a fitted curve ($r^2 \geq 0.95$) was created using C_{fruit} up to cumulative temperature after anthesis<900°C d, and the estimated fruit size at 300, 500, and 800° C d was obtained ($E_{fruit_800,300}$, $E_{fruit_800,500}$, $E_{fruit_800,800}$, Figure 2D). Since the pericarp cell division occurs 10–25 days after flowering and the cumulative temperature after anthesis is assumed to be<500°C d, fruit size ($E_{fruit_500,t}$) was collected during this period. Additionally, 2 weeks before harvest, the cumulative temperature after anthesis was assumed to be ≥800°C d, and fruit size data ($E_{fruit_800,t}$) was gathered to predict the harvest size by that time.

2.3 Model development

The data set was created using estimated fruit sizes ($E_{fruit_500,t}$, $E_{fruit_800,t}$) as explanatory variables and actual harvested fruit sizes as objective variables. For the estimated fruit size data sets ($E_{fruit_500,t}$), 401, 516, and 417 fruits were used for "CF Momotaro York," "Zayda," and "Adventure," respectively. For the estimated fruit size data sets ($E_{fruit_800,t}$), 404, 524, and 421 fruits were used for "CF Momotaro York," "Zayda," and "Adventure," respectively. The data analysis flow is shown in Figure 1. We used PyCaret 3.3, an open-source low code Python library that automates machine learning (AutoML) models (Moez, 2020). The library manages algorithms for regression and classification. The PyCaret library evaluates and compares these models based on specific metrics. Data were normalized using Z-score normalization and randomly divided into 80% for training and 20% for testing. Automated model selection was performed with PyCaret 3.3, where all existing regression models were trained and compared automatically based on the defined preprocessing pipeline for the dataset. From the recommended models, we selected and refined three: (1) Ridge Regression, (2) Extra Tree Regression, and (3) CatBoost Regression. Hyperparameters were optimized by repeated 10-fold cross-validation to maximize the determination coefficient (R^2) in the training data with grid search (PyCaret's default parameter). The model was then fitted to maximize the R^2 for all training data. These analyses were conducted using PyCaret. Test data were used to verify prediction accuracy. To evaluate model performance, we used MAPE, R^2, and root mean squared error (RMSE), which can be calculated as follows:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{P_i - H_i}{H_i}\right| \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(P_i - H_i)^2}{\sum_{i=1}^{n}(P_i - \bar{H})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - H_i)^2} \quad (5)$$

where n represents the number of observations, P_i represents the predicted tomato size, H_i represents the harvested tomato size, and $\bar{H}$ represents the mean harvested fruit size.

2.4 Model improvement

To improve the prediction model, a new dataset was created that incorporated not only fruit size at each cumulative temperature after anthesis but also the average temperature during the prediction period as an additional explanatory variable. Using this dataset, the predictive model was developed with Ridge Regression. The model's performance was evaluated by calculating the MAPE, R^2 and RMSE following the same procedures described in Material and Methods Section 2.3.

3 Results

3.1 Morphological characteristics of the tomato cultivars

We examined the morphological characteristics of the fruits from each cultivar (Table 1). "CF Momotaro York" and "Adventure" had larger fruit size, long diameter and short diameter compared to "Zayda." The standard deviation (SD) for fruit size, the ranking of cultivars from largest to smallest was "CF Momotaro York," "Adventure," and "Zayda." The number of fruit sets per truss for each cultivar was also analyzed (Figure 3). "Zayda" showed particularly stable fruit set, with no fruit loss observed at nodes 4, 5, 10, and 12 across a survey of nine plants. The fruit volume index (V_{fruit}) for each cultivar was calculated using the long and short diameters of the fruit (Figure 4). The results indicated that the calculated fruit densities (d) for "CF Momotaro York," "Zayda," and "Adventure" were 1.005, 1.072, and 0.970, respectively.

TABLE 1 Fruit size, long diameter, and short diameter of each tomato cultivar.

Cultivar	Fruit size (g) Mean ± SD	Long diameter (cm) Mean ± SD	Short diameter (cm) Mean ± SD
"CF Momotaro York"	154.3 ± 51.5a	6.7 ± 0.9a	6.4 ± 0.8a
"Zayda"	140.9 ± 27.4b	6.4 ± 0.5b	6.2 ± 0.5b
"Adventure"	152.0 ± 39.2a	6.7 ± 0.7a	6.6 ± 0.7a

Data are presented as the means ± SD. Different letters in a column indicate significant differences at the 1% level according to the Tukey–Kramer test.

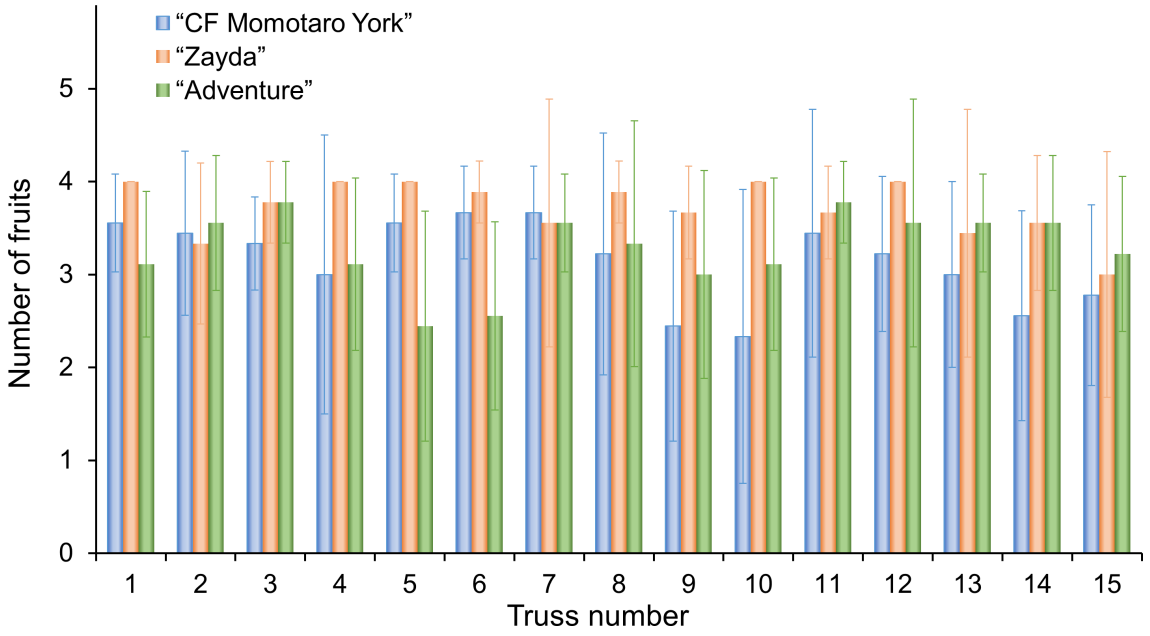


FIGURE 3
Number of fruits per truss for "CF momotaro York," "Zayda," and "Adventure." Vertical bars indicate the SD of the means (n = 9).

3.2 Evaluation of prediction models

Estimated fruit size data ($E_{fruit_500,t}$, $E_{fruit_800,t}$) were used as explanatory variables to predict actual harvest fruit size through machine learning, utilizing three different regression models. The MAPE, R^2, and RMSE of "CF Momotaro York," "Zayda," and "Adventure" are shown in Table 2, based on the predicted fruit size data. Ridge Regression consistently demonstrated stable and highly MAPE, R2, and RMSE for each cultivar. When $E_{fruit_500,t}$ was used as the explanatory variable, "Zayda" had the lowest MAPE values, followed by "Adventure," and "CF Momotaro York," regardless of the regression model. When $E_{fruit_800,t}$ was used, "Zayda" again had the lowest MAPE, with minimal differences between "CF Momotaro York" and "Adventure," both showing MAPE values around 10%. These results indicate that prediction accuracy improves as the fruit develops and that the performance of the models varies by cultivar.

3.3 Improvement of prediction models

We investigated the relationship between the average temperature and cumulative temperature from anthesis to harvest (Figure 5). The results indicated that for all cultivars, as the average temperature during the fruit growth period increased, the cumulative temperature until harvest decreased. The correlation coefficients were −0.681, −0.716, and −0.474 for "CF Momotaro York," "Zayda," and "Adventure," respectively, with a p-value of less than 0.01 for all cultivars, indicating strong statistical significance. In an effort to improve the predictive model, we incorporated data

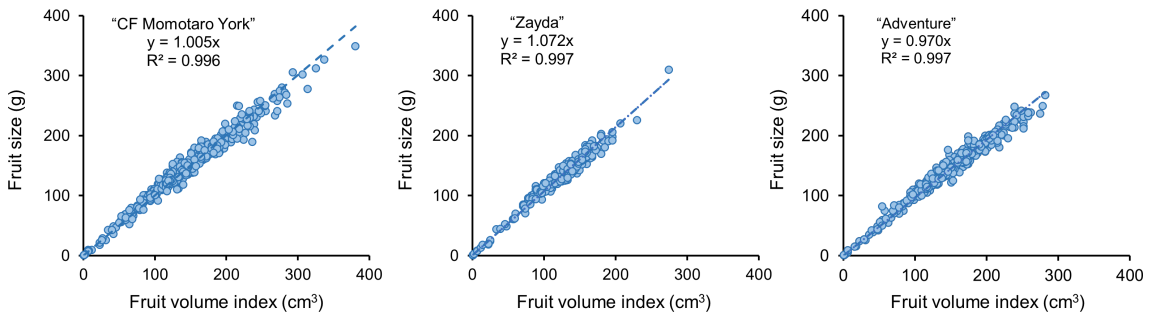


FIGURE 4
Relationship between fruit volume index (V_{fruit}) and actual fruit size in "CF momotaro York" (n = 435), "Zayda" (n = 547), and "Adventure" (n = 449).

TABLE 2 Results of fruit size prediction at harvest.

Cultivar	Model	MAPE (%)		R^2		RMSE	
		Explanatory variables[1]					
		$E_{fruit_500,t}$	$E_{fruit_800,t}$	$E_{fruit_500,t}$	$E_{fruit_800,t}$	$E_{fruit_500,t}$	$E_{fruit_800,t}$
"CF Momotaro York"	Ridge Regression	17.2	10.1	0.48	0.8	32.5	19.6
	Extra Tree Regrreion	17.1	9.7	0.51	0.81	31.9	19.3
	CatBoost Regression	17.0	10.0	0.44	0.78	33.8	20.6
"Zayda"	Ridge Regression	9.8	8.8	0.58	0.75	19.1	13.6
	Extra Tree Regrreion	10.0	8.9	0.43	0.71	22.4	14.8
	CatBoost Regression	10.3	8.5	0.42	0.74	22.5	14.1
"Adventure"	Ridge Regression	15.5	10.0	0.57	0.81	27.1	16.4
	Extra Tree Regrreion	16.5	10.3	0.44	0.77	30.9	18.0
	CatBoost Regression	16.7	9.6	0.48	0.81	26.7	16.7

[1]Explanatory variables are those used for fruit size predictions. "$E_{fruit_500,t}$" represents the estimated fruit size at cumulative temperatures after 200°C, 300°C and 500°C d anthesis were used an explanatory variable. "$E_{fruit_800,t}$" represents the estimated fruit size at cumulative temperatures after 300°C, 500°C and 800°C d anthesis was used as explanatory variable.

showing that the time to harvest varies depending on the average temperature during fruit enlargement. By adding average temperature during the prediction period as an explanatory variable for fruit size prediction, the model's performance improved for some cultivars and specific periods, however, the overall improvement was not substantial (Table 3).

4 Discussion

4.1 Morphological characteristics suitable for fruit size prediction

In this study, the morphological characteristics of the fruits of "CF Momotaro York" (a Japanese cultivar), "Zayda," and "Adventure" (Dutch cultivars) were found to be differ from one another. Although the number of fruit sets per stem was limited to four in this experiment, "Zayda" exhibited the smallest fruit size. SD in fruit size was smallest for "Zayda," followed by "Adventure,"and "CF Momotaro York." "Zayda" showed the most stable fruit set (Table 1, Figure 3). The initial fruit growth rate and size differed depending on fruit load in a tomato truss (Bertin, 2005). These minimal fluctuations in the number of fruits suggest low variation in the morphological characteristics of fruits in "Zayda," which may explain the high prediction accuracy and the machine learning model's ability to perform well even at early stages of cumulative temperature after anthesis (Tables 2, 3). When $E_{fruit_500,t}$ was used as an explanatory variable, "Adventure" outperformed "CF Momotaro York" in terms of MAPE and RMSE in three machine learning algorithms, also indicating that variation in fruit size can influence prediction accuracy (Table 2). It is well-known that the number of days to harvest and fruit dry matter weight in tomatoes can vary depending on the growing season (Heuvelink, 1995b). In this study, the cumulative temperature from anthesis to harvest varied significantly with changes in the average temperature during this period (Figure 5), resulting in corresponding changes in fruit size (data not shown). This

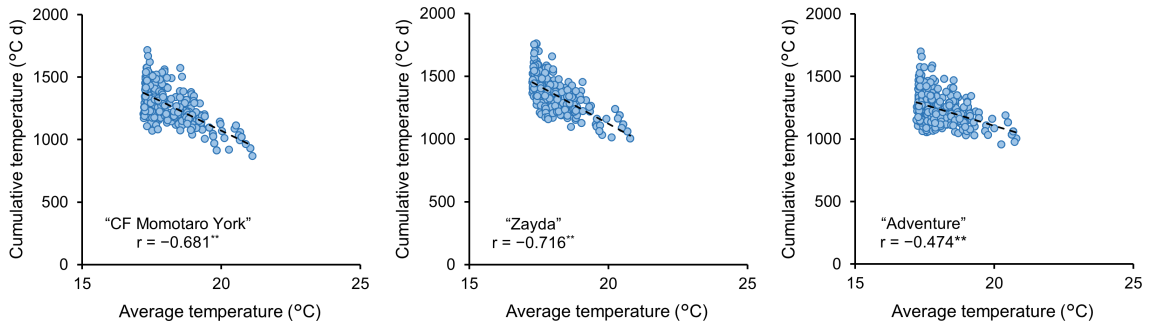


FIGURE 5
Relationship between average temperature and cumulative temperature from anthesis to harvest in "CF momotaro York" (n = 416), "Zayda" (n = 530), and "Adventure" (n = 431). **: significant negative correlation between average temperature and cumulative temperature (P< 0.01). Periods over which average temperature and cumulative temperature are calculated from anthesis to harvest date for each fruit.

10.3389/fpls.2025.1516255

TABLE 3 Results of fruit size prediction with the addition of average temperature as an explanatory variable.

Cultivar	Model	Average temperature[1]	MAPE (%)		R^2		RMSE	
			Explanatory variables[2]					
			$E_{fruit_500,t}$	$E_{fruit_800,t}$	$E_{fruit_500,t}$	$E_{fruit_800,t}$	$E_{fruit_500,t}$	$E_{fruit_800,t}$
"CF Momotaro York"	Ridge Regression	+	17.2	9.8	0.50	0.82	32.1	19.0
		–	17.2	10.1	0.48	0.80	32.5	19.6
"Zayda"		+	9.4	7.6	0.63	0.81	17.9	12.0
		–	9.8	8.8	0.58	0.75	19.1	13.6
"Adventure"		+	14.8	10.0	0.59	0.81	26.3	16.3
		–	15.5	10.0	0.57	0.81	27.1	16.4

[1]Average temperature represents the period from anthesis to the recorded fruit diameter. "+" and "–" indicate whether the variable was used as an explanatory factor or not. [2]Explanatory variables used for fruit size predictions. "$E_{fruit_500,t}$" represents the estimated fruit size at cumulative temperatures after 200°C, 300°C and 500°C d anthesis were used as explanatory variable. "$E_{fruit_800,t}$" represents the estimated fruit size at cumulative temperatures after 300°C, 500°C and 800°C d anthesis was used as explanatory variable.

suggests that prediction accuracy could be further enhanced by selecting cultivars optimized for fruit size prediction and developing seasonal models that align with fruit size trends at harvest.

4.2 Improvement of the model by adding average temperature

Given that average temperature significantly affects the time to harvest, we aimed to enhance the accuracy of our prediction model by incorporating average temperature during the fruit size estimation as an explanatory variable. Previous research found that the influence of maximum fruit diameter was minimal within the stable temperature range typically maintained in commercial greenhouses (Tijskens et al., 2016). Earlier studies demonstrated that tomatoes grown at varying temperatures of 14°C, 18°C, 22°C, and 26°C experienced longer harvest times at lower temperatures, with smaller fruit sizes observed at both 14°C and 26°C (Adams, 2001). This suggests that different average temperatures can lead to varying fruit growth stages when cumulative temperatures of 500°C d and 800°C d after anthesis are reached. While incorporating temperature data improved the accuracy of certain predictions, the overall improvement was not statistically significant (Table 3). The average temperature during data collection should be recorded, as the cumulative temperature after anthesis is essential for making these predictions. Thus, it is recommended that this data be included in future datasets for operational use.

4.3 Use of predictive technologies

In Japan, tomatoes are graded based on their shape and size, but there is no unified standard for grading; it varies depending on the cultivar, growing region, and consumer preferences. Generally, consumers tend to prefer slightly larger tomatoes, while smaller ones are often considered unmarketable. To produce larger, high-value fruits, tomato cultivation involves limiting the number of fruits per plant (Cockshull and Ho, 1995; Heuvelink, 1995a). Early prediction of fruit size distribution can optimize crop management practices to reduce fruit load. With this in mind, statistical methods have been employed to develop predictive models for fruit size. For example, a model for kiwifruit was created to estimate fruit size based on fruit diameter measurements (Minchin et al., 2003). For effective crop management, predicting model for the size distribution of the harvested fruit have been developed for apples, pears and citrus (Lötze and Bergh, 2004; Tanimoto and Yoshida, 2024). Pruning and thinning can be used to manipulate the crop to achieve the desired size distribution at harvest (Guardiola and García-Luis, 2000). In this study, to estimate the size of individual tomato fruits and apply it to crop management, fruit size predictions were based on the estimated fruit size ($E_{fruit_500,t}$) at cumulative temperatures after anthesis, ranging from 200°C d to 500°C d. For three tomato cultivars, MAPE was below about 17%, which allowed predicting size distribution based on measurements during the early stage of fruit enlargement. To facilitate rapid information sharing with supply chain stakeholders, fruit size predictions were made based on the estimated fruit size ($E_{fruit_800,t}$) at cumulative temperatures after anthesis, ranging from 300°C d to 800°C d. Under the assumption of an average temperature of 18°C d, a cumulative temperature of 800°C d after anthesis would typically correspond to about two weeks before harvest in most growing seasons. When the average temperature drops below 18°C, however, the cumulative temperature required to reach harvest increases (Figure 5). Predicting fruit size at this later stage of accumulated temperature still allows for timely information sharing before shipment. Depending on the season, it may be necessary to adjust the amount of data over time for the explanatory variables to improve prediction accuracy.

4.4 Automation of data collection

In this study, fruit diameter data was collected using calipers. To apply this approach in practical production settings, fruit diameter

data will need to be obtained through image analysis. Accurate sensing of flowering dates will also be crucial, as fruit diameter is closely linked to cumulative temperature after anthesis. These technologies, aimed at achieving high precision, are currently under development (Nyalala et al., 2019; Lee et al., 2022). Computer vision systems can be used to quantify increases in fruit diameter, length, and volume (Song et al., 2014; Mildenhall et al., 2021). In this study, the fruit size in each cultivar was calculated based on the long and short diameters, but predicting fruit size using just one diameter would facilitate the process. The coefficient of determination between harvested fruit size and long fruit diameter was 0.91, 0.92, and 0.94 for "CF Momotaro York," "Zayda," and "Adventure," respectively (data not shown). Although using only long diameters provides less accuracy compared to using both long and short diameters, focusing on a single diameter measurement may simplify data collection using computer vision. Automatic measurement of fruit diameter offers the advantage of high-frequency data collection. Frequent measurements through computer vision could enable earlier predictions of fruit size at harvest, potentially even sooner than the cumulative temperature benchmarks used in this study, which would reduce the labor required for monitoring. This technology will allow growers to adjust thinning practices to ensure the proportion of small-sized fruits at harvest remains low. In the future, integrating automated fruit diameter data collection with fast, predictive crop management strategies will be key to improving production efficiency.

5 Conclusion

In this study, we proposed a method for predicting the tomato fruits size at harvest time by analyzing time-series data on the diameter of tomato fruits at cumulative temperatures after anthesis using machine learning algorithms. We developed a model that can be used for cultivation management, such as fruit thinning, using fruit diameter data from the early fruit growth period, and a model that can be used for information sharing with supply chain stakeholders, assuming a prediction two weeks before harvest. The MAPE for fruit size prediction of the three tomato cultivars ranged from 9.8% to 17.2% and from 8.5% to 10.3%, respectively, and these predictions could be used for tomato producers. The difference in accuracy between cultivars was related to the SD of fruit size and diameter, and the prediction accuracy was higher when used with cultivar that had less variation in individual fruits. By adding the average temperature during the fruit size prediction period as an explanatory variable, in addition to fruit size, the performance improved depending on the cultivar and period. In this study, fruit diameter was measured using a caliper, but in the future, we are planning to use computer vision to measure diameter more frequently in order to estimate harvest size even earlier than the reference value for cumulative temperature used in this study. This will enable tomato producers to adjust thinning work and improve production efficiency and profitability.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

MT: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. YK: Methodology, Resources, Writing – review & editing. HN: Formal analysis, Methodology, Writing – review & editing. UL: Methodology, Writing – review & editing. KY: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by JSPS KAKENHI Grant Number JP22K20607.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adams, S. (2001). Effect of temperature on the growth and development of tomato fruits. *Ann. Bot.* 88, 869–877. doi: 10.1006/anbo.2001.1524

Afonso, M., Fonteijn, H., Fiorentin, F. S., Lensink, D., Mooij, M., Faber, N., et al. (2020). Tomato fruit detection and counting in greenhouses using deep learning. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.571299

Apolo-Apolo, O. E., Martínez-Guanter, J., Egea, G., Raja, P., and Pérez-Ruiz, M. (2020). Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV. *Eur. J. Agron.* 115, 126030. doi: 10.1016/j.eja.2020.126030

Berrueta, C., Heuvelink, E., Giménez, G., and Dogliotti, S. (2020). Estimation of tomato yield gaps for greenhouse in Uruguay. *Sci. Hortic.* 265, 109250. doi: 10.1016/j.scienta.2020.109250

Bertin, N. (2005). Analysis of the tomato fruit growth response to temperature and plant fruit load in relation to cell division, cell expansion and DNA endoreduplication. *Ann. Bot.* 95, 439–447. doi: 10.1093/aob/mci042

Bertin, N., Causse, M., Brunel, B., Tricon, D., and Genard, M. (2009). Identification of growth processes involved in QTLs for tomato fruit size and composition. *J. Exp. Bot.* 60, 237–248. doi: 10.1093/jxb/ern281

Cockshull, K. E., and Ho, L. C. (1995). Regulation of tomato fruit size by plant density and truss thinning. *J. Hortic.* 70, 395–407. doi: 10.1080/14620316.1995.11515309

Egi, Y., Hajyzadeh, M., and Eyceyurt, E. (2022). Drone-computer communication based tomato generative organ counting model using YOLO V5 and Deep-Sort. *Agriculture* 12, 1290. doi: 10.3390/agriculture12091290

Fishman, S., and Génard, M. (2002). A biophysical model of fruit growth: simulation of seasonal and diurnal dynamics of mass. *Plant Cell Environ.* 21, 739–752. doi: 10.1046/j.1365-3040.1998.00322.x

Ge, Y., Lin, S., Zhang, Y., Li, Z., Cheng, H., Dong, J., et al. (2022). Tracking and counting of tomato at different growth period using an improving YOLO-deepsort network for inspection robot. *Machines* 10, 489. doi: 10.3390/machines10060489

Giovannoni, J. J. (2004). Genetic regulation of fruit development and ripening. *Plant Cell* 16, S170–S180. doi: 10.1105/tpc.019158

Gongal, A., Karkee, M., and Amatya, S. (2018). Apple fruit size estimation using a 3D machine vision system. *Inf. Process. Agric.* 5, 498–503. doi: 10.1016/j.inpa.2018.06.002

Guardiola, J. L., and García-Luis, A. (2000). Increasing fruit size in Citrus. Thinning and stimulation of fruit growth. *Plant Growth Regul.* 31, 121–132. doi: 10.1023/A:1006339721880

Heuvelink, E. (1995a). Effect of plant density on biomass allocation to the fruits in tomato (Lycopersicon esculentum Mill.). *Sci. Hortic.* 64, 193–201. doi: 10.1016/0304-4238(95)00839-X

Heuvelink, E. (1995b). Growth, development and yield of a tomato crop: periodic destructive measurements in a greenhouse. *Sci. Hortic.* 61, 77–99. doi: 10.1016/0304-4238(94)00729-Y

Heuvelink, E. (1997). Effect of fruit load on dry matter partitioning in tomato. *Sci. Hortic.* 69, 51–59. doi: 10.1016/S0304-4238(96)00993-4

Higashide, T. (2022). Review of dry matter production and growth modelling to improve the yield of greenhouse tomatoes. *Hortic. J.* 91, 247–266. doi: 10.2503/hortj.UTD-R019

Itoh, M., Goto, C., Iwasaki, Y., Sugeno, W., Ahn, D.-H., and Higashide, T. (2020). Production of high soluble solids fruits without reducing dry matter production in tomato plants grown in salinized nutrient solution controlled by electrical conductivity. *Hortic. J.* 89, 403–409. doi: 10.2503/hortj.UTD-148

Khojastehnazhand, M., Omid, M., and Tabatabaeefar, A. (2010). Determination of tangerine volume using image processing methods. *Int. J. Food Prop.* 13, 760–770. doi: 10.1080/10942910902894062

Lee, J., Nazki, H., Baek, J., Hong, Y., and Lee, M. (2020). Artificial intelligence approach for tomato detection and mass estimation in precision agriculture. *Sustainability* 12, 9138. doi: 10.3390/su12219138

Lee, U., Islam, M. P., Kochi, N., Tokuda, K., Nakano, Y., Naito, H., et al. (2022). An automated, clip-type, small internet of things camera-based tomato flower and fruit monitoring and harvest prediction system. *Sensors* 22, 2456. doi: 10.3390/s22072456

Li, T., Heuvelink, E., and Marcelis, L. F. M. (2015). Quantifying the source–sink balance and carbohydrate content in three tomato cultivars. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00416

Liu, H. F., Genard, M., Guichard, S., and Bertin, N. (2007). Model-assisted analysis of tomato fruit growth in relation to carbon and water fluxes. *J. Exp. Bot.* 58, 3567–3580. doi: 10.1093/jxb/erm202

Lötze, E., and Bergh, O. (2004). Early prediction of harvest fruit size distribution of an apple and pear cultivar. *Sci. Hortic.* 101, 281–290. doi: 10.1016/j.scienta.2003.11.006

Marcelis, L. F. M., and Gijzen, H. (1998). Evaluation under commercial conditions of a model of prediction of the yield and quality of cucumber fruits. *Sci. Hortic.* 76, 171–181. doi: 10.1016/S0304-4238(98)00156-3

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2021). NeRF: representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 99–106. doi: 10.1145/3503250

Minchin, P. E. H., Richardson, A. C., Patterson, K. J. , and Martin, P. J. (2003). Prediction of final weight for Actinidia chinensis'Hort1 6A' fruit. *N. Z. J. Crop Hortic. Sci.* 31, 147–157. doi: 10.1080/01140671.2003.9514247

Moez, A. (2020). *PyCaret: An open source, low-code machine learning library in Python*. Available online at: https://www.pycaret.org (Accessed July 9, 2024).

Naor, A., Hupert, H., Greenblat, Y., Peres, M., Kaufman, A., and Klein, I. (2001). The response of nectarine fruit size and midday stem water potential to irrigation level in stage III and crop load. *J. Amer. Soc Hortic. Sci.* 126, 140–143. doi: 10.21273/JASHS.126.1.140

Nuruddin, M. M., Madramootoo, C. A., and Dodds, G. T. (2003). Effects of water stress at different growth stages on greenhouse tomato yield and quality. *HortScience* 38, 1389–1393. doi: 10.21273/HORTSCI.38.7.1389

Nyalala, I., Okinda, C., Nyalala, L., Makange, N., Chao, Q., Chao, L., et al. (2019). Tomato volume and mass estimation using computer vision and machine learning algorithms: Cherry tomato model. *J. Food. Eng.* 263, 288–298. doi: 10.1016/j.jfoodeng.2019.07.012

Saito, T., Kawasaki, Y., Ahn, D.-H., Ohyama, A., and Higashide, T. (2020). Prediction and improvement of yield and dry matter production based on modeling and non-destructive measurement in year-round greenhouse tomatoes. *Hortic. J.* 89, 425–431. doi: 10.2503/hortj.UTD-170

Seo, D., Cho, B.-H., and Kim, K.-C. (2021). Development of monitoring robot system for tomato fruits in hydroponic greenhouses. *Agronomy* 11, 2211. doi: 10.3390/agronomy11112211

Song, Y., Glasbey, C. A., Horgan, G. W., Polder, G., Dieleman, J. A., and van der Heijden, G. W. A. M. (2014). Automatic fruit recognition and counting from multiple images. *Biosyst. Eng.* 118, 203–215. doi: 10.1016/j.biosystemseng.2013.12.008

Tanimoto, Y., and Yoshida, S. (2024). A method of constructing models for estimating proportions of citrus fruit size grade using polynomial regression. *Agronomy* 14, 174. doi: 10.3390/agronomy14010174

Tijskens, L. M. M., Unuk, T., Okello, R. C. O., Wubs, A. M., Šuštar, V., Šumak, D., et al. (2016). From fruitlet to harvest: Modelling and predicting size and its distributions for tomato, apple and pepper fruit. *Sci. Hortic.* 204, 54–64. doi: 10.1016/j.scienta.2016.03.036

Tran, D., Hertog, M. L. A. T. M., Tran, T. L. H., Quyen, N. T., Van de Poel, B., Mata, C. I., et al. (2017). Population modeling approach to optimize crop harvest strategy. The case of field tomato. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00608

Warrington, I., Fulton, T., Halligan, E., and De Silva, H. (1999). Apple fruit growth and maturity are affected by early season temperatures. *J. Amer. Soc Hortic. Sci.* 124, 468–477. doi: 10.21273/JASHS.124.5.468

Zadravec, P., Veberic, R., Stampar, F., Eler, K., and Schmitzer, V. (2013). Fruit size prediction of four apple cultivars: Accuracy and timing. *Sci. Hortic.* 160, 177–181. doi: 10.1016/j.scienta.2013.05.046

Zhao, Y., Gong, L., Zhou, B., Huang, Y., and Liu, C. (2016). Detecting tomatoes in greenhouse scenes by combining AdaBoost classifier and colour analysis. *Biosyst. Eng.* 148, 127–137. doi: 10.1016/j.biosystemseng.2016.05.001

Ziaratban, A., Azadbakht, M., and Ghasemnezhad, A. (2017). Modeling of volume and surface area of apple from their geometric characteristics and artificial neural network. *Int. J. Food Pro.* 20, 762–768. doi: 10.1080/10942912.2016.1180533

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 12 February 2025
DOI 10.3389/fpls.2025.1520297

Check for updates

OPEN ACCESS

EDITED BY
Liujun Xiao,
Nanjing Agricultural University, China

REVIEWED BY
Chenglong Huang,
Huazhong Agricultural University, China
Yue Mu,
Nanjing Agricultural University, China

*CORRESPONDENCE
Xinyu Guo
guoxy73@163.com
Chunjiang Zhao
zhaocj@nercita.org.cn

†These authors have contributed equally to this work

RECEIVED 31 October 2024
ACCEPTED 20 January 2025
PUBLISHED 12 February 2025

CITATION
Song H, Wen W, Zhang Y, Zhao Y, Guo X and Zhao C (2025) Two-dimensional semantic morphological feature extraction and atlas construction of maize ear leaves.
Front. Plant Sci. 16:1520297.
doi: 10.3389/fpls.2025.1520297

COPYRIGHT
© 2025 Song, Wen, Zhang, Zhao, Guo and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Two-dimensional semantic morphological feature extraction and atlas construction of maize ear leaves

Hongli Song[1,2†], Weiliang Wen[2,3†], Ying Zhang[2,3], Yanxin Zhao[4], Xinyu Guo[2,3*] and Chunjiang Zhao[1,2,3*]

[1]College of Information Engineering, Northwest A&F University, Yangling, China, [2]Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China, [3]Beijing Key Lab of Digital Plant, National Engineering Research Center for Information Technology in Agriculture, Beijing, China, [4]Beijing Key Laboratory of Maize DNA (DeoxyriboNucleic Acid) Fingerprinting and Molecular Breeding, Maize Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

Maize ear leaves have important roles in photosynthesis, nutrient partitioning and hormone regulation. The morphological and structural variations observed in maize ear leaves are numerous and contribute significantly to the yield. Nevertheless, research on the fine-scale morphology of maize leaves is less, particularly the quantitative methods to characterize the morphology of leaves in two-dimensional (2D) space is absent. This makes it challenging to accurately identify 2D leaf shape of their cultivars. Therefore, this study presents the methods of 2D semantic morphological feature extraction and atlas construction, with the ear leaf in silking stage of maize association analysis population serving as an example. A three-dimensional (3D) digitizer was employed to obtain data from 1,431 leaves belonging to 518 inbred lines. The data was then processed using mesh subdivision and planar parameterization to create 2D leaf models with area-preserving characteristics. Additionally, averaged 2D leaf models of all the inbred lines were constructed, and 29 2D leaf features were quantified. Based on this, 11 features were extracted as semantic features of 2D leaf shape through clustering and correlation analysis. A comprehensive 2D leaf shape indicator L_{2D} based on the 11 semantic features was proposed, and a 2D leaf shape atlas was constructed in accordance with the L_{2D} ordering. Inbred line identification of 2D leaf shape in maize was achieved using the atlas. The results of maize leaf inbred line identification can determine the probability that the corresponding true inbred line ranked within the top 10 of the predicted results is 0.706, within the top 20 is 0.810, and within the top 45 is 0.900. This enables the generation of the corresponding maize 2D leaf shape through the matching of semantic features. The methodology presented in this study offers novel insights into the construction of semantic models for the morphology of maize and the identification of cultivars. It also provides a theoretical and technical foundation for the generation and drawing the leaf shape based on semantic 2D morphological and structural features.

KEYWORDS

maize, two-dimensional, leaf shape, phenotyping, semantic features

1 Introduction

Plant architecture (Godin, 2000) is used to describe the observable characteristics of an individual or parts of plant. These characteristics encompass leaf shape, stem morphology, tassel structure, and root architecture, etc. Plant architecture exhibit variation among different species, different cultivars of the same species, and even among different individuals of the same cultivar. These differences are closely associated with the diverse genetic makeup. To a certain extent, plant architecture can be used as a group of indicators of the growth status of plants, which is of great significance to agricultural research and production. The plant architecture (Jafari et al., 2024) can be described by phenotypic big data, including topology, spatial geometric relationships of multiple organs in the multi-dimensional perspective. Leaf shape constitutes an essential element of plant architecture, encompassing aspects such as leaf spatial disposition, local blade folds, leaf margin amplitude, and others. The discrepancies in leaf shape directly influence the canopy structure and radiation utilization efficiency.

Maize is an important food and energy crop. The rapid acquisition of maize leaf shapes and analyzing the differences among cultivars is an important component of maize breeding (Tian et al., 2019) and dense density planting for high-yield cultivation (Li et al., 2021). In earlier research, simple one-dimensional (1D) measurement was the predominant way for obtaining leaf shape data, including leaf length, width, and aspect ratio. In further research of leaf shape, the researchers placed the maize leaf on a horizontal plane and attempted to flatten it as much as possible, then they used equipment, including RGB cameras (Han et al., 2014; Robil et al., 2021), hyperspectral cameras (Gao et al., 2021), etc., to capture images of the leaf from a position perpendicular to the horizontal plane. These images were used to extract 2D phenotypic features including the leaf area, leaf profile perimeter, etc. However, the process of flattening results loss 3D structural information, such as leaf bending and folding, which is not conducive to more in-depth research of the leaf shape from the 3D perspective. The advent of LiDAR, 3D digitizers, multi-view stereo (MVS), and other 3D data acquisition technologies has made the complete acquisition of the 3D leaf morphology feasible. The research of the leaf shape based on 3D data, in addition to obtaining the 1D and 2D phenotypic features, can also obtain 3D phenotypic features, such as the leaf curvature and the inclination angle (Bailey and Mahaffee, 2017; Hosoi et al., 2011). Nevertheless, the point cloud data obtained through LiDAR and other 3D techniques (Chen et al., 2023) necessitate intricate processing to generate 3D models comprising semantic information and to extract more 3D leaf shape features (Su et al., 2018; Wen et al., 2024b). In contrast, 3D digitizers are capable of directly acquiring data containing semantic information through the manual operation of the instrument acquisition method (Wen et al., 2021). The utilization of digitizers for the acquisition of 3D structural data pertaining to plants has been a subject of extensive research and development over a long time. This methodology has been employed in the investigation of diverse plant organs, including roots (Danjon and Reubens, 2008; Wu et al., 2015), leaves (Zheng et al., 2022), and the entire plant (Zheng et al., 2008).

The quantitative analysis of crop leaf shape features is closely related to the means of data acquisition. Wang et al. (2022) first segmented the foreground and background by converting color images to greyscale using an adaptive thresholding algorithm. After segmentation, features such as color and compactness were extracted in addition to conventional phenotypic features such as length, width and surface area. Wu et al. (2022) extracted 3D structure-related leaf features such as leaf inclination, leaf curvature, etc. from a 3D model of a maize leaf acquired by a 3D digitizer. Wen et al. (2024b) achieved mesh generation from the 3D point cloud of maize leaves by a series meshing technologies. The ARAP algorithm was used to convert the 3D mesh into 2D planar mesh, and finally both the 2D and 3D semantic leaf mesh model was generated. Based on the 3D semantic leaf model, 3D leaf area, leaf surface flatness and other phenotypic features that cannot be obtained based on 2D images can be extracted. Wen et al. (2024a) extracted 3D leaf features such as leaf inclination angle, blade-included angle, blade self-twisting, blade planarity, margin amplitude from a maize leaf model acquired by a 3D digitizer. In addition to maize leaf phenotypes, there are also related 3D quantitative analysis studies on wheat leaf phenotypes (Zheng et al., 2022). The current rapid development of 2D and 3D acquisition techniques and related quantitative analysis techniques has made it possible to acquire crop leaf-shape data in multiple dimensions and with high quality, thus enabling the completion of the analysis. The extraction of phenotypic features from crop leaves allows researchers to mine related physiological and ecological knowledge. In the previous research, Wang et al. (2022) validated the correlation between the sixth leaf sheath color phenotypic traits and the corresponding candidate genes by integrating the leaf sheath phenotypic data of the maize association analysis population (Yang et al., 2011) through GWAS. Wu et al. (2022) proposed a novel classification method based on the spatial morphology of the midvein curves in maize leaves. The analysis was also performed using GWAS to obtain the midvein curve of the leaves. GWAS was also employed to analyze the association between leaf midvein curves and genes.

Yang et al. (2011) constructed a large association panel in maize, and assembled a comprehensive maize association analysis population comprising 527 globally diverse lines representing tropical, subtropical, and temperate germplasm. The population is a collection of major inbred lines from around the world, which are believed to represent the principal morphological features of the maize genes and phenotypes. The use of the collection as a research object ensures comprehensive coverage of the diverse leaf shapes observed in maize leaves, thus facilitating subsequent genotype-phenotype correlation studies. To date, numerous genotype-phenotype related studies have been conducted on the maize association analysis population, with all of them yielding positive results (Gao et al., 2024; Song et al., 2024; Wang et al., 2024a, b; Wu et al., 2024; Xia et al., 2024; Yang et al., 2024).

At present, quantitative characterization methods for 2D leaf shape features of maize leaves are absent, and the interactions between leaf shape features remain unclear. Furthermore, there is a pressing need for big data analysis and knowledge mining for maize inbred populations within a phenomics perspective. The underlying laws

governing leaf shape features embedded in maize association analysis population materials remain elusive. Consequently, it is not yet feasible to achieve inbred line identification and visualization based on leaf shape features. In this study, we utilize the maize ear position leaves of the population proposed by Yang et al. (2011) as a case study to investigate the quantitative extraction of maize 2D leaf shape features, the construction of a leaf shape atlas, and the identification inbred line according to the leaf shapes. It is anticipated to facilitate the extraction of knowledge from the big data pertaining to leaf shapes.

2 Materials and methods

2.1 Experimental design and data acquisition

The experiment was conducted at the Nanfan breeding station of the Maize Research Center, Beijing Academy of Agriculture and Forestry Sciences (BAAFS), in Yazhou District, Sanya City, Hainan Province, China (longitude 109.1832, latitude 18.3623). The material selected for analysis in this research was the collection mentioned in section 1 (Yang et al., 2011). The specific planting times and planting settings can be found in Wu et al. (2022).

Data acquisition was carried out when the maize plants reached the silking stage, and the process was conducted from May 17 to May 27, 2021. The acquisition process was (1) Excavation of the root system and soil around the maize plants within a diameter of 25 cm and a depth of 20 cm, followed by arranging each individual plant in a pot (all sampling was done before 10 am) and transporting them indoors and watering them to minimize significant morphological changes caused by plant water deprivation (Wu et al., 2019). (2) 3D data of the ear leaves of each plant were collected by a digital probe using FastScan & FastRak 3D digitizers in combination with a Tx4 calibration transmitter. The 3D coordinates of selected leaf points were obtained manually (Figure 1A). The average acquisition time for each leaf was 5-8 minutes. If the plant had multiple ears, the leaf of the largest ear was selected as the ear leaf. Each data was checked visually to ascertain its accuracy. The data acquisition rule was to collect five points at even intervals using a digital probe starting from the leaf base perpendicular to the direction of the leaf midvein; then continue in the same manner upward along the direction of the leaf midvein to the tip of the leaf, with an unique point to indicate the tip of the leaf (Figure 1B). The number of points collected varied for different leaves due to different leaf lengths, but the number of points collected conformed to (5×n+1) as determined by the collection rules. The number of sampling points for most of the leaves ranged from 66 to 91 points, and the detailed sampling results are shown in Figure 1C.

A total of 518 distinct inbred lines were involved, with three samples collected for each inbred line. Finally, a total of 1,522 maize leaf model were obtained.

2.2 Overview

The general progression of the methodology is illustrated in Figure 2. The process begins with data acquisition and processing. Mesh subdivision and parameterization methods are employed to transform the original rough 3D mesh data acquired by the 3D digitizer into fine 2D mesh data. Subsequently, the average leaf shape models of each inbred line were constructed using the 2D leaf mesh model from the same inbred line. The morphological leaf features were quantified, then clustered and screened to determine the semantic features. A new phenotypic indicator L_{2D} that comprehensively reflect the 2D leaf shape was proposed and the 2D leaf shape atlas was constructed accordingly. For application, the

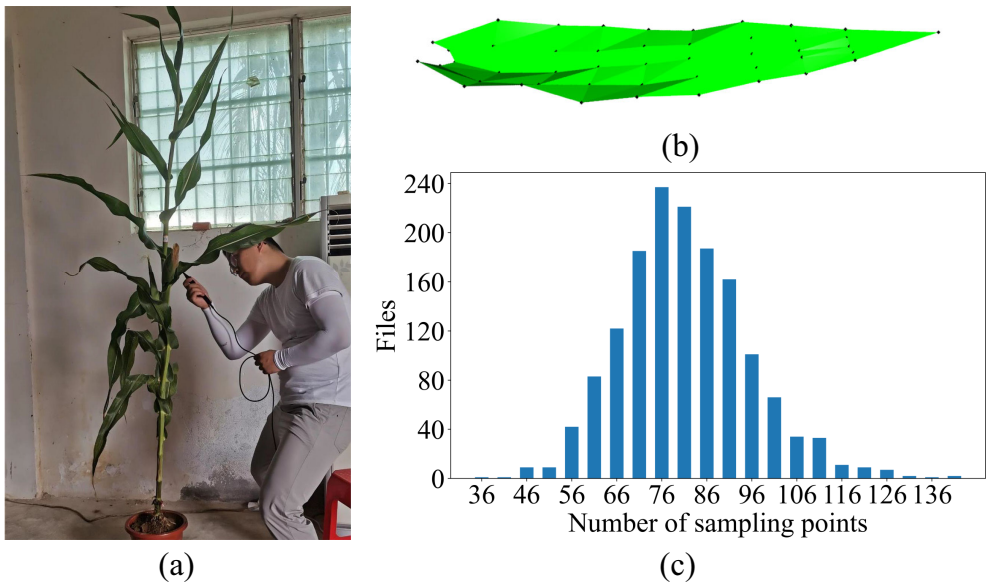


FIGURE 1
Raw data acquisition process and results. **(A)** Data acquisition, **(B)** the schematic of raw data acquisition, **(C)** statistics of the number of raw data points.

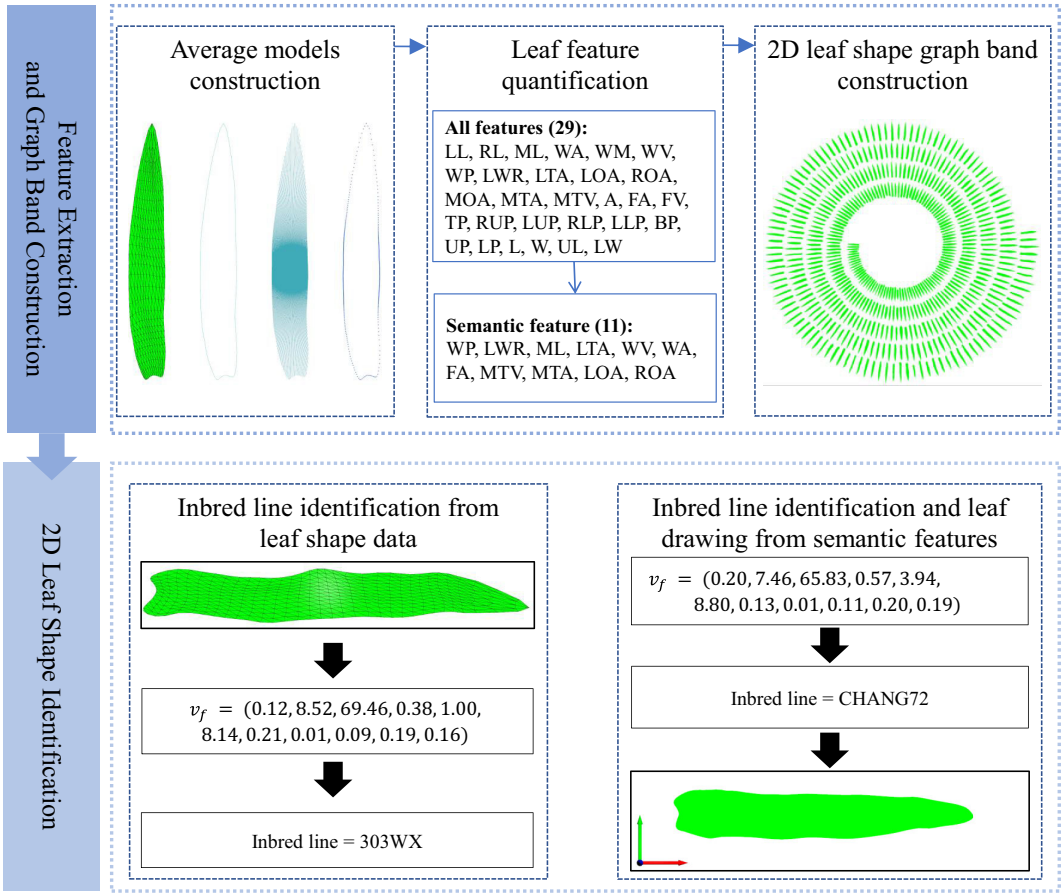


FIGURE 2
Overview of the overall process.

atlas was used to identify the inbred line according to a given leaf, or drawing the 2D leaf shape according to a given semantic feature.

2.3 Data processing and morphological feature quantification of 2D leaf shapes in maize

2.3.1 Maize leaf data preprocessing

The raw mesh data obtained by the 3D digitizer must undergo data preprocessing prior to its utilization in subsequent operations. The data preprocessing involves three steps: comparative examinations of the data within the same inbred line, data normalization, and raw mesh data subdivision.

2.3.1.1 Comparative examinations of the data within the same inbred line

The objective is to ascertain the distinctions between the three samples for each of the 518 inbred lines, with the exclusion of samples exhibiting significant discrepancies. The four phenotypic features of leaf length, leaf width, aspect ratio, and raw mesh average angle were initially estimated based on the raw leaf mesh. The similarity was calculated based on the aforementioned four features, employing the Euclidean distance metric. The resulting similarity of the samples within the same inbred line were then compared. If there is sample with a similarity difference exceeding 0.1 with other samples within the same group, while the similarity difference between the remaining samples is less than 0.1, the sample exhibiting the greatest divergence is excluded. Duplicate groups with only two samples and one remaining sample were not subjected to screening. A total of 91 samples with excessive differences were excluded in this step, leaving 1,431 maize leaf data from 518 inbred lines for subsequent analysis.

2.3.1.2 Data normalization

The principal axis direction of the maize leaf mesh model was initially obtained through principal component analysis (PCA). Subsequently, all principal axis directions were rotated so that the direction of the leaf tip pointed to the positive x-axis, with the center of mass designated as the origin [0, 0, 0]. Secondly, all leaves were maintained in a consistent proportion relative to one another, and the coordinate values of each point were uniformly scaled to the range of [-1, 1].

2.3.1.3 Mesh subdivision

The normalized mesh is then subdivided using the $\sqrt{3}$ (sqrt3) subdivision method (Kobbelt, 2000). The sqrt3 subdivision method is an efficient triangular mesh subdivision algorithm that belongs to

the facet-splitting surface approximation mode. It has the advantages of a slower increase in mesh complexity during the refinement process and a certain degree of adaptive refinement. Following the subdivision of the raw leaf mesh, the resulting leaf model exhibits a shape that is more closely aligned with the actual leaf morphology (Figure 3B) compared to the raw data (Figure 3A). This enhanced resemblance to a more accurate reflection of the observed features. In this study, two iterations of the sqrt3 subdivision were performed on the original mesh. Following the operations, the number of triangular meshes was found to be nine times that of the original data, resulting in a more uniform and smooth mesh model.

2.3.2 2D Flattening of 3D leaf mesh models

The process of flattening a 3D mesh model in 2D is referred to as planar parameterization. In this study, the as-rigid-as-possible (ARAP) (Liu et al., 2008; Sorkine-Hornung and Alexa, 2007) method is employed.

In the planar parameterization process, an energy function is used to delineate the discrepancy between the potential transformations and the target transformation.

$$E(u,\ L) = \sum_{t=1}^{T} A_t \,||\, J_t(u) - L_t \,||_F^2,$$

In the function, the area of the 3D triangles are $A_t\ \ (1 \le t \le T)$, For triangle t, $u_t = \left\{u_t^0\ ,u_t^1\ ,u_t^2\right\}$ denotes the 2D coordinates, and $x_t = \left\{x_t^0\ ,x_t^1\ ,x_t^2\right\}$ denotes the 3D coordinates. The relationship between x_t and u_t can be represented by a 2×2 Jobabian matrix $J_t(u)$. L_t is assigned as an auxiliary linear transformation (2×2 matrix). $||\cdot||_F$ is the Frobenius norm. The variables in the energy function are the coordinates u of the mapping on the 2D plane and the transformation matrix L_t. To minimize the energy function and restrict L_t to the rotation matrix in the ARAP algorithm, the problem can be transformed into the following optimization problem after reconstruction and derivation:

$$(u,\ t) = argmin(u,\ t)E(u,\ t),\ \ L_t \in M$$

$$M = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},\ \ \theta \in [0, 2\pi)$$

This method maintains the angle of the triangles with the greatest possible constancy during the 3D to 2D mapping process, ensuring that the triangles are not distorted. Furthermore, the areas of the triangles are only slightly affected, thus preserving the invariance of the area to the greatest extent possible. Subsequently, the 3D mesh model is transformed into a corresponding 2D planar mesh model (Figure 3C). As illustrated in Figure 3, the planar parameterization of the 3D mesh model results in a deformed 2D mesh model due to the wrinkles and distortions inherent to the original 3D model. Consequently, the flattened 2D mesh model is not entirely symmetric along the axis of the leaf midvein. The 2D leaf model resulting from planar parameterization retains the basic phenotypic features of maize leaves, such as leaf length and width, to the greatest extent possible. Additionally, the deformations introduced by planar parameterization can also reflect certain 3D maize leaf phenotypic features, such as leaf twist. Accordingly, the 2D maize leaf mesh model obtained through the ARAP method is a more suitable means of extracting relevant phenotypic features for the analysis and research of leaf morphology.

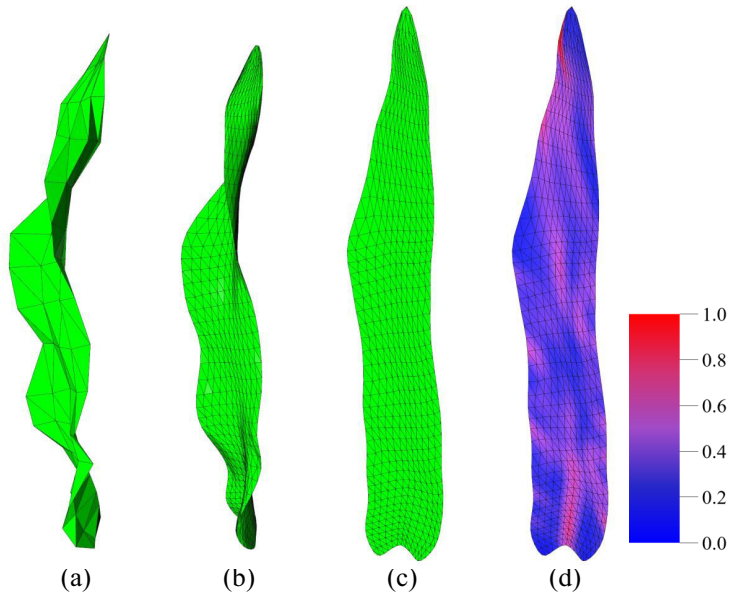


FIGURE 3
Illustration of key stages in data processing. **(A)** Raw 3D leaf data. **(B)** 3D leaf data after mesh subdivision (Kobbelt, 2000). **(C)** Parameterized 2D leaf mesh (Liu et al., 2008; Sorkine-Hornung and Alexa, 2007). **(D)** Fold visualization of a 2D leaf (more blue color means less folds, more red color means more folds).

2.3.3 Construction of averaged 2D leaf models

Average leaf shape models were constructed for each of the distinct inbred lines of the 2D leaf mesh models. In total, 518 models were constructed. The leaf shape model is primarily concerned with the representation of leaf contour information, which is ultimately conveyed through a set of contour lines. The construction of the average model involves three steps: leaf contour extraction, contour sampling, and line set model construction.

2.3.3.1 Leaf contour extraction

2D leaf mesh model is a non-closed mesh (Figure 4A), and the contour edges of the leaf mesh are the boundary edges. The edge contour is obtained by collecting all the boundary edges in the leaf mesh model. This process yields a line set comprising 1,431 leaves, as illustrated in Figure 4B.

2.3.3.2 Contour sampling

A ray is initiated from the center point ([0, 0]) of the leaf and rotated around the center, intersecting with the leaf contour line set to obtain the sampling point set (Figure 4C). To guarantee the precision of the sampling, the sampling resolution was set to 600, thus ensuring the attainment of optimal sampling outcomes at the leaf tip and the base, which are situated at considerable distances from the center point (Figure 4D).

2.3.3.3 Averaged 2D leaf model construction

The average point set of specific inbred line and the average point set of the entire leaves can be calculated from the sample results. The mean value of the coordinates of the sampling points at each position belonging to the same inbred line of samples is estimated. The overall leaf average point set is calculated in the same way. The points in the point set are connected sequentially in order to finally obtain the 518 inbred lines average model.

2.3.4 Extraction of 2D leaf shape features from 2D mesh models

A total of 29 2D phenotypic leaf features were extracted from 1,431 2D mesh models of maize leaves, with reference to previous research (Wen et al., 2024a; Wu et al., 2022; Zheng et al., 2022). These features included 17 conventional phenotypic features and 12 leaf contour features. The conventional leaf parameters include: leaf length, leaf width, leaf tip angle, leaf area, etc. Additionally, the length of the left and right edges of the leaf, the offset degree of the left and right edges, and the width of the leaf at different positions were also extracted according to the 2D maize leaf model. The specific phenotypic features are presented in Table 1.

The remaining 12 features pertain to the leaf contour, as illustrated in Table 2. By dividing the leaf contour into six parts (Figure 5), namely, the tip part, the upper (left/right) part, the lower (left/right) part, and the base part, and calculating the distance between each part and the center point, a quantitative value of the leaf contour reflecting the shape of each part was obtained. The quantitative values of the various parts were used to obtain two additional features: the "Length/Width" and the "Upper - Lower". This resulted in a total of 12 features.

2.3.4.1 Calculation of 1D features (length, width and related features)

In the case of a 1D features, the calculation is based on the distance and angles between the points in the 2D mesh model. These 1D features include mid length (ML), maximum width (WM), average width (WA), variance, and widest position (WP), etc.

2.3.4.2 Calculation of 2D features (area, leaf tip angle, offset and tortuosity)

Leaf area was calculated by summing the areas of all triangular mesh facets. Compared to traditional methods based on the formula (leaf length × maximum leaf width × 0.75) (Valentinuz and Tollenaar, 2006) or (leaf length × maximum leaf width × 0.765) (Lizaso et al., 2003), the approach of calculating the area of the mesh facets individually yields a more precise estimation of the leaf area.

Leaf tip angle, offset and tortuosity are angle features. The left and right edge angles at the leaf tip represents the leaf tip angle. The offset values of the left and right edges and midvein of the leaf are calculated by summing the angles obtained from the point-by-point line calculation of the mesh vertices at the corresponding positions. This emphasizes the overall degree of curvature of the leaf edge contour. The tortuosity of the midvein is calculated using the mean and deviation between the midvein and the line between the base to the tip of the leaf. It emphasizes the curvature of the leaf after the planar parametrization.

2.3.4.3 Calculation of the folding degree

Folding is an important morphological feature of maize leaves, and a quantitative calculation method for folding based on data obtained by 3D digitizer was proposed. For the folding extracted from each sample, the average values of phenotypic features representing the value of each inbred line were obtained by intra-group averaging based on 518 different inbred lines. The FA mainly reflect the level of leaf folds comprehensively. The FA parameters were calculated as follows:

$$FA = \frac{\sum_{i=1}^{m} \frac{\sum_{j=1}^{n_i} \alpha(\vec{N}_i, \vec{N}_{ij})}{n_i}}{m},$$

$$\alpha(\vec{x}, \vec{y}) = arccos\left(\left| \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| \cdot |\vec{y}|} \right| \right)$$

where m represents the total number of mesh patches in the maize leaf mesh model, n_i represents the number of patches adjacent to patch i, $\vec{N}_i$ represents the normal vector of the patch i, $\vec{N}_{ij}$ represents the normal vector of the j_{th} patch neighboring patch i.

2.4 Maize leaf semantic feature extraction and 2D leaf shape atlas construction

2.4.1 2D leaf shape semantic feature determination

While the aforementioned 29 features provide comprehensive coverage of leaf shape, some features exhibit similarities to a certain extent. This correlation between features may result in redundancy,

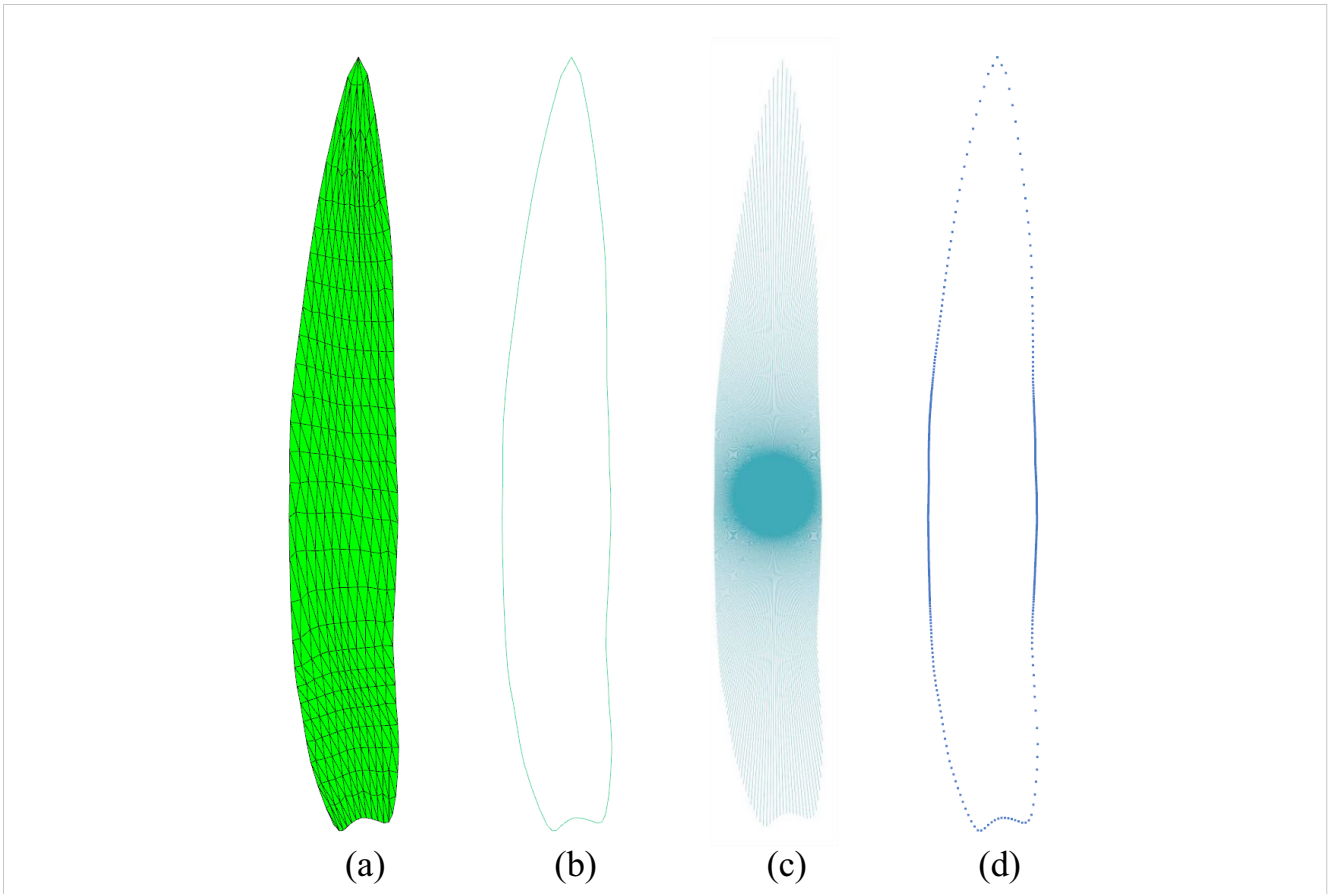


FIGURE 4
Leaf shape sampling. **(A)** Subdivided 2D maize leaf model. **(B)** Mesh contour. **(C)** Mesh contour sampling by emitting rays from the leaf center. **(D)** Sampled points of the leaf contour.

which could impede the simplicity of the feature representation. To address this, we have selected a subset of features as the semantic features of the 2D leaf shape, aiming to maximize feature coverage while simplifying the representation.

The correlation and clustering analyses were conducted using the 29 2D leaf shape features obtained above to determine the key features. The hierarchical clustering approach employs a specified method to quantify the degree of affinity between the features. The process involves initially clustering the more closely related features into one class, and then repeating this until all the features are clustered into one class. The selection of a suitable distance metric is paramount for accurately measuring the difference between phenotypic features. In this study, Pearson Correlation Coefficient is employed as the distance metric, which reflects the degree of linear correlation between the two features under investigation. The Pearson Correlation Coefficient is calculated as follows:

$$1 - \frac{(u - \bar{u}) \cdot (v - \bar{v})}{||u - \bar{u}||_2 ||v - \bar{v}||_2}$$

where $\bar{v}$ is the mean of the elements of vector v, and $x \cdot y$ is the dot product of x and y. A high correlation between two phenotypic features indicates a significant degree of overlap and similarity in their ability to reflect leaf phenotype. Consequently, the more representative features were selected for characterization, with the objective of simplifying and refining the feature information. Furthermore, Average Linkage is chosen as the linkage algorithm, as it has been demonstrated to effectively calculate the average distance between all pairs of points in two clusters, thereby facilitating the generation of compact clusters. The Average Linkage (average) algorithm assigns:

$$d(u, v) = \sum_{ij} \frac{d(u[i], v[j])}{(|u| * |v|)}$$

for all points i and j where $|u|$ and $|v|$ are the cardinalities of clusters u and v, respectively.

2.4.2 2D leaf shape atlas construction

Once the semantic features of the 2D leaf shape have been determined, a n-dimensional feature vector v_f can be constructed by weighting and combining all the semantic features to reflect the main characteristics of each leaf. The feature vectors were used to sequencing the 518 inbred lines of leaves according to the differences in the features. In this sequencing process, phenotypic features should be comprehensively considered. The sorted result was used to construct a atlas for 2D leaf shape, which reflected the differences in the leaf shapes of the maize in comparison with the evolutionary process.

TABLE 1 Phenotypic features extracted from 2D maize leaf model.

ID	Feature name	Identifier	Explanation of the feature	Unit
1	Left length	LL	Length from the left point of the base part along the left margin to the point of the leaf tip	*cm*
2	Right length	RL	Length from the right point of the base part along the right margin to the point of the leaf tip	*cm*
3	Mid length (Leaf length)	ML	Length from the midpoint of the base part along the midvein to the point of the leaf tip (Leaf length)	*cm*
4	Width (average)	WA	The average width of the leaf obtained perpendicular to the direction of the midvein from the leaf base to the leaf tip	*cm*
5	Width (max)	WM	The max width of the leaf obtained perpendicular to the direction of the midvein from the leaf base to the leaf tip	*cm*
6	Width variance	WV	The variance of leaf width obtained perpendicular to the direction of the midvein from the leaf base to the leaf tip	cm^2
7	Widest position	WP	Distance from the position of maximum leaf width to the leaf base/leaf length (WP=0 if the position of maximum leaf width is at the leaf base; WP=1 if the position of maximum leaf width is at the leaf tip)	–
8	Length-width ratio	LWR	ML/WA	–
9	Leaf tip angle	LTA	The angle value between the leaf tip point and the neighboring points on both sides	*rad*
10	Left offset angle	LOA	Sum of the angle values between the lines between the points on the left edge of the leaf	*rad*
11	Right offset angle	ROA	Sum of the angle values between the lines between the points on the right edge of the leaf	*rad*
12	Mid offset angle	MOA	Sum of the angle values between the lines between the points on the midvein of the leaf	*rad*
13	Mid tortuosity (average)	MTA	Mean value of the angle between the line between the points on the midvein of the leaf and the line between the leaf base and the leaf tip.	*rad*
14	Mid tortuosity variance	MTV	The variance of the angle between the line between the points on the midvein of the leaf and the line between the leaf base and the leaf tip.	rad^2
15	Area	A	Total leaf mesh area	cm^2
16	Folding (average)	FA	Mean value of all triangular mesh angles of the leaf	*rad*
17	Folding variance	FV	The variance of all triangular mesh angles of the leaf	rad^2

$$v_f = (f_1, f_2, \cdots, f_n)$$

$$\boldsymbol{w} = (w_1, w_2, \cdots, w_n)$$

where f_i is the i_{th} semantic feature and w_i is the weight coefficient of the i_{th} feature.

In the determination of w, it should be noted that the number of inbred lines included in this research is 518, with the data for each inbred line consisting of only one to three samples. Given the limited number of samples within each group, it is challenging to achieve a superior training outcome when the sample size is insufficient utilizing machine learning techniques. Consequently, when determining the optimal weights, we employ the greedy algorithm to enumerate and calculate the most suitable weights for each semantic feature, and then as the weighting coefficients for the final leaf feature vector.

2.5 Inbred line identification based on 2D leaf shape atlas

Accurately identify a given inbred line based on the observed leaf shape data is a crucial aspect of crop breeding, representing one of the practical applications of the proposed method. The process of identifying a inbred line based on 2D leaf shape atlas can be divided into two main aspects: identifying the inbred line from the given leaf, and identifying the 2D leaf shape from the given semantic features.

2.5.1 Inbred line identification from given leaf shape data

The cosine similarity is a difference measure between two vectors in an n-dimensional space. A feature vector was constructed for each 2D mesh model of a maize leaf based on the semantic features of the leaf. Given an example leaf shape data, the cosine similarity was calculated by comparing the vector of this example similar leaves in the atlas. The inbred line with the highest similarity was selected and identified as the inbred line of the given leaf.

2.5.2 Inbred line identification and leaf shape drawing from given semantic feature

As with the aforementioned method for identifying a given leaf shape, the closest maize inbred line to a given feature vector is determined by calculating the cosine similarity. This allows the leaf shape of that inbred line to be identified in the constructed leaf shape atlas, which in turn permits the unique drawing (visualization) of the 2D leaf shape for a given semantic feature.

TABLE 2 2D leaf contour features.

ID	Feature name	Identifier	Explanation of the feature	
1	Tip part distance	TP	Mean distance of sampling points of the tip part from leaf center point	*cm*
2	Right upper part distance	RUP	Mean distance of sampling points of the upper right part from leaf center point	*cm*
3	Left upper part distance	LUP	Mean distance of sampling points of the upper left part from leaf center point	*cm*
5	Left lower part distance	LLP	Mean distance of sampling points of the lower left part from leaf center point	*cm*
6	Base part distance	BP	Mean distance of sampling points of the base part from leaf center point	*cm*
7	Upper part distance	UP	Mean distance of sampling points of the upper left part and upper right part from leaf center point	*cm*
8	Lower part distance	LP	Mean distance of sampling points of the lower left part and lower right part from leaf center point	*cm*
9	Length	L	Overall leaf length calculated from TP and BP correspondence	*cm*
10	Width	W	Overall leaf width calculated from RUP, LUP, RLP, LLP correspondingly	*cm*
11	Upper - Lower	UL	Difference between the widths of the upper and lower parts of the leaf calculated from the correspondence between UP and LP	*cm*
12	Length/Width	LW	Length width ratio of leaf calculated from L and W	–

2.5.3 Evaluation metric

The Top-X Accuracy metric r_{Top-X}, is employed to evaluate the performance of inbred line identification from either a specific 2D leaf shape or a semantic feature vector. The value of r_{Top-X} indicates that the ground truth inbred line labels of the leaf instance data are within the top X of the predicted similarity results. In this context, r_{Top-1} indicates that the inbred line with the highest predicted similarity is the ground truth. r_{Top-3} indicates that the ground truth is among the top three results for similarity.

3 Results

3.1 Results and analysis of 2D flattening from 3D leaves

The ARAP planar parameterization method is designed to minimize changes in mesh area and shape distortion during 2D flattening. A comparison of the area of 3D and 2D mesh models before and after planar parameterization indicates that the area of the mesh models after planar parameterization is generally reduced by a very small amount. The average leaf area ratio of the 2D mesh models to the 3D mesh models for the 1,431 leaf mesh models is 99.88%, and there is only a change of 0.12%. As illustrated in Figure 6, the area ratio of the 2D and 3D mesh models is consistently above 0.990 and below 1.000. The majority of the data points are situated within the interval between 0.998 and 1.000, with only a few instances where the ratio is below 0.998.

As illustrated in Figure 7, the ARAP planar parameterization technique is capable of accurately retaining the morphology of the 3D leaf mesh, as evidenced by the zigzagging of the 2D leaf mesh edges, while the projection method is susceptible to influences such as the projection angle. The ARAP planar parameterization technique also accurately restores the morphometrics of the entire leaf, including the tip and the base. In contrast, the 2D mesh based on the projection method is unable to accurately reproduce the leaf tip and the base due to the excessive curvature of the leaf. Additionally, the side edges of the 2D mesh lack smoothness due to the mutation of edge contour caused by the projection, which hinders the accurate reflection of the true shape of the original 3D leaf. It is challenging to accurately represent the true shape of the original leaf. The extracted leaf contour, obtained by manually flattening the leaf and photographing it, is shown in Figure 7F. Due to the inability to eliminate wavy folds of the leaves when flattening the leaves manually, excessive wavy zigzags can be observed in the extracted contour. In comparison, the contour obtained by the ARAP method exhibits a smoother contour and circumvents substantial errors in estimating phenotypic features such as leaf area.

Furthermore, the 2D mesh model, following ARAP planar parameterization, can effectively visualize and demonstrate the folds of the 3D mesh model (Figures 3B, 7H). The visualization of 3D leaf folds on a 2D leaf after planar parameterization reveals that the red regions correspond to larger folds. Firstly, it can be observed in Figure 7H that the majority of leaves exhibit a large percent red area in the base midline portion, this distinct feature is also evident in Figure 7I. This phenomenon can be attributed to the bending morphology exhibited by ear leaves, which is evident from the base of the leaf midvein to the leaf sheath connection. Given that ear leaves are longer and wider, this bending structure is more pronounced, providing greater support for the leaves. It should be noted that, with the exception of the base of the leaf, the folds can be observed in the position of the midvein. This is due to the fact that all the leaves exhibit a bent morphology at the position of the leaf midveins. Secondly, as illustrated in Figure 7I, the folds of the maize leaves are predominantly concentrated in the lower middle portion of the leaf, where the overall distribution of the folds is characterized by a dense area with a curved boundary. This boundary is a consequence of the widespread bending of the left and right edges of the middle and lower parts of the leaf towards the back of the leaves, as well as the pronounced wavy folds of the leaf perpendicular to the direction of the leaf veins in this region.

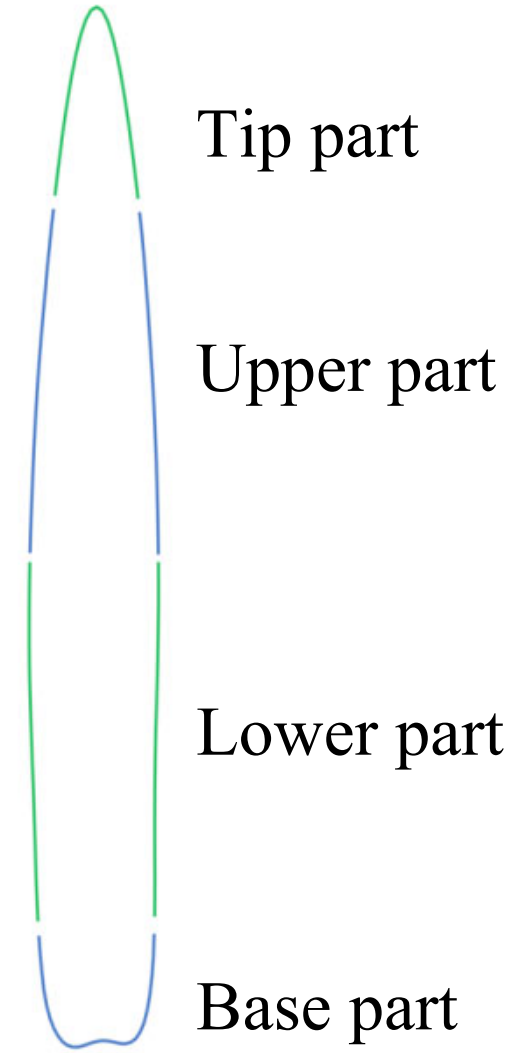


FIGURE 5
Leaf Contour Segmentation.

Thirdly, a specific distribution of folds is observed at the tip of the leaf. This distribution is not so dense than that observed at the lower middle portion of the leaf, primarily due to the overall amplitude of the folds at the tip being smaller than that observed at the lower middle portion of the leaf. The formation of folds at the tip of the leaf can be attributed to the fact that the tip is more susceptible to rotation and deflection than other regions. At last, the upper-middle region exhibits the lowest degree of folding when compared to other regions. Except the bend of the midveins in the middle region displays a notable degree of folding, the surface of this portion is more smoothed than that of the remainder.

3.2 Semantic feature of 2D leaf shape

11 out of 29 2D leaf-shape features are identified as semantic features (WP, LWR, L, LTA, WV, W, FA, MTV, MTA, LOA, ROA) by setting the clustering threshold to 0.25 and using the "average" method and the "correlation" metric in hierarchical clustering, as shown in Figure 8. The determination of an optimal clustering threshold is a critical step. Through experimental analysis, a threshold value of 0.25 has been determined to achieve optimal distinction between phenotypic features based on their correlation. That is, most of the 2D leaf-shape characteristics can be characterized using these 11 semantic features.

Figure 9 depicts the fitted distribution functions of the 11 semantic features. To determine the weights of each semantic feature, seven values were taken from the interval of [0, 3] with step length of 0.5 for each semantic feature to validate the optimal weight. We sequentially traversed to test the identification ability corresponding to seven different weights with the weights of the other semantic features had been set to 1, then we selected the weight with the best identification ability to be determined as the corresponding weight of the semantic feature. The 11 semantic features were systematically traversed to ascertain the optimal weights, and a set of optimal weights is determined. Following the validation and normalization of the

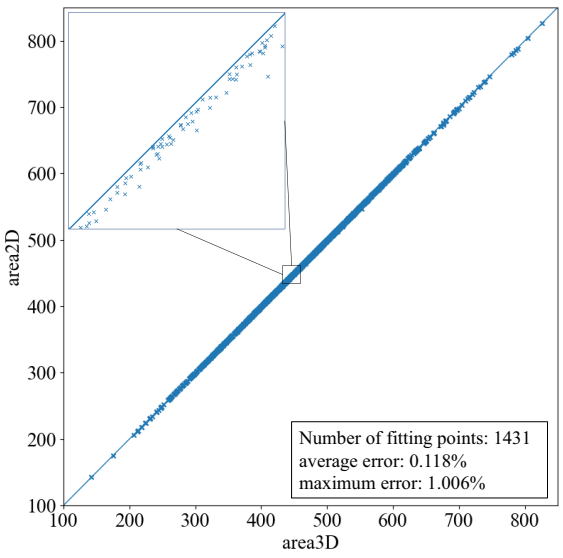


FIGURE 6
Leaf area ratio of 2D model to 3D model after planar parameterization.

features, the correspondence between the semantic features and the corresponding optimal weights is presented in Table 3. Notably, L, which reflects the length of the leaf, plays a pivotal role in the identification process. Additionally, the two features, MTA and MTV, measure the degree of tortuosity of the leaf and the variation degree of tortuosity, both of which are crucial for identify between different leaf shapes. Additionally, the W reflects the width of the leaf. Consequently, the aforementioned four parameters are assigned the highest weights.

3.3 Results of 2D leaf shape atlas construction

In accordance with the methodology in section 2.4.2, the semantic feature vectors of 2D leaf shape can be determined by the 11 semantic features with the corresponding weights. The complete semantic feature vectors v_f and the corresponding weight vectors w are as follows:

$$v_f = (WP,\ LWR,\ L,\ LTA,\ WV,\ W,\ FA,\ MTV,\ MTA,\ LOA,\ ROA)$$

$$w = (0.028,\ 0.056,\ 0.167,\ 0.111,\ 0.028,\ 0.111,\ 0.083,\ 0.139,\ 0.167,\ 0.056,\ 0.056)$$

To establish a rule for ranking the 2D leaf shapes, a weighted vector v_{f_w} is defined using the feature vector v_f with the weight vector w. $L_{2D} = ||v_{f_w}||$, the modulus of the weighted vector v_{f_w}, is then calculated as the rule for ranking the 2D leaf shape of all the inbred lines, thus generating a atlas. L_{2D} serves as a comprehensive indicator reflecting the overall 2D morphology features of a maize leaf. L_{2D} increases along with the semantic characteristics of the leaf become more pronounced (e.g., longer length, wider width, more curved edges, etc.). Conversely, L_{2D} also decreases as the phenotypic characteristics of the leaf become less pronounced.

$$L_{2D} = ||v_{f_w}|| = \sqrt{\sum_{i=1}^{n}(w_i \cdot v_{f_i})}$$

$$\sum_{i=1}^{n} w_i = 1$$

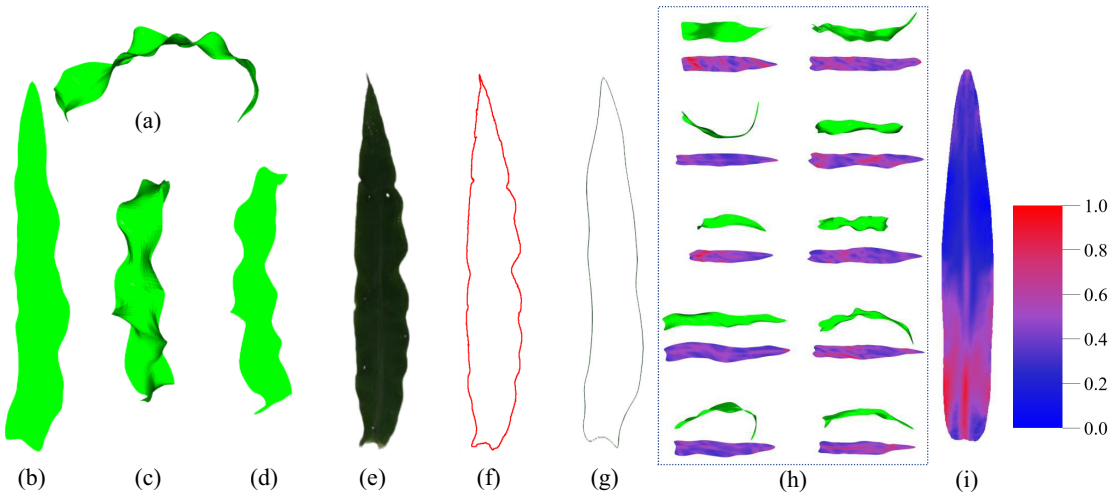


FIGURE 7
Comparison of the results of the planar parameterization. **(A)** 3D mesh model of the maize leaf. **(B)** 2D mesh model after ARAP planar parameterization. **(C)** Front view of the 3D mesh model of the maize leaf. **(D)** Front view projection of the 3D mesh model. **(E)** Top view image of a manually flattened maize leaf. **(F)** Extracted leaf contour of **(E)**. **(G)** Extracted leaf contour after ARAP flattened. **(H)** The result of planar parameterization for different leaves. **(I)** Fold visualization of the average model for all leaves (more blue means less folds, more red means more folds).

where w_i is the i_{th} parameter in the weight vector w, and v_{f_i} is the i_{th} parameter in the feature vector v_f.

The 518 inbred lines were ranked according to the L_{2D} to form the final 2D leaf shape atlas (Figure 10). As illustrated in Figure 10, the maize leaves in the inner circle are relatively diminutive in both overall length and width, and the leaf contour is relatively smooth, exhibiting no discernible zigzagging at the edges. In contrast, the leaves in the outer circle are larger in both overall length and width, displaying pronounced zigzagging at the edges, and the phenotypic characteristics of the leaves are more pronounced.

3.4 2D leaf shape identification results

3.4.1 Inbred line identification results for given leaf shape

A total of 1,431 leaves were utilized to conduct inbred line identification, and using r_{Top-X} as the performance evaluator. For comparison, experiments were conducted in three setups: (1) using the full 29 features, (2) using unweighted 11 semantic features, and (3) using 11 semantic features with trained weights according to the random forest. The results of the experiment are presented in Figure 11 and Table 4. The weighted features, calculated using the greedy algorithm, demonstrated the optimal performance. Furthermore, the proposed method outperforms the other three feature vector design methods in r_{Top-1}, and the accuracy remains superior as X increases. The r_{Top-X} of this method exhibited accelerated improvement with increasing of X, attaining r_{Top-10} of 0.706, r_{Top-20} of 0.810, and r_{Top-45} of 0.900. Due to the considerable number of inbred lines included in this study (a total of 518), r_{Top-10} has been possible to identify the inbred lines within 2% of the total number of inbred lines. Furthermore, r_{Top-20} and r_{Top-45} have been achieved, whereby the inbred lines have been identified within 4% and 9% of the total number of inbred lines, respectively. Given the limited number of samples obtained from the various inbred lines and the observed variability in the characteristics of the leaves, the outcomes of this study are noteworthy.

3.4.2 Identification results for given semantic features

For semantic features of a 2D leaf, the model with the highest weighted cosine similarity among the 518 inbred line models can be identified as the one that most closely aligns with and represents the corresponding 2D model of the leaf (Figure 12). To guarantee that the leaf illustration accurately reflects the semantic features, the matching threshold is set to 0.9. This implies that when the highest match between a specific semantic feature and the inbred line model library is less than 0.9, it is assumed that this particular structure of a maize leaf does not

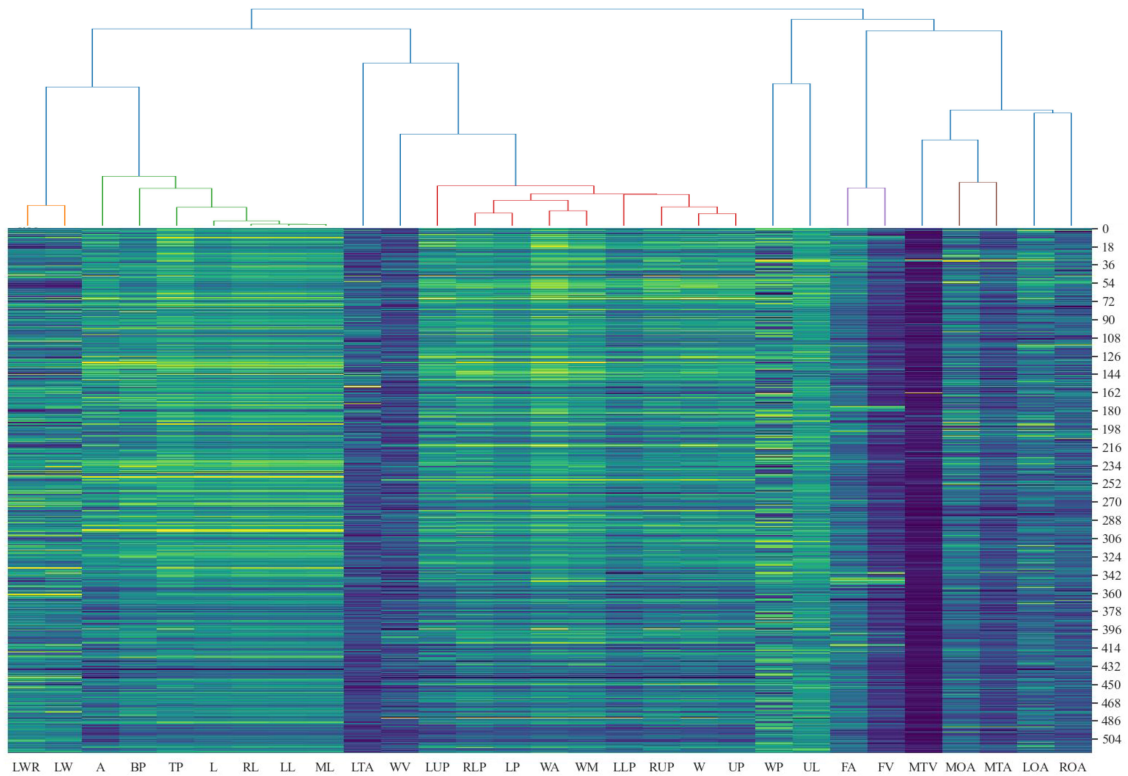


FIGURE 8
Cluster analysis of 2D leaf shape features.

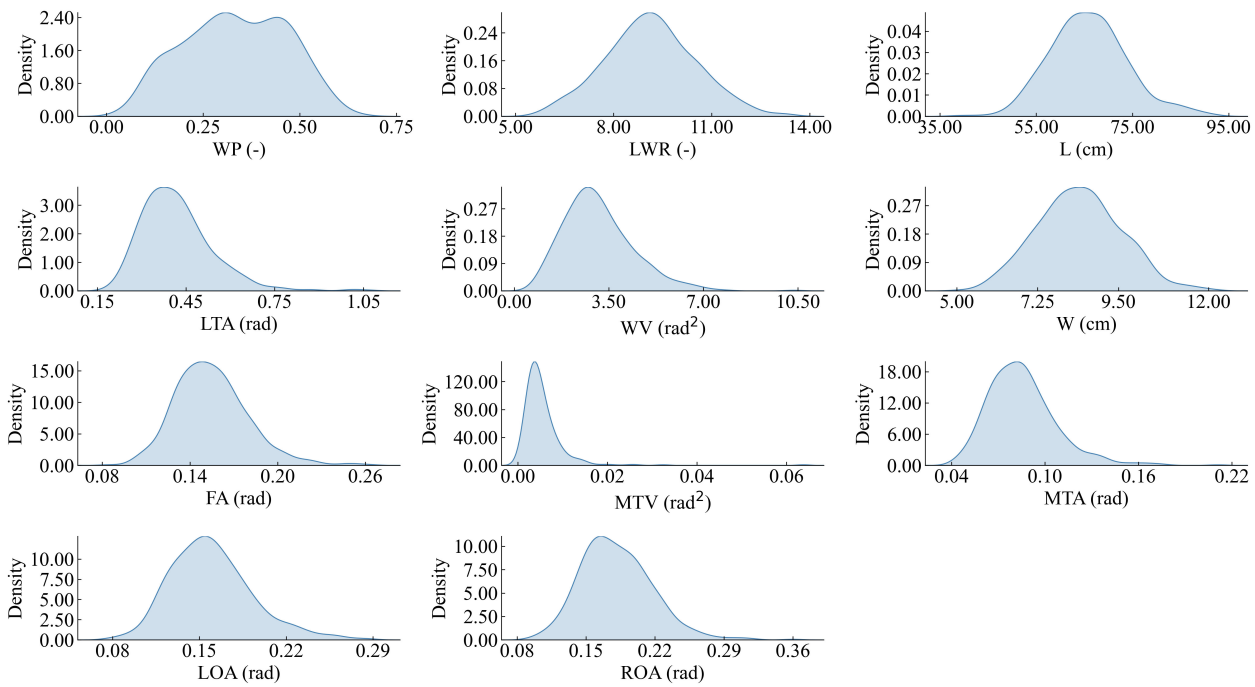


FIGURE 9
Distribution function of the 11 semantic features.

exist, and thus, the specified maize leaf model cannot be generated. This is a straightforward method for generating 2D leaf models based on feature matching. The data template library in this study provides comprehensive coverage of maize 2D leaf models, and the feature vectors constructed in this study encompass the majority of features observed in maize 2D leaf models. Furthermore, the 518 inbred line models are all generated based on true leaves, ensuring that the generated results are rich, authentic, and of significant research value.

TABLE 3 Semantic features with corresponding weights.

ID	Identifier	Physiological significance	Weight
1	WP	Widest position of leaf	0.028
2	LWR	Leaf aspect ratio	0.056
3	L	Average leaf length based on leaf contour	0.167
4	LTA	Leaf tip angle	0.111
5	WV	Variance of the width of the leaf	0.028
6	W	Average leaf width based on leaf contour	0.111
7	FA	Fold	0.083
8	MTV	Variance of leaf midvein tortuosity	0.139
9	MTA	Mean value of leaf midvein tortuosity	0.167
10	LOA	Degree of curvature of the left edge of the leaf	0.056
11	ROA	Degree of curvature of the right edge of the leaf	0.056

4 Discussion

4.1 Leaf shape atlas promotes plant phenotypic identification

This study benefits from "human face recognition" technology to identify and categorize 2D maize leaves. The key point is that to quantify the morphological features of 2D leaf, and ascertain the typical characteristics as semantic features, which then allows for focused exploration and classification. In comparison to existing research on crop leaf shapes and phenotypic identification, this study presents the following novel contributions and advances:

(1) Construction method of leaf 2D mesh model. This study applied the ARAP planar parameterization in converting 3D mesh model to 2D planar mesh. The method allows for the more precise retention of the 3D structure, surpassing the traditional methods such as projection or manual flattening (Wang et al., 2022). Furthermore, the method avoids the potential loss of information due to projection angle (Figures 7B, D). For example, the leaf area is highly accurately reserved in the 2D mesh model obtained by ARAP planar parameterization method.

(2) Quantification and determination of semantic leaf features. A multitude of phenotypes for expressing leaves have been put forth in previous studies. However, the divergence of research directions have resulted in numerous redundancies of phenotypic features. For example, the terms "upper leaf midrib sag" and "leaf midrib curvature" in the work of Wu et al. (2022) were used to describe the curvature of the

10.3389/fpls.2025.1520297

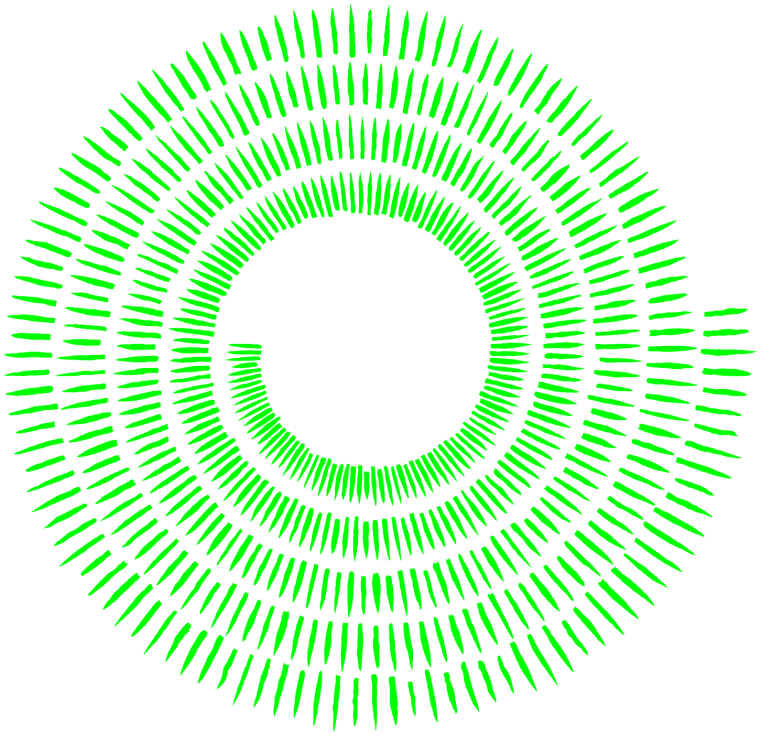


FIGURE 10
2D leaf shape atlas ranked using L_{2D} (L_{2D} increasing from inside to outside).

leaf midrib from different perspectives. Zheng et al. (2022) put forth the terms "leaf sag" and "leaf bending" to describe the sagging curvature and overall bending curvature of the leaf, respectively. The two aforementioned metrics exhibit a high degree of similarity. Consequently, the ability to effectively screen the features, ensuring both concise representation and comprehensive coverage, becomes a pivotal determinant of the success of subsequent research. In this study, a hierarchical clustering method was employed to screen and categorize the phenotypic features proposed in previous studies, resulting in the determination of 11 parameters of 2D maize leaf features that were not duplicated by each other and could be more comprehensively covered. This approach enabled the condensation and compression of the 2D leaf features, laying the foundation for the construction of feature vectors and the identification of leaf cultivars or inbred lines.

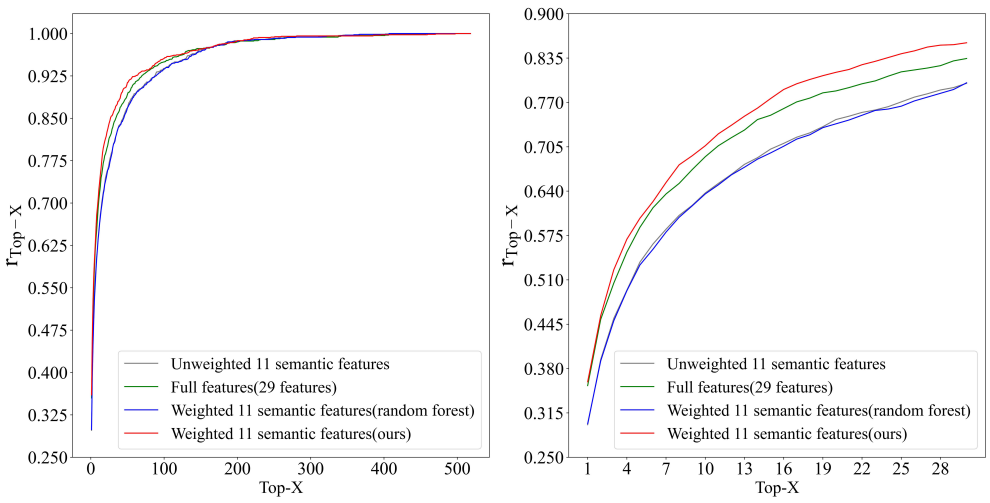


FIGURE 11
Top-X Accuracy trend in inbred line identification using leaf shape features.

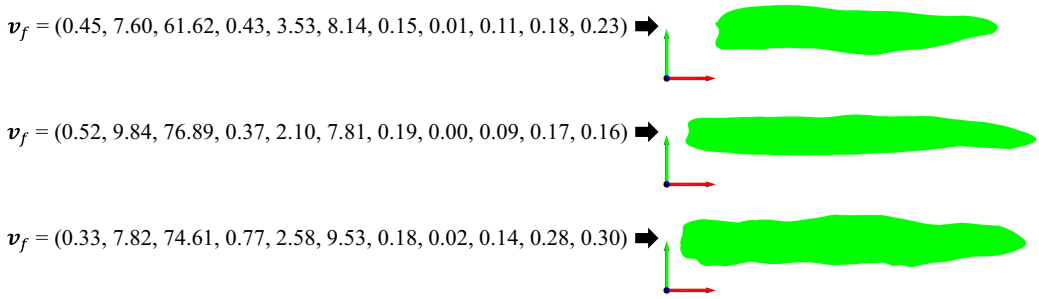


FIGURE 12
Three examples of generating and drawing 2D leaf models from feature vectors.

(3) 2D leaf shape atlas construction. Given the considerable number of phenotypic features employed to describe leaf shapes, a comprehensive ranking of all leaf shapes is essential to fully reflect the observed variation between leaf shapes. In previous studies (Åström et al., 2015; Cohen, 2011; Dubey et al., 2024), the screening of features are remain qualitative, and comprehensive quantitative ranking integration is lacking. In this study, a comprehensive 2D leaf shape indicator L_{2D} was obtained by weighting the 11 semantic features and estimating the module length of the vector. The 2D leaf shape atlas was obtained by ranking the 2D leaves according to L_{2D}. This approach allows for sequencing leaf shapes based on the main leaf features. Additionally, it demonstrates the relationship between the expression of the 2D leaf shape features, from weak to strong, in a population comprising a vast array of inbred lines.

(4) Inbred line identification and 2D leaf shape generation. In contrast to previous identification studies, which have been conducted across species (Kumar et al., 2019) and plants with substantial variation (Dubey and Thanikkal, 2023), the identification of the inbred lines to which a given leaf shape belongs has enabled phenotyping to reach "cultivar-level resolution," which is crucial for the advancement of phenotyping. The creation of leaf shape models based on semantic features is a highly interpretable process, and this study is of great significance in promoting the generation of both 2D and 3D models based on semantic features. This study presents a research pipeline for the analysis of crop phenotypic big data and the generation of models based on phenotypic big data. The pipeline includes feature extraction, semantic feature extraction and screening, construction of comprehensive indexes and atlas, and inbred line identification based on atlas. It offers novel methodologies and strategies for the extraction of knowledge from phenotypic big data, as well as a comprehensive workflow that serves as a valuable reference for the identification of plant cultivars.

TABLE 4 The r_{TopX} comparison obtained using the four methods.

r_{Top-X}	Unweighted 11 semantic features	Full features (29)	Weighted 11 semantic features (random forest)	Weighted 11 semantic features (ours)
r_{Top-1}	0.298	0.355	0.299	**0.361**
r_{Top-2}	0.393	0.452	0.391	**0.458**
r_{Top-3}	0.454	0.505	0.450	**0.525**
r_{Top-4}	0.495	0.551	0.494	**0.570**
r_{Top-5}	0.536	0.587	0.532	**0.600**
r_{Top-6}	0.563	0.616	0.555	**0.625**
r_{Top-7}	0.584	0.636	0.580	**0.653**
r_{Top-8}	0.604	0.651	0.602	**0.679**
r_{Top-9}	0.619	0.672	0.618	**0.692**
r_{Top-10}	0.637	0.690	0.636	**0.706**

Optimal performance is indicated by the use of bold text.

4.2 Limitations and future works

The study focuses on the process of analyzing 2D shape of maize leaves, with less attention on the 3D morphology and related analyses. In subsequent studies, the identification based on 2D features will be expanded to 3D features. The feature extraction and quantitative analysis will be conducted based on the 3D mesh of maize leaves. The phenotypic features obtained from the 2D and 3D meshes will be combined to realize more efficient and accurate variety identification. Additionally, further studies on the physiological properties of maize leaves, such as light interception capacity (Song et al., 2023) and the relationship between phenotype, genes, and environment, based on the 3D structure, will be conducted. For instance, the correlation between leaf phenotype and genome through GWAS will be conducted. The study also has the advantage of non-destructive acquisition of 3D models of leaves and extraction of phenotypic features. The substantial number of samples collected ensures comprehensive coverage of leaf shape diversity, and the results indicate that the 3D flattening method is area-preserving. Consequently, the method should be applied in leaf monitoring at different growth stages.

Moreover, the 2D leaf shape generation method employed in this study, while maintaining interpretability, requires further enhancement through the integration of deep learning and other techniques to augment the degree of freedom and flexibility in generation. This study serves as an initial investigation of deep learning-based 2D leaf generation, with the objective of providing accurate leaf parameter metrics and 2D leaf shape reference standards for subsequent studies. Subsequently, we will employ a deep learning-based approach to achieve the objective of 2D leaf shape generation by integrating techniques associated with generative AI, such as the diffusion model and CLIP. Concurrently, upon completion of the analysis and research of the 3D maize leaf shapes, in conjunction with the findings of the 2D models generation, the generation of 3D maize leaves with a sense of realism will be pursued as the ultimate objective of the research.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The data will be made available upon request. Requests to access these datasets should be directed to guoxy73@163.com.

Author contributions

HS: Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. WW: Data curation, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing. YiZ: Methodology, Writing – review & editing. YaZ: Data curation, Resources, Writing – review & editing. XG: Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing – original draft, Writing – review & editing. CZ: Funding acquisition, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was partially supported by the National Natural Science Foundation of China (32330075), the Construction of Collaborative Innovation Center of Beijing Academy of Agricultural and Forestry Sciences (KJCX20240406). Research Innovation Platform Construction of Beijing Academy of Agriculture and Forestry Sciences (PT2024-23).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Åström, H., Metsovuori, E., Saarinen, T., Lundell, R., and Hänninen, H. (2015). Morphological characteristics and photosynthetic capacity of *Fragaria vesca* L. winter and summer leaves. *Flora* 215, 33–39. doi: 10.1016/j.flora.2015.07.001

Bailey, B. N., and Mahaffee, W. F. (2017). Rapid measurement of the three-dimensional distribution of leaf orientation and the leaf angle probability density function using terrestrial LiDAR scanning. *Remote Sens. Environ.* 194, 63–76. doi: 10.1016/j.rse.2017.03.011

Chen, H. B., Liu, S. B., Wang, C. Y., Wang, C. F., Gong, K. Y., Li, Y. H., et al. (2023). Point cloud completion of plant leaves under occlusion conditions based on deep learning. *Plant Phenom.* 5, 13. doi: 10.34133/plantphenomics.0117

Cohen, J. I. (2011). A phylogenetic analysis of morphological and molecular characters of Lithospermum L. (Boraginaceae) and related taxa: evolutionary relationships and character evolution. *Cladistics* 27, 559–580. doi: 10.1111/j.1096-0031.2011.00352.x

Danjon, F., and Reubens, B. (2008). Assessing and analyzing 3D architecture of woody root systems, a review of methods and applications in tree and soil stability, resource acquisition and allocation. *Plant Soil* 303, 1–34. doi: 10.1007/s11104-007-9470-7

Dubey, A. K., and Thanikkal, J. G. (2023). Deep learning based identification of aquatic and semi aquatic plants. *Expert Syst.* 42 (1), 9. doi: 10.1111/exsy.13473

Dubey, A. K., Thanikkal, J. G., Sharma, P., and Shukla, M. K. (2024). A unique morpho-feature extraction algorithm for medicinal plant identification. *Expert Syst.* 9.

Gao, D. H., Li, M. Z., Zhang, J. Y., Song, D., Sun, H., Qiao, L., et al. (2021). Improvement of chlorophyll content estimation on maize leaf by vein removal in hyperspectral image. *Comput. Electron. Agric.* 184, 9. doi: 10.1016/j.compag.2021.106077

Gao, L., Pan, L. L., Shi, Y. T., Zeng, R., Li, Z. Y., Zhang, X., et al. (2024). Genetic variation in a heat shock transcription factor modulates cold tolerance in maize. *Mol. Plant* 17, 1423–1438. doi: 10.1016/j.molp.2024.07.015

Godin, C. (2000). Representing and encoding plant architecture: A review. *Ann. For. Sci.* 57, 413–438. doi: 10.1051/forest:2000132

Han, W. T., Yu, S., Xu, T. F., Chen, X. W., and Ooi, S. K. (2014). Detecting maize leaf water status by using digital RGB images. *Int. J. Agric. Biol. Eng.* 7, 45–53.

Hosoi, F., Nakabayashi, K., and Omasa, K. (2011). 3-D modeling of tomato canopies using a high-resolution portable scanning lidar for extracting structural information. *Sensors* 11, 2166–2174. doi: 10.3390/s110202166

Jafari, F., Wang, B. B., Wang, H. Y., and Zou, J. J. (2024). Breeding maize of ideal plant architecture for high-density planting tolerance through modulating shade avoidance response and beyond. *J. Integr. Plant Biol.* 66, 849–864. doi: 10.1111/jipb.13603

Kobbelt, L. (2000). *√3-subdivision, Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (ACM Press/Addison-Wesley Publishing Co.), 103–112. doi: 10.1145/344779.344835

Kumar, M., Gupta, S., Gao, X. Z., and Singh, A. (2019). Plant species recognition using morphological features and adaptive boosting methodology. *IEEE Access* 7, 163912–163918. doi: 10.1109/Access.6287639

Li, R. F., Zhang, G. Q., Liu, G. Z., Wang, K. R., Xie, R. Z., Hou, P., et al. (2021). Improving the yield potential in maize by constructing the ideal plant type and optimizing the maize canopy structure. *Food Energy Secur.* 10, 13. doi: 10.1002/fes3.v10.4

Liu, L., Zhang, L., Xu, Y., Gotsman, C., and Gortler, S. (2008). A local/global approach to mesh parameterization. *Computer Graphics Forum: J. Euro. Assoc. Computer Graphics.* 27 (5). doi: 10.1111/j.1467-8659.2008.01290.x

Lizaso, J. I., Batchelor, W. D., and Westgate, M. E. (2003). A leaf area model to simulate cultivar-specific expansion and senescence of maize leaves. *Field Crop Res.* 80, 1–17. doi: 10.1016/S0378-4290(02)00151-X

Robil, J. M., Gao, K., Neighbors, C. M., Boeding, M., Carland, F. M., Bunyak, F., et al. (2021). grasviq: an image analysis framework for automatically quantifying vein number and morphology in grass leaves. *Plant J.* 107, 629–648. doi: 10.1111/tpj.v107.2

Song, Q. F., Liu, F. S., Bu, H. Y., and Zhu, X. G. (2023). Quantifying contributions of different factors to canopy photosynthesis in 2 maize varieties: development of a novel 3D canopy modeling pipeline. *Plant Phenom.* 5, 16. doi: 10.34133/plantphenomics.0075

Song, A. R., Wang, C. Y., Wen, W. L., Zhao, Y., Guo, X. Y., and Zhao, C. J. (2024). Predicting the oil content of individual corn kernels combining NIR-HSI and multi-stage parameter optimization techniques. *Food Chem.* 461, 13. doi: 10.1016/j.foodchem.2024.140932

Sorkine, O., and Alexa, M. (2007). *As-rigid-as-possible surface modeling.* Paper presented at the Proceedings of the fifth Eurographics symposium on Geometry processing, Barcelona, Spain. 109–116.

Su, W., Zhang, M. Z., Liu, J. M., and Sun, Z. P. (2018). Automated extraction of corn leaf points from unorganized terrestrial LiDAR point clouds. *Int. J. Agric. Biol. Eng.* 11, 166–170. doi: 10.25165/j.ijabe.20181103.3177

Tian, J. G., Wang, C. L., Xia, J. L., Wu, L. S., Xu, G. H., Wu, W. H., et al. (2019). Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. *Science* 365, 658. doi: 10.1126/science.aax5482

Valentinuz, O. R., and Tollenaar, M. (2006). Effect of genotype, nitrogen, plant density, and row spacing on the area-per-leaf profile in maize. *Agron. J.* 98, 94–99. doi: 10.2134/agronj2005.0111

Wang, Y. B., Luo, Y., Guo, X., Li, Y. F., Yan, J. L., Shao, W. W., et al. (2024b). A spatial transcriptome map of the developing maize ear. *Nat. Plants* 10, 16. doi: 10.1038/s41477-024-01683-2

Wang, J. L., Wang, C. Y., Lu, X. J., Zhang, Y., Zhao, Y. X., Wen, W. L., et al. (2022). Dissecting the genetic structure of maize leaf sheaths at seedling stage by image-based high-throughput phenotypic acquisition and characterization. *Front. Plant Sci.* 13, 17. doi: 10.3389/fpls.2022.826875

Wang, R. F., Zhong, Y. T., Han, J. N., Huang, L. L., Wang, Y. Q., Shi, X. G., et al. (2024a). NIN-LIKE PROTEIN3.2 inhibits repressor Aux/IAA14 expression and enhances root biomass in maize seedlings under low nitrogen. *Plant Cell* 36, 4388–4403. doi: 10.1093/plcell/koae184

Wen, W. L., Wang, Y. J., Wu, S., Liu, K., Gu, S. H., and Guo, X. Y. (2021). 3D phytomer-based geometric modelling method for plants-the case of maize. *Aob Plants* 13, 13. doi: 10.1093/aobpla/plab055

Wen, W., Wang, J., Zhao, Y., Wang, C., Liu, K., Chen, B., et al. (2024a). 3D morphological feature quantification and analysis of corn leaves. *Plant phenom. (Washington D.C.)* 6, 0225. doi: 10.34133/plantphenomics.0225

Wen, W., Wu, S., Lu, X., Liu, X., Gu, S., and Guo, X. (2024b). Accurate and semantic 3D reconstruction of maize leaves. *Comput. Electron. Agric.* 217, 108566. doi: 10.1016/j.compag.2023.108566

Wu, J., Pagès, L., Wu, Q., Yang, B., and Guo, Y. (2015). Three-dimensional architecture of axile roots of field-grown maize. *Plant Soil* 387, 363–377. doi: 10.1007/s11104-014-2307-2

Wu, S., Wang, J. L., Zhao, Y. X., Wen, W. L., Zhang, Y., Lu, X. J., et al. (2022). Characterization and genetic dissection of maize ear leaf midrib acquired by 3D digital technology. *Front. Plant Sci.* 13, 19. doi: 10.3389/fpls.2022.1063056

Wu, S., Wen, W. L., Xiao, B. X., Guo, X. Y., Du, J. J., Wang, C. Y., et al. (2019). An accurate skeleton extraction approach from 3D point clouds of maize plants. *Front. Plant Sci.* 10, 14. doi: 10.3389/fpls.2019.00248

Wu, S., Zhang, Y., Zhao, Y. X., Wen, W. L., Wang, C. Y., Lu, X. J., et al. (2024). Using high-throughput phenotyping platform MVS-Pheno to decipher the genetic architecture of plant spatial geometric 3D phenotypes for maize. *Comput. Electron. Agric.* 225, 15. doi: 10.1016/j.compag.2024.109259

Xia, X. Y., Wei, Q. H., Wu, H. X., Chen, X. Y., Xiao, C. X., Ye, Y. P., et al. (2024). *Bacillus* species are core microbiota of resistant maize cultivars that induce host metabolic defense against corn stalk rot. *Microbiome* 12, 19. doi: 10.1186/s40168-024-01887-w

Yang, X. H., Gao, S. B., Xu, S. T., Zhang, Z. X., Prasanna, B. M., Li, L., et al. (2011). Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breed.* 28, 511–526. doi: 10.1007/s11032-010-9500-7

Yang, J., Liu, Z., Liu, Y. B., Fan, X. J., Gao, L., Li, Y. P., et al. (2024). Genome-wide association study identifies quantitative trait loci and candidate genes involved in deep-sowing tolerance in maize (*Zea mays* L.). *Plants-Basel* 13, 15. doi: 10.3390/plants13111533

Zheng, B. Y., Shi, L. J., Ma, Y. T., Deng, Q. Y., Li, B. G., and Guo, Y. (2008). Comparison of architecture among different cultivars of hybrid rice using a spatial light model based on 3-D digitising. *Funct. Plant Biol.* 35, 900–910. doi: 10.1071/FP08060

Zheng, C. X., Wen, W. L., Lu, X. J., Chang, W. S., Chen, B., Wu, Q., et al. (2022). Three-dimensional wheat modelling based on leaf morphological features and mesh deformation. *Agronomy-Basel* 12, 21. doi: 10.3390/agronomy12020414

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 10 March 2025
DOI 10.3389/fpls.2025.1449626

Check for updates

OPEN ACCESS

EDITED BY
Wenyu Zhang,
Jiangsu Academy of Agricultural Sciences Wuxi Branch, China

REVIEWED BY
Adarsh Krishnamurthy,
Iowa State University, United States
Qian Wu,
Jiangsu Academy of Agricultural Sciences (JAAS), China

*CORRESPONDENCE
Shengyong Xu
xsy@mail.hzau.edu.cn

RECEIVED 15 June 2024
ACCEPTED 06 February 2025
PUBLISHED 10 March 2025

CITATION
Wu S, Hu C, Tian B, Huang Y, Yang S, Li S and Xu S (2025) A 3D reconstruction platform for complex plants using OB-NeRF.
Front. Plant Sci. 16:1449626.
doi: 10.3389/fpls.2025.1449626

COPYRIGHT
© 2025 Wu, Hu, Tian, Huang, Yang, Li and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A 3D reconstruction platform for complex plants using OB-NeRF

Sixiao Wu[1], Changhao Hu[1], Boyuan Tian[1], Yuan Huang[2], Shuo Yang[3], Shanjun Li[1] and Shengyong Xu[1]*

[1]China College of Engineering/Key Laboratory of Agricultural Equipment for the Middle and Lower Reaches of the Yangtze River, Ministry of Agriculture, Huazhong Agricultural University, Wuhan, China, [2]College of Horticulture and Forestry Sciences, National Key Laboratory for Germplasm Innovation & Utilization of Horticultural Crops, Huazhong Agricultural University, Wuhan, China, [3]Xianning Academy of Agricultural Science, Xianning, China

Introduction: Applying 3D reconstruction techniques to individual plants has enhanced high-throughput phenotyping and provided accurate data support for developing "digital twins" in the agricultural domain. High costs, slow processing times, intricate workflows, and limited automation often constrain the application of existing 3D reconstruction platforms.

Methods: We develop a 3D reconstruction platform for complex plants to overcome these issues. Initially, a video acquisition system is built based on "camera to plant" mode. Then, we extract the keyframes in the videos. After that, Zhang Zhengyou's calibration method and Structure from Motion(SfM)are utilized to estimate the camera parameters. Next, Camera poses estimated from SfM were automatically calibrated using camera imaging trajectories as prior knowledge. Finally, Object-Based NeRF we proposed is utilized for the fine-scale reconstruction of plants. The OB-NeRF algorithm introduced a new ray sampling strategy that improved the efficiency and quality of target plant reconstruction without segmenting the background of images. Furthermore, the precision of the reconstruction was enhanced by optimizing camera poses. An exposure adjustment phase was integrated to improve the algorithm's robustness in uneven lighting conditions. The training process was significantly accelerated through the use of shallow MLP and multi-resolution hash encoding. Lastly, the camera imaging trajectories contributed to the automatic localization of target plants within the scene, enabling the automated extraction of Mesh.

Results and discussion: Our pipeline reconstructed high-quality neural radiance fields of the target plant from captured videos in just 250 seconds, enabling the synthesis of novel viewpoint images and the extraction of Mesh. OB-NeRF surpasses NeRF in PSNR evaluation and reduces the reconstruction time from over 10 hours to just 30 Seconds. Compared to Instant-NGP, NeRFacto, and NeuS, OB-NeRF achieves higher reconstruction quality in a shorter reconstruction time. Moreover, Our reconstructed 3D model demonstrated superior texture and geometric fidelity compared to those generated by COLMAP and Kinect-based reconstruction methods. The R^2 was

0.9933,0.9881 and 0.9883 for plant height, leaf length, and leaf width, respectively. The MAE was 2.0947, 0.1898, and 0.1199 cm. The 3D reconstruction platform introduced in this study provides a robust foundation for high-throughput phenotyping and the creation of agricultural "digital twins".

KEYWORDS

neural radiance fields, 3D reconstruction, plant phenotyping, digital twins, mesh

1 Introduction

3D reconstruction of plants facilitates the high-throughput phenotypic and the realization of "digital twins" in agriculture (Kim and Heo, 2024; Peladarinos et al., 2023; Pylianidis et al., 2021). High-throughput phenotypic analysis is an essential component of plant science and plays a significant role in agricultural production and genotype-phenotype studies. In agricultural production, the revolutionary integration of emerging sensors with artificial intelligence (AI) has transformed precision agriculture (PA), significantly enhancing the efficiency, effectiveness, and productivity of agricultural industry breeding and primary production (Sishodia et al., 2020). Consequently, the comprehensive analysis of plant phenotypes for monitoring plant growth has become increasingly crucial (Fu et al., 2020). In the field of genotype-phenotype studies, recent years have seen tremendous advances in plant genome sequencing, propelling the research of crop improvement that combines genotyping with phenotyping (Sun et al., 2022). However, progress in phenotypic analysis has been slow, hindering development. Traditional phenotypic methods largely rely on manual measurements, which are time-consuming, labor-intensive, and cannot guarantee accuracy. Furthermore, manual measurements cannot track the entire plant lifecycle comprehensively (Xiao et al., 2022). High-throughput phenotypic analysis based on 3D reconstruction overcomes the aforementioned drawbacks of traditional methods, emerging as a potent tool for evaluating plant phenotypes.

3D reconstruction technologies can be categorized into active and passive vision systems (Lu et al., 2020). Active vision is a reconstruction method that utilizes its light source to measure distances by projecting it onto the object. LiDAR and depth cameras are the mainstream active vision imaging devices. LiDAR is a commonly employed technique for acquiring 3D plant data (Luo et al., 2021; Tsoulias et al., 2023; Walter et al., 2019), renowned for its high precision and resolution; however, it is costly and time-consuming. In recent years, with Microsoft's introduction of the cost-effective Kinect series of depth cameras, methods based on depth cameras for 3D reconstruction have been proposed (Song et al., 2023; Teng et al., 2021a; Yang et al., 2019). Nevertheless, the data resolution obtained from depth cameras is low, and due to the unique structure of plants, the 3D representation capability of depth cameras is significantly degraded by distortion and noise, this significantly hampers the accuracy of reconstructing individual plants, especially at the finer organ scale.

Passive vision systems employ cameras to capture images of objects, extracting 3D data through image analysis. Conventional passive vision techniques include Structure from Motion (SfM) paired with MultiView Stereo (MVS), as well as voxel carving. Voxel carving utilizes segmentation masks to reconstruct objects from various perspectives, and it has been widely adopted for 3D reconstruction and the phenotypic analysis of plants (Das Choudhury et al., 2020; Golbach et al., 2015; Scharr et al., 2017). However, voxel carving is generally confined to conventional plant science laboratory environments where multiple camera positions are static and can be precisely calibrated (Feng et al., 2023). Moreover, when applied to larger plants, such as maize, voxel carving can only proceed within a limited number of views. it struggles to robustly reconstruct extensive 3D structures from numerous angles (e.g., 15 or more), which compromises the quality of the reconstruction (Tross et al., 2021).

SfM-MVS is extensively utilized in the analysis of morphological and structural plant phenotypes and is acknowledged as an optimal approach for creating a high-throughput, cost-effective platform for individual plant phenotyping (Wu et al., 2022). Nevertheless, it exhibits certain drawbacks: (1) To improve the quality and efficiency of the reconstruction process, it is imperative to acquire plant masks and provide uniform illumination conditions (Wu et al., 2022). (Gao et al., 2021) (2) Since the camera's intrinsic parameters and poses are derived through SfM, an additional calibration tool is essential to convert the algorithm-generated models into metric reconstructions (Li et al., 2022). (3) The time required for reconstruction is considerable.

In recent years, the advent of NeRF (Neural Radiance Fields) (Mildenhall et al., 2020), an innovative passive vision technology based on deep learning, has garnered considerable interest for its capacity to render high-fidelity reconstructions of intricate plant and agricultural environments (Hu et al., 2024). Traditional 3D reconstruction methodologies yield "explicit" 3D models, including point clouds, meshes, and voxel arrays (Schönberger and Frahm, 2016; Seitz et al., 2006). These "explicit" forms are inherently discrete, which can result in a loss of geometric and textural fidelity. In contrast, NeRF falls under the category of implicit neural representation methods, boasting the advantage of continuous

implicit functions, which not only allow for the synthesis of new viewpoint images of plants but also enable the extraction of textured mesh models and colored point clouds. So, NeRF holds promise as a key bridge between two-dimensional imaging and 3D reconstruction —two methods of acquiring plant phenotypes. It thereby facilitates the establishment of low-cost, high-throughput, and non-destructive plant phenotyping analysis systems (Hu et al., 2024). Moreover, by leveraging this neural implicit representation, NeRF is capable of mapping real crops into a computational virtual space, facilitating a VR-style immersive interaction for users. This innovative method offers a new avenue for creating "digital twins" within the agricultural industry (Zhang et al., 2024). Despite its potential, NeRF demonstrates several limitations. 1. Reconstructing an individual plant with NeRF can take upwards of 10 hours. 2. It cannot automatically segregate a single plant from a neural radiance field that encompasses a background. 3. There is still room for improvement in NeRF's ability to reconstruct highly detailed plant geometry. 4. Its performance is highly susceptible to lighting conditions (Cui et al., 2024).

As an early attempt in the field of 3D reconstruction of individual plants using NeRF, a new 3D reconstruction platform based on Object-Based NeRF(OB-NeRF) is presented for complex plants such as citrus fruit tree seedlings, this study makes the following contributions:

1. We propose an improved version of the NeRF model, called Object-Based NeRF (OB-NeRF), which enables high-throughput, automated, and high-precision 3D reconstruction of individual plants under complex backgrounds and uneven lighting conditions.
2. We propose a camera poses global calibration strategy, which uses the predetermined camera imaging trajectory as prior information to automatically calibrate the camera pose so that the pose and size of the reconstructed plant model can be restored to the true value.
3. A plant reconstruction platform based on the multi-view images to reconstruct high-precision and proportional mesh models for complex plants is constructed, integrating a "camera to plant" video acquisition system, as well as algorithms for Keyframe extraction, estimation and calibration of camera parameters, and 3D reconstruction based on OB-Nerf.

2 Materials

The saplings used in this study were obtained from the digital orchard demonstration base at Huazhong Agricultural University (114°366403'N, 22°756488'E). The variety selected was the mandarin orange (Citrus reticulata 'Yura'), a species widely cultivated in China. The ages of the saplings ranged from 6 months to 2 years, with an average age of 18 months. The average height of the saplings was 120 cm. All saplings were grown under the same standard management practices, which included regular irrigation, fertilization, and pest control. As shown in Figure 1, we categorized saplings into three groups based on their height: the small size group with heights ranging from 60 to 90 cm, the medium size group from 90 to 130 cm, and The large-size group is defined as individuals with a height ranging from 130 cm to 170 cm. It was observed that as the saplings grew, the complexity of their canopy structure increased. The study included a total of 20 fruit tree saplings: 7 in the small-size group, 7 in the medium-size group, and 6 in the large-size group. The tree seedlings were transferred into flower pots before the experiment. the plant height, leaf width, and leaf width are initially recorded by manual measurement. Subsequently, a video acquisition system obtained multi-view videos of the plants, serving as the data source for subsequent experiments.

3 Methods

The pipeline of the 3D reconstruction method proposed in this paper is shown in Figure 2. Firstly, an image acquisition system automatically captures Multi-view videos of plants. Then, transmit videos to the computer. Secondly, the Zhengyou Zhang calibration method (Zhang, 1999) is used to obtain the parameters of cameras. Moreover, the camera poses are estimated by SFM. Thirdly, the camera poses are calibrated by the camera poses global calibration strategy based on Hardware-Software Co-Design. Finally, we reconstruct the target plant's neural radiance field using the proposed OB-Nerf algorithms and the Marching Cubes algorithm extracts the mesh model from the neural radiation field. The development and testing of the reconstruction pipeline were conducted on a computer equipped with an Intel Core i7-12700H processor, 16GB memory, and an NVIDIA GeForce RTX 3070Ti Laptop GPU. The computer operating system is Windows 11.



FIGURE 1
Experimental materials.

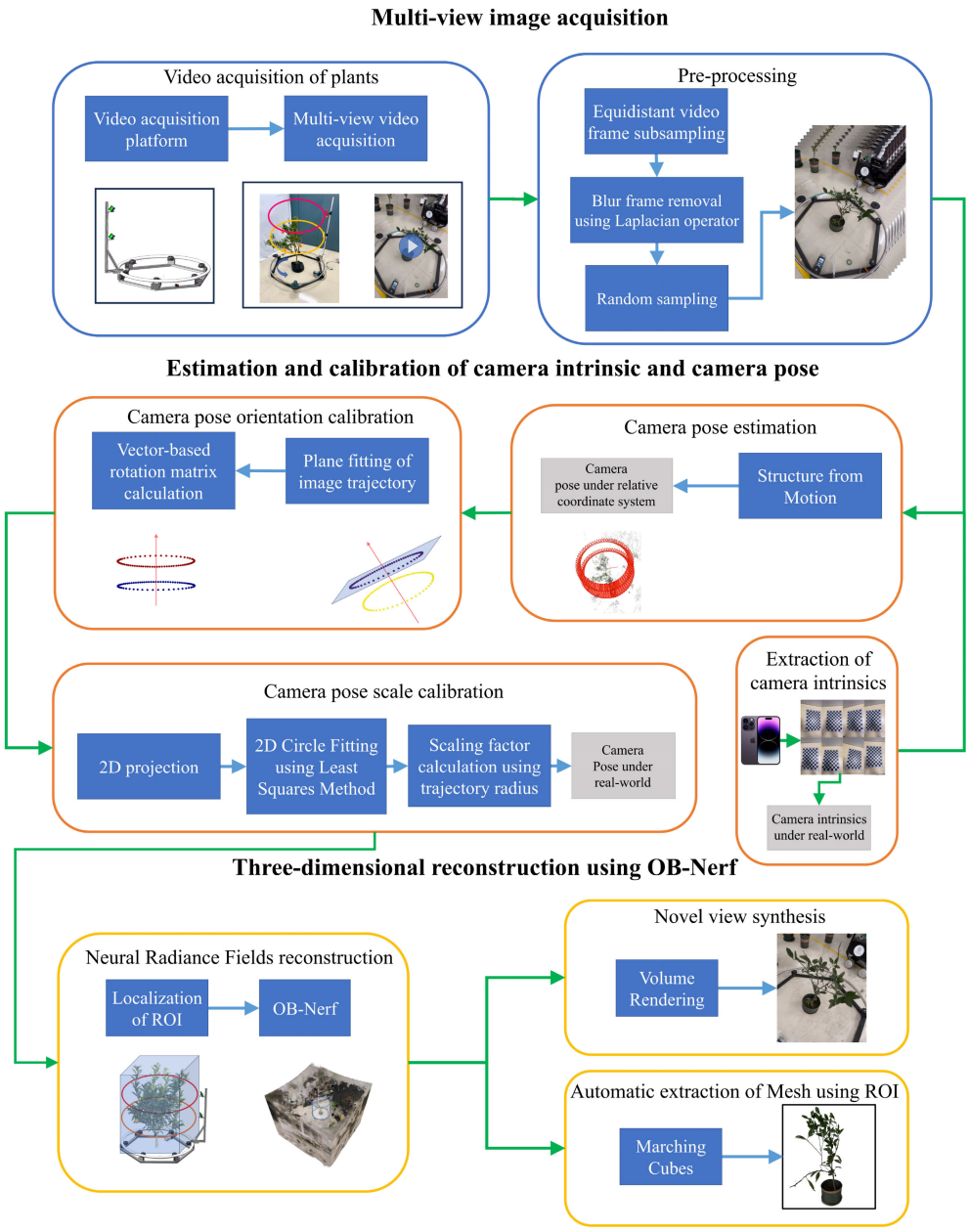


FIGURE 2
The pipeline of 3D reconstruction method.

3.1 Multi-view images acquisition

3.1.1 Video acquisition system

An automated system for video acquisition is presented, as depicted in Figure 3. Our system comprises several components: a controller, limit switches, a hexagonal base, a direct current (DC) motor, a driven pulley, a metal ring, a camera bracket, a camera, and counterweights. Notably, the metal ring has a radius of 75 cm, and the base of the camera bracket is made from an aluminum profile with a slotted design. This design allows for the radius of the imaging trajectory to be adjustable, providing a range of up to 30 cm. The height of the two cameras is adjustable within a range of 0-1.5 meters. This flexibility enables us to adjust the radius and height of the camera's imaging trajectory to accommodate plants of varying sizes. This paper proposes a linear procedure for video acquisition as follows:

3.1.1.1 Device setup

The target plant is centrally positioned within the system. Adjusting the radius and camera height to accommodate the size of the target plant, we recorded the radius r_{real}, the height of the lower camera from the ground h_1, and the height h_2 from the upper camera to the top of the bracket as prior knowledge for the reconstruction algorithm.

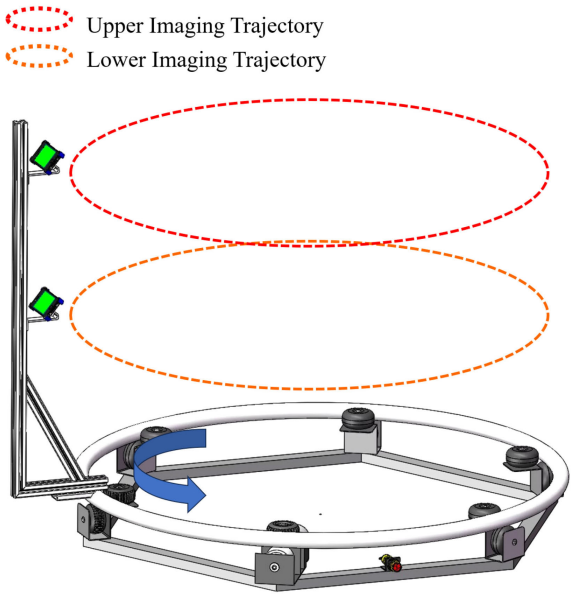


FIGURE 3
Hardware of video acquisition system.

3.1.1.2 Video capture

Once the system receives the start signal sent by the remote control, the DC motor drives the metal ring equipped with the camera bracket to rotate. The camera orbits the plant, capturing a 360-degree video. Once the video capture is complete, the DC motor automatically stops. Each capture is completed within 15 seconds. Our System is compatible with any standard RGB imaging device. For our experimental setup, we selected two iPhone 14 Pro smartphones, which offer a video resolution of 1080x1920 at a frame rate of 30fps.

3.1.2 Keyframe extraction

In the Structure from Motion (SFM) process, we observe that the relationship between the number of images and the reconstruction time follows a quadratic function, with the reconstruction time increasing dramatically as the number of images increases. Insufficient image quantity and image blur both negatively impact the reconstruction quality when using OB-NeRF for reconstruction. Therefore, this research aims to find the optimal balance between reconstruction quality and time cost through a series of experiments, which are detailed in sections 4.2.1 and 4.2.2. The experimental results demonstrate that high-quality 3D reconstruction can be achieved in a relatively short time by excluding blurry frames and limiting the number of images to approximately 90, which is a significant improvement compared to using all available images. To achieve this, we develop an intelligent frame extraction algorithm that can automatically select 90 clear frames from the input video for subsequent reconstruction, effectively reducing the computational burden while maintaining the reconstruction quality. The algorithm workflow is as follows:

1. Perform equidistant downsampling on the binocular videos obtained by the video acquisition system to acquire 50 images from each perspective, totaling 100 images. Uniform downsampling can reduce the number of images and effectively decrease redundancy by avoiding the repeated selection of similar frames.
2. Eliminate blurry frames. To quantitatively assess the sharpness of the images, this study adopted a method based on the variance of the Laplacian operator. First, the RGB image is converted to grayscale, and then the Laplacian operator L is calculated using an 8-neighborhood convolution kernel.Finally, we calculate the variance of the resulting Laplacian image to obtain a variance value, which serves as a measure of sharpness. The greater the variance, the higher the image sharpness. The mathematical expression for variance is shown in Equation 1:

$$\mathrm{Var}(L) = \frac{1}{N}\sum_{i=1}^{N}(L_i - \mu)^2 \tag{1}$$

We calculate the average variance of all images and eliminate those that are less than 20% of the average value. If the number of images removed exceeds 10, it is determined that the data quality is not up to standard, and image acquisition needs to be repeated.

3. To ensure a total of 90 images, we conduct a random sampling of the remaining images after eliminating blurry frames, extracting 45 images from each perspective, a total of 90 images.

To enhance efficiency, we employed multithreaded parallel computing techniques for equidistant downsampling and the removal of blurred frames. Consequently, the overall preprocessing duration was reduced to roughly 3 seconds.

3.2 Estimation and calibration of camera parameters

3.2.1 Camera parameters estimation

The camera projection model defines the mapping from a 3D world to a 2D image plane. A is the camera intrinsic matrix, R and t are the rotation and translation, k_1, k_2, k_3 are the distortion coefficients, that describe the change of coordinates from world to camera coordinate systems.

3.2.1.1 Camera intrinsic parameters extraction

The precise camera projection model was established in this study using Zhang Zhengyou's calibration method. The method involves the following steps: preparing a calibration board, capturing calibration images, extracting corners of the calibration board, establishing correspondences between corners, computing the camera's intrinsic parameters, estimating distortion parameters,

and Optimizing the parameters mentioned above with the L-M algorithm. We Prepared a chessboard with a grid size of 40 mm * 40 mm and dimensions of 12 × 9 corners and obtain the internal parameters $A = \begin{bmatrix} 872.2 & 0 & 540.6 \\ & 873.1 & 957.9 \\ 0 & 0 & 1 \end{bmatrix}$ and distortion coefficients $k_1 = 6.2 \times 10^{-3}$, $k_2 = -3.95331432 \times 10^{-2}$, $k_3 = 7.5 \times 10^{-5}$, $p_1 = -1.63556310 \times 10^{-4}$, $p_2 = 4.11907012 \times 10^{-2}$.

3.2.1.2 Camera pose estimation

SFM can estimate the camera poses, which are inputs to the OB-NeRF. The input of SFM includes a set of plant image sequences, and the camera intrinsic parameters. Initializing camera intrinsic parameters at the beginning of SFM provides a robust initial estimate, which reduces calculated errors, accelerates algorithmic convergence, and enhances the accuracy of pose estimation. SFM outputs the camera pose matrix corresponding to each image.

3.2.2 Camera poses global calibration strategy based on Hardware-Software Co-Design

The camera poses obtained through the SFM are estimated within a relative coordinate system, hereafter referred to as the "virtual coordinate system". Since the virtual coordinate system is not directly aligned with the real-world coordinate system, the reconstructed plant model's dimensions and orientation may not match the real-world counterparts. Aligning the coordinate systems is crucial for accurate camera pose calibration, as it establishes the transformation relationship between the virtual coordinate system and the real-world coordinate system. The transformation relationship between coordinate systems encompasses both orientation and scale relationships. The orientation relationship is characterized by the rotation matrix R_{vtr}. The scale relationship is denoted by the scaling factor k. The calibration formulae for the camera's poses as delineated in Equations 2 and 3:

$$R = R_{vtr} \cdot R_{\text{virtual}}, t_{virtual} = R_{vtr} \cdot t_{virtual} \tag{2}$$

$$t = k \cdot t_{\text{virtual}} \tag{3}$$

We propose an automatic camera pose global calibration strategy based on Hardware-Software Co-Design. It leverages the camera's predetermined imaging trajectory as a reference. More precisely, the normal vector of the trajectory plane is utilized as a direction reference for the coordinate systems, Concurrently, the diameter of the trajectory is utilized as a scale reference for the coordinate systems. Moreover, the process of this method is as follows:

1. Orientation calibration of the camera pose using the Least squares method and Rodrigues' rotation equation. The imaging trajectory is horizontal, suggesting that the normal vector of the trajectory plane is parallel to the z-axis of the real world. Initially, we use the camera positions $A = \{(x_1, y_1, z_1),(x_2, y_2, z_2),\ldots,(x_n, y_n, z_n)\}$ restored by SFM as a discrete representation of the camera imaging trajectory, fit the imaging trajectory to a plane using the Least squares method to acquire the normal vector (a, b, c). Thirdly, calculate the rotation matrix R_{vtr} based on the normal vector (a, b, c) of the imaging trajectory plane and the direction vector (0, 0, 1) of the z-axis using Rodrigues' rotation equation. Finally, apply the orientation calibration formulae to the camera poses.
2. Calculation of scale factor using imaging trajectory. The scale factor k is calculated based on the actual radius of the trajectory as prior information. First, we project the camera trajectory points after orientation calibration onto the XY plane. Secondly, fit the 2D trajectory points to a circle $(x - x_c)^2 + (y - y_c)^2 = r^2_{virtual}$ using the Least squares. Thirdly, calculate the scale factor k based on $r_{virtual}$ and the imaging trajectory's real radius r_{real} according to Equation 4. Finally, apply the scale calibration formula to the camera poses.

$$k = \frac{r_{real}}{r_{virtual}} \tag{4}$$

Moreover, to accelerate the training process of OB-Nerf, we translate the coordinate system, thereby shifting the target plant to the center of the coordinate system. First, we separate the camera imaging positions of the upper and lower tracks by utilizing the height difference and calculate the average zcoordinate of the two-point sets, denoted as z_{up} and z_{low}. Second, we calculate the translation vector $\vec{t}_s$ according to Equation 5. Finally, apply the translation formula (Equation 6) to the camera poses.

$$t_s = (x_c, y_c, z_c)^T \tag{5}$$

Where: $z_c = \frac{z_{up}+z_{low}}{2}$.

$$t = t - t_s \tag{6}$$

3.3 3D reconstruction using object-based Nerf

3.3.1 Localization of region of Interest

Traditional multi-view image-based plant 3D reconstruction methods typically utilize background removal to enhance reconstruction accuracy and computational efficiency. But in natural scenes, it poses a challenge, thus necessitating setting up a solid-colored studio, which is difficult and costly for "camera to plant" mode. To achieve high-quality reconstruction of the target plant without performing background segmentation, it is essential to determine the spatial region in which the plant is located within the scene. This region is called the Region of Interest (ROI), which acts as prior knowledge for the OB-Nerf algorithm (Figure 4). Due to its high computational efficiency in the ray sampling process, we use the axis-aligned bounding box (AABB) to represent the ROI. An AABB is a rectangular box defined by its minimum and maximum coordinates $(x_{\min}, y_{\min}, z_{\min})$ and $(x_{\max}, y_{\max}, z_{\max})$, which are parallel to the coordinate axes. In section 3.3.2, we computed the heights of two imaging trajectories, z_{up} and z_{low}. The calculation

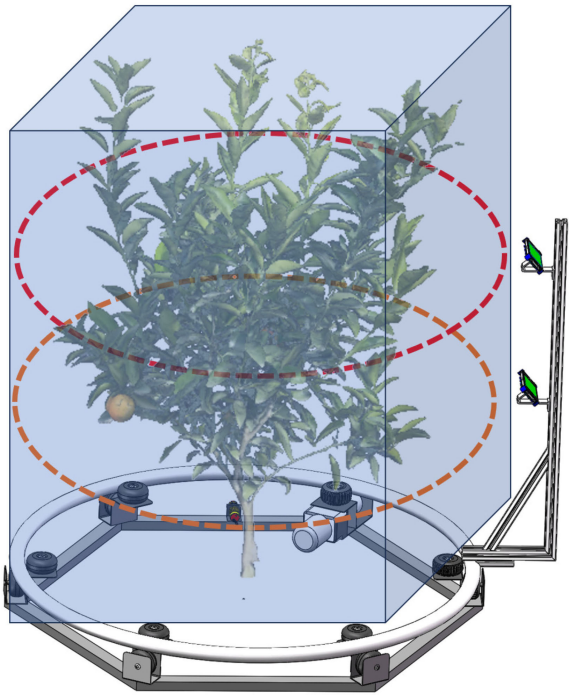


FIGURE 4
Location of the region of interest.

formula for AABB as illustrated in Equation 7:

$$\begin{cases} (x_{\min}, y_{\min}, z_{\min}) = (-r_{real}, -r_{real}, z_{low} - h_1) \\ (x_{\max}, y_{\max}, z_{\max}) = (r_{real}, r_{real}, z_{up} + h_2) \end{cases} \quad (7)$$

3.3.3 Object-based NeRF

Neural Radiance Fields implicitly represent 3D scenes through neural network approximations of 3D density fields and 5D color fields. The radiation field describes the volumetric density at every point in a scene, as well as the color of each point from every observational direction, expressible by the function: $F_\theta : (x,d) \rightarrow (c, \sigma)$. Initially, the light rays emitted from the camera are sampled to determine the positions of each sampling point. Subsequently, these positions are encoded and mapped into a high-dimensional space. Subsequently, these encoded inputs are fed into a multilayer perceptron, which is utilized to approximate the 3D density field, thereby obtaining the volume density corresponding to each point. Next, the volume density output is concatenated with the view direction vector, and these combined inputs are fed into another multilayer perceptron, designed to fit the 5D light field, thereby retrieving color for each point in the observed direction. Finally, the image is synthesized through volume rendering, and the network is optimized by minimizing the loss between the synthesized and actual view color values using backpropagation. This study proposes targeted enhancement strategies to address the issues encountered by the NeRF model in plant 3D reconstruction:

1. A critical requirement for achieving high-throughput reconstruction is completing the process of reconstructing individual plants within a short time frame. We have employed a shallow multilayer perceptron to approximate the density and color fields.
2. In phenotyping, reconstructing individual plants requires high fidelity not only at the macroscopic level but also at the organ level. Plant organs often have small and intricate structures; for example, a 1.5meter-tall sapling may have branches with diameters as small as 2 millimeters.Existing NeRF algorithms have limitations in representing fine geometric details. To address this issue, we propose three steps: a novel ray sampling strategy, Hash encoding to map input points to a higher-dimensional space, and an optimization framework to refine camera poses.
3. Removing image backgrounds enhances plant reconstruction but obtaining masks in natural scenes is challenging. Traditional solid-colored studio setups are impractical for "plant-to-camera" modes. We propose a ray sampling strategy integrated with prior knowledge in OB-NeRF to improve reconstruction without plant masks.
4. A significant hurdle in reconstruction under natural scene is the presence of uneven spatial lighting. To address this issue, we incorporate an exposure adjustment phase into OB-NeRF.

These strategies result in the creation of the OB-NeRF model, as shown by its internal structure in Figure 5.

3.3.3.1 Ray sampling strategy based on region of Interest

To optimize the neural radiance field, the first step is to generate a set of sampled 3D points along the camera rays passing through each pixel of the input images. The scene can be divided into ROI and background. ROI is the area where the target plant for reconstruction is located and is dense in volume in space. The quality of the neural field in ROI determines the reconstruction's quality in the target plant. Moreover, there are many empty spaces in the scene. Sampling points falling in these empty spaces do not contribute to the optimization process. Instead, they increase computational and memory overhead. We utilize a hierarchical sampling approach, and optimize two networks to reconstruct the scene: a "coarse" network and a "fine" network, with the output of the "coarse" network guiding the sampling point selection for the "fine" network. Therefore, we implement a new ray sampling strategy that focuses the neural field optimization process on the ROI area and skips the empty spaces, thereby improving the reconstruction quality of the target plant and reducing reconstruction time.

First, we calculate the intersection point between the ray with origin $o = (o_x, o_y, o_z)$ and direction $d = (d_x, d_y, d_z)$ and the ROI's AABB. The formula is presented in Equation 8:

$$\begin{cases} t_{\min} = \max\left(\frac{x_{\min}-o_x}{d_x}, \frac{y_{\min}-o_y}{d_y}, \frac{z_{\min}-o_z}{d_z}\right) \\ t_{\max} = \min\left(\frac{x_{\max}-o_x}{d_x}, \frac{y_{\max}-o_y}{d_y}, \frac{z_{\max}-o_z}{d_z}\right) \end{cases} \quad (8)$$

Where: t_{max} and t_{min} represent the distance from the intersection point to the origin of the ray. The ray intersects the

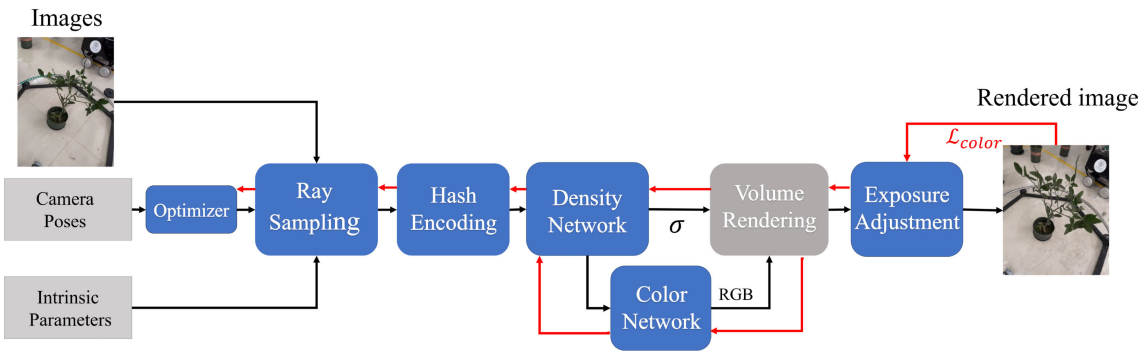


FIGURE 5
The pipeline of OB-Nerf.

AABB if and only if $t_{\min} \leq t_{\max}$ and $t_{\max} \geq 0$. Because the camera imaging position is contained within the AABB, $t_{\min} \leq 0$. We assign a value of 0 to $t_{\min}$.

Secondly, the intersection point divides the ray into two regions, $(0, t_{max})$ and (t_{max}, t_b), where t_b represents the maximum length of the ray. We generate sampling points with different sampling intervals in two regions and obtain a total of N_c sampling points(Figure 6a). Thirdly, we obtain the output ω_i by inputting the n_c sampling points into the coarse network, as illustrated in Equation 9, and normalization is applied to ω_i to determine the respective constant probability density function $\hat{\omega}_i = \frac{\omega_i}{\sum_{j=1}^{N_c} \omega_j}$ in two regions. Subsequently, total N_f points are obtained through inverse transform sampling.

$$\omega_i = T_i(1 - \exp(\sigma_i \delta_i)), \text{where} T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right) \tag{9}$$

where $\delta_i = t_{i+1} - t_i$ is the distance between adjacent samples

Finally, we combine N_c and N_f points as "fine" network input points.

3.3.3.2 Shallow MLP

To enhance the speed of the 3D reconstruction pipeline, we employ shallow MLP networks for approximating the density and color fields. The architecture of the two fully connected networks is illustrated in Figure 6b, input vectors are represented by green rectangular blocks, intermediate hidden layers by blue rectangular blocks, and output vectors by red rectangular blocks. The number within each block indicates the vector's dimension. All layers consist of standard fully connected layers. Black arrows signify layers with ReLU activation, dashed black arrows represent layers with sigmoid activation, and dotted black arrows with a black dot indicate layers with ELU activation. The "+" symbol denotes vector concatenation. The density network takes the positional encoding of the input location and produces a volume density and a 15-dimensional geometric feature vector. The color network takes the 16-dimensional geometric feature along with the positional encoding of the viewing direction to output RGB values.

3.3.3.3 Multi-resolution hash encode

We employ a shallow multi-layer perceptron to fit color and density fields. However, shallow MLP tends to learn low-frequency functions, resulting in poor reconstruction of high-frequency details in color and geometric shapes, failing to effectively represent the plant organs. To address this issue, the study adopted multi-resolution hash techniques from Instant-NGP (Müller et al., 2022) for position encoding (Figure 6c). The hash table parameters are shown in Table 1. In terms of implementation details, to avoid increasing the computational load by simply augmenting the number of hash table layers, this study optimizes the design of the hash table. In the final level of the hash table, space is divided into 2048 hash voxels instead of the conventional 512. This design does not alter the dimension of the hash encoding output. Moreover, it enhances the encoding capacity at finer scales, thereby improving the fine-scale reconstruction performance of OB-NeRF.

3.3.3.4 Exposure adjustment

In natural environments, scene lighting conditions are not uniform, leading to varying exposure levels in images from different perspectives. Conventional 3D reconstruction often involves setting up controlled lighting environments to provide stable and uniform lighting. To achieve 3D reconstruction under uncontrolled lighting conditions, this study incorporates the exposure rate E_i of each image as a learnable parameter into the optimization process, incorporating the exposure adjustment into the forward propagation. By adjusting color values during image synthesis with the learned exposure rates, the study compensates for the exposure variations in images. Through the backpropagation algorithm, the study dynamically updates the exposure rate for each image based on gradient information from the loss function. An L2 regularization term is utilized to supervise the optimization of exposure rates to prevent overfitting. With the exposure adjustment step, we substantially enhance the accuracy of 3D reconstruction under uneven lighting conditions and improve the algorithm's robustness to complex and variable lighting scenarios.

10.3389/fpls.2025.1449626

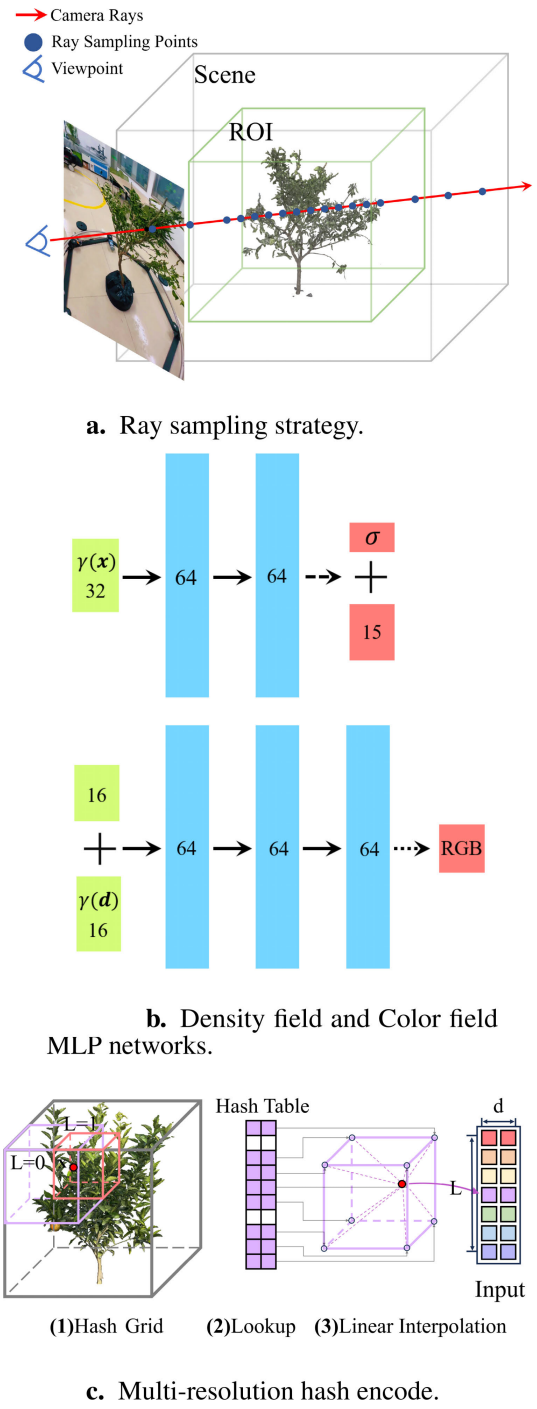


FIGURE 6
Improvement strategies of OB-NeRF. **(a)** Ray sampling strategy. **(b)** Density field and Color field MLP networks. **(c)** Multi-resolution hash encode.

In our proposed method, the pixel values of the synthesized images are adjusted by calculating the scaling factors $S_i$corresponding to the exposure rate E_i of each image. The exposure rate E_i, i represents the exposure level of image i, while the scaling factor S_i is a multiplicative factor used to adjust the pixel values of the synthesized image i. The adjustment of pixel values can be expressed as Equations 10 and 11:

$$S_i = e^{\ln(2)\times E_i} \tag{10}$$

$$RGB_{\text{target}} = S_i \cdot RGB \tag{11}$$

TABLE 1 Hash table parameters.

Parameter	Symbol	Value
Number of levels	L	16
Max. entries per level (hash table size)	T	219
Number of feature dimensions per entry	F	2
Coarsest resolution	$N_{\min}$	16
Finest resolution	$N_{\max}$	2048

3.3.3.5 Camera pose optimizer

Precise camera pose ensures the accurate transformation from the world coordinate system to the camera coordinate system, serving as the basis of the entire rendering and training process. Although we have already globally calibrated the camera poses, we have observed potential minor deviations between the calibrated camera poses and their true values on the level of a single image. In this study, we construct a camera pose optimizer for each image that incorporates two trainable parameters: the camera's position offset and rotation offset. The optimizer, integrated into the network's forward propagation process, iteratively updates these parameters based on the loss function using the backpropagation algorithm. An L2 regularization term is utilized to supervise the optimization of the camera pose to prevent overfitting.

3.3.4 Mesh extraction from neural radiance fields

NeRF is an implicit continuous representation of the 3D scene. To discretely sample the radiance field and generate a mesh model, we employ the Marching Cubes algorithm, a classic method to extract iso-surfaces from volumetric data by approximating the surface with a polygonal mesh.

Firstly, Given the predefined 3D ROI, a set of spatial points $P = \{p_1, p_2, \ldots, p_n\}$ is generated via dense volumetric sampling. To represent the region of interest more accurately, we use the inscribed cylinder of the AABB as the sampling area. Secondly, For each point $p_i \in P$, using the OB-NeRF model to obtain the density values, $\sigma(p_i) = \text{NeRF}\sigma(p_i)$. Finally, The Marching Cubes algorithm (Lorensen and Cline, 1987) identifies the isosurface by thresholding the density values, as illustrated in Equation 12:

$$M = \text{MarchingCubes}(P, \sigma_{\text{threshold}}) \tag{12}$$

Where: M is the resultant mesh and $\sigma_{\text{threshold}}$ is an optimal density value demarcating the object's boundary.

4 Result and discussion

4.1 Evaluation of 3D reconstruction performance

In this section, we employed our 3D reconstruction platform to reconstruct twenty saplings from three size categories, representing distinct growth stages, for verifying platform performance. This categorization ensures a comprehensive analysis across the growth stages. After acquiring the sapling mesh models, We extracted the plant height from the Mesh model and used CloudCompare software to measure leaf length and width. In addition, we manually measured plant height, leaf length, and leaf width.

We evaluate the 3D reconstruction performance of our platform using the PSNR of the generated synthetic images, along with the accuracy, spatial resolution, and texture resolution of the generated mesh models.

Peak Signal-to-Noise Ratio (PSNR) is a widely recognized quantitative metric for assessing image quality. Higher PSNR values denote better image fidelity and closer similarity to the original image. We utilized PSNR to evaluate the quality of the synthesized novel views, with PSNR calculations based on the Mean Squared Error (MSE), which quantifies the average squared difference in pixel values between the reference image and the image under assessment. Its calculation method is shown in Equations 13 and 14. As depicted in Figure 7, we present the rapid reconstruction capability and exceptional effects of the OB-NeRF algorithm, as evidenced by the synthesized images and extracted 3D mesh models. We employ PSNR as the metric to evaluate the quality of synthesized images of target plants. Figure 7 also demonstrates our capability in exposure adjustment, with the exposure values indicating deviations from the original exposure rate. These results were achieved after merely 30 s of OB-NeRF training, substantiating the proposed algorithm's significant speed advantage. By introducing a new ray sampling strategy, we embedded prior knowledge into OB-NeRF, enabling the algorithm to reconstruct target plants amidst complex backgrounds efficiently. Notably, the analysis indicates that the target plants can still be reconstructed with high fidelity (PSNR = 29.95 dB), even when the background reconstruction exhibits significant distortion (PSNR < 20 dB).

$$\mathrm{MSE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[I_{\mathrm{real}}(i,j) - I_{\mathrm{synthesized}}(i,j)\right]^2 \tag{13}$$

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}_i^2}{\mathrm{MSE}}\right) \tag{14}$$

As demonstrated in Figure 8, we present different organs of the plant mesh model, including normal leaves, curled leaves, damaged or perforated leaves, leaf clusters, and branches. The average reconstruction error of these organs is less than 2mm, and the texture resolution reaches 0.5 mm/pixel, exhibiting high geometric and texture fidelity. The reconstruction results demonstrate that our platform also has good reconstruction capabilities at the organ level of plants and validates its applicability and efficiency in handling plants with varying growth states and complexity.

The accuracy of the mesh model directly reflects the quality of the geometric reconstruction. To verify the precision of the 3D reconstruction platform developed in this study, we selected three key phenotypic parameters: tree height, leaf length, and leaf width, and analyzed the error between their estimated and true values. These parameters can reflect the overall growth status and leaf development level of seedlings and are important indicators for evaluating seedlings' phenotypes. The estimated value of the tree height is obtained by calculating the difference between the maximum and minimum values of the model on the z-axis. We manually measured each seedling using a tape measure to obtain the true values. Each fruit tree was measured three times, and the average value was taken as the final true value. The estimated leaf length and width values were obtained using the CloudCompare software. We converted the Mesh model into a point cloud and manually segmented the target leaves. In the measurement tool of CloudCompare, we selected the points at both ends of the long axis and short axis of the leaf, and the software automatically calculated the Euclidean distance between the two points, which were used as the estimated values of leaf length and leaf width. For the measurement of true values, we picked the target leaves and measured them three times each in the long-axis and short-axis directions using a vernier caliper (accuracy 0.01 mm), and the average values were taken as the true values of leaf length and leaf width. We used mean absolute error and coefficient of determination to evaluate the accuracy of the reconstruction results. MAE reflects the average size of the deviation between the estimated and true values, with smaller values indicating smaller deviations and 0 being the ideal value. R_2 reflects the goodness of fit between the estimated and true values, ranging from 0 to 1, with values closer to 1 indicating a higher goodness of fit. The MAE of tree height was 2.0947 cm, with an R_2 of 0.9933 (Figure 9a); the MAE of leaf length was 0.1899 cm, with an R^2 of 0.9881 (Figure 9b); and the MAE of leaf width was 0.1199 cm, with an R^2 of 0.9883 (Figure 9c). Experimental results demonstrate that our method achieves reconstruction accuracy at the millimeter level. The results demonstrate that our reconstruction platform has high accuracy at both the individual plant and organ levels.

Spatial resolution represents whether the model retains sufficient detail information, while texture resolution affects the perceived visual quality of the model. The spatial resolution reached 0.0019 mm, and the texture resolution reaches 0.5 mm/pixel, exhibiting high geometric and texture fidelity.

4.2 Ablation experiment

4.2.1 Correlation between the numbers of images and SFM reconstruction time consumption

To reduce the time consumption of the 3D reconstruction pipeline, we investigated the impact of the number of images on the time consumption using the SfM method. We established experimental groups with varying numbers of images, specifically 45, 60, 75, 90, 105, 120, 135, 150, 165, and 180 images. Each group included six fruit trees of different sizes as experimental subjects. To minimize random errors in the results, each fruit tree underwent 10 independent SfM reconstructions, resulting in a total of 60 reconstructions per group (10 reconstructions × 6 trees). We



FIGURE 7
3D reconstruction results.

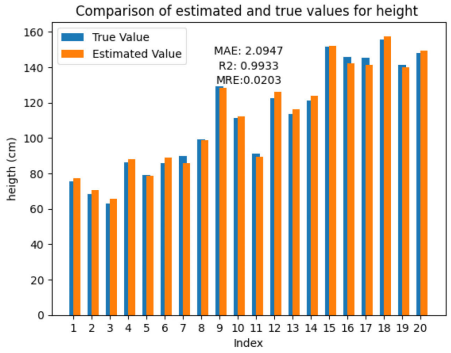
recorded the time taken for each reconstruction and calculated the average reconstruction time for each group to assess the relationship between the number of images and reconstruction duration. Ultimately, we employed Origin software for data analysis to quantitatively explore the specific impact of the number of images on reconstruction time. This analysis aided us in understanding how to select the optimal number of images to optimize the efficiency of the entire reconstruction process while ensuring the quality of the reconstruction.

we conducted a systematic analysis of the time required for the three primary steps in Structure from Motion method: feature extraction, feature matching, and sparse reconstruction. Our findings reveal that for datasets ranging from 30 to 180 images, the time needed for feature extraction scales linearly with the number of images and represents a smaller fraction of the overall processing time (Figure 10a). In contrast, the time dedicated to feature matching increases quadratically as the number of images grows, accounting for the largest portion of the total processing

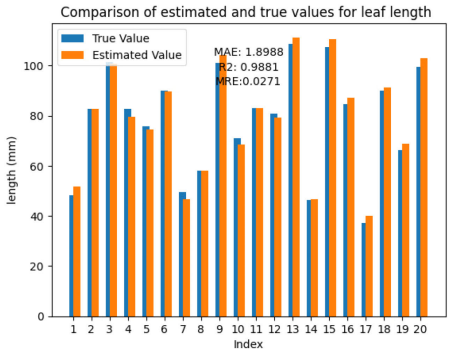
10.3389/fpls.2025.1449626



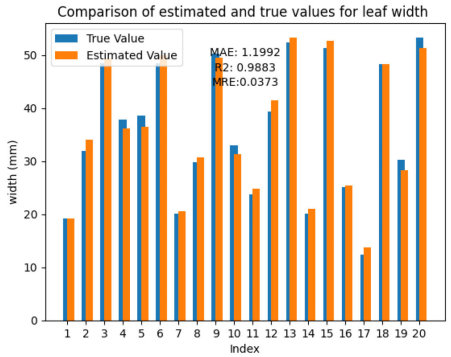
FIGURE 8
Organ-level reconstruction results. **(a)** Norma leaf. **(b)** Curled leaf. **(c)** Severely curled leaf. **(d)** Damaged or perforated leaves. **(e)** Leaf Cluster. **(f)** Branching.



a. Correlation analysis between estimated values and true values of plant height.



b. Correlation analysis between estimated values and true values of leaf length.



c. Correlation analysis between estimated values and true values of leaf width.

FIGURE 9
Correlation analysis between estimated values and true values. **(a)** plant height. **(b)** leaf length. **(c)** leaf width.

time (Figure 10b). Regarding the sparse reconstruction phase, while its time consumption also follows a linear trend with the image count, this stage exhibits considerable variability in individual reconstruction efforts (Figure 10c).

To establish a high-throughput 3D reconstruction pipeline, it is crucial to minimize the number of images used while ensuring the reconstruction's quality is not excessively compromised.

4.2.2 Correlation between the numbers of images and reconstruction quality

We assess the quality of the reconstruction from two aspects: the synthesized images and the extracted mesh models. To more intuitively represent the geometric shape of the mesh, we opt to display it using an uncolored mesh. This approach enhances the

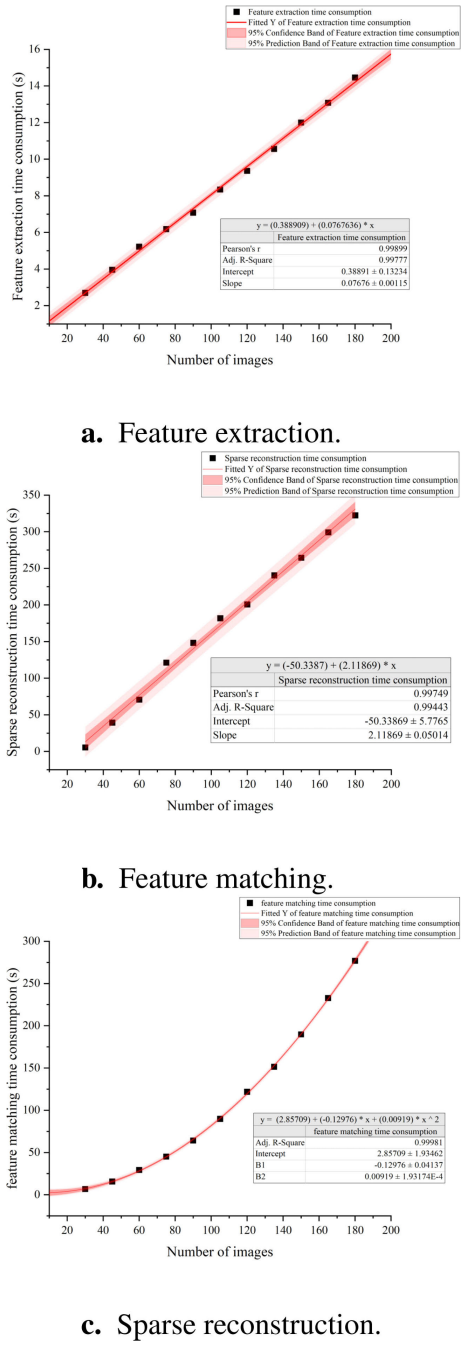


FIGURE 10
Correlation between SFM time consumption and the number of images. **(a)** Feature extraction. **(b)** Feature matching. **(c)** Sparse reconstruction.

clarity of its structural details. Figure 11 clearly illustrates the relationship between the number of images and the quality of reconstruction. The analysis indicates that a relatively high quality of reconstruction is already achieved with 90 images. However, when the number of images is increased to 180, there is no significant improvement in reconstruction quality. We conclude that selecting 90 images as the input for the reconstruction pipeline is optimal. When processing 90 images, SFM takes about 210s, the pipeline takes about 250s.

4.2.3 Assessment of training efficiency and reconstruction quality for OB-NeRF

To accelerate the training speed and enhance the reconstruction quality of the NeRF model for the target plant, we employed an optimized multi-resolution hash encoding, a shallow MLP network, exposure adjustments, camera extrinsics optimization, and a new ray sampling strategy. We designed comparative experiments between OB-NeRF, Based-NeRF, Mip-NeRF (Barron et al., 2021), Neus (Wang et al., 2023a), NeRFacto (Tancik et al., 2023) and Instant-NGP. For Neus, we use an accelerated version based on multiresolution hash encoding and open-source library NerfAcc (Wang et al., 2023b). The results, as shown in Table 2, indicate that the training time requirement for Based-NeRF ranged between 9 to 12 hours, whereas OB-NeRF significantly reduced training time. Compared with Mip-NeRF,NeRFacto and Neus, our method shows advantages in both efficiency and reconstruction quality. Compared to Instant-NGP, which also utilizes multi-resolution hash encoding, our network demonstrated a faster training speed and reconstruction quality in reconstructing target plants.

We compared the geometric reconstruction performance of NeRFacto, Instant-NGP, NeuS, and our method, with the results shown in Tables 3 and 4. Experimental results demonstrate that our method achieves higher reconstruction accuracy than NeRFacto and Instant-NGP, and is comparable to NeuS. However, NeuS requires significantly more reconstruction time, indicating lower efficiency.

4.3 Comparison with other 3D reconstruction methods

Conventional image-based plant 3D reconstruction require the acquisition of plant masks to accelerate reconstruction speed and enhance the quality of the reconstruction. Moreover, they are sensitive to lighting conditions, necessitating stable and uniform illumination. However, achieving these conditions in real scenarios is challenging. Our 3D reconstruction platforms accomplish 3D reconstruction of small to medium-sized plants against complex backgrounds without requiring plant masks. Additionally, it is robust to variations in lighting, eliminating the need for stable, uniform illumination.

The mainstream algorithm for image-based 3D reconstruction is SfM-MVS. COLMAP is widely recognized as an advanced and efficient tool for implementing SfM-MVS. We utilized COLMAP to reconstruct the same datasets. As illustrated in Figure 12a, certain saplings underwent incorrect reconstruction. In contrast, Figure 12b demonstrates that even the correctly reconstructed saplings exhibit low-quality results. The point cloud derived from the reconstruction process contains numerous floating noise points, and severe artifacts are evident along the edges of the branches and leaves. These issues contribute to the generation of low quality meshes. Furthermore, the reconstruction time using COLMAP exceeds two hours, whereas our algorithm completes the reconstruction in less than five minutes.

In recent years, the field of 3D reconstruction of plants has seen the emergence of Kinect as a novel trend. Our laboratory previously experimented with the use of Kinect for quad-view imaging, combined

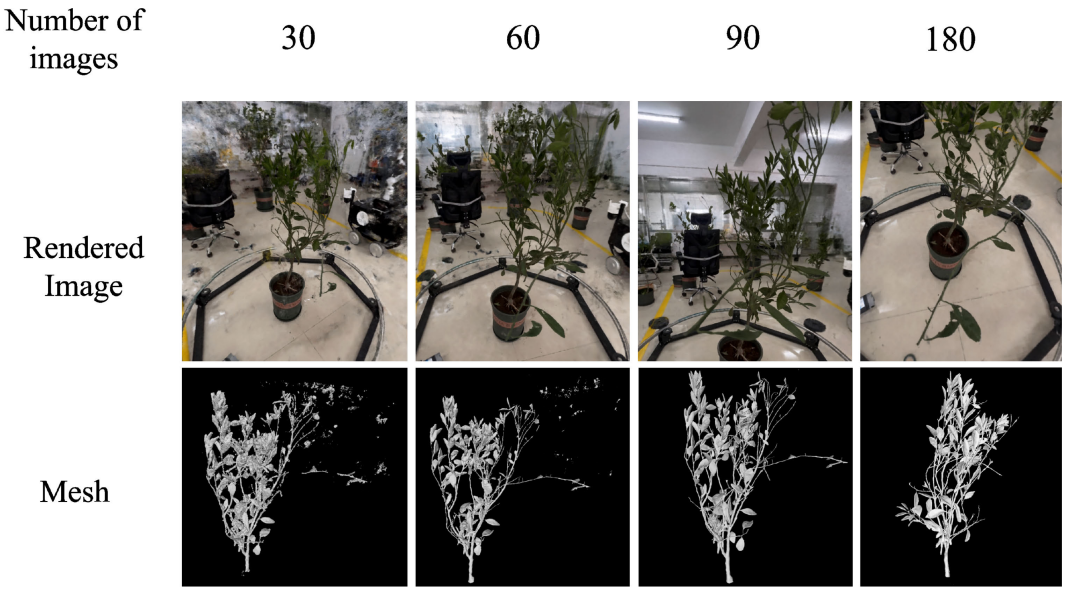


FIGURE 11
Correlation between the numbers of images and reconstruction quality.

TABLE 2 Comparison between Based-NeRF, Mip-NeRF, Instant-NGP, Neus, NeRFacto and Ours.

Method	Average Training Time	Average PSNR
Based-NeRF	10 h	24.71dB
Mip-NeRF	3 h	25.81 dB
Instant-NGP	30 s	26.79 dB
Neus	30min	27.31dB
NeRFacto	5min	27.64dB
Ours	30 s	29.95 dB

TABLE 3 H measurement results of NeRFacto, Instant-NGP, Neus and Ours.

Method	R^2	MRE (%)	MAE (cm)
Instant-NGP	0.963	2.60	2.67
Neus	0.992	2.05	2.11
NeRFacto	0.987	2.47	2.54
Ours	0.993	2.03	2.09

with Iterative Closest Point (ICP) registration techniques, successfully reconstructing a 3D model of a single rapeseed (Teng et al., 2021b). We subsequently attempted to apply this technique to the 3D reconstruction of saplings(Figure 13). However, for saplings with complex structures and larger sizes, the reconstruction results obtained with Kinect were not satisfactory. We observed that the point clouds reconstructed using Kinect exhibit significant data gaps in many areas, particularly in sections with complex structures. The point clouds are heavily affected by noise, and artifacts appear at the edges of branches and leaves. These issues prevent the reconstructed models from accurately reflecting the precise geometric structure of the plants. During the ICP registration process, we also encountered several instances of failure.

TABLE 4 Leaf width measurement results of NeRFacto, Instant-NGP, Neus and Ours.

Method	R^2	MRE (%)	MAE (cm)
Instant-NGP	0.947	4.20	0.2357
Neus	0.990	2.15	0.1208
NeRFacto	0.961	2.84	0.1593
Ours	0.988	2.14	0.1199

Objective metrics evaluate the performance of reconstruction algorithms by considering the number of points, the Spatial resolution, and time efficiency. Figure 14 and Table 5 show the result of the three methods' average performance on the same metrics. In Figure 15, we intuitively show the comparison of the reconstruction results of the three methods. The reduction in time consumption achieved by our method, presented in Figure 14a, is substantial. Specifically, our method reduces time consumption by 96.08% relative to Colmap, and 49.73% relative to Kinect-based These findings underscore the high time efficiency of the algorithm proposed in this paper. As delineated in Figure 14b, The model's points generated by our method outperform Colmap by 159.26% and Kinect-based by 393.72%. As shown in Table 5, the spatial resolution of the model generated by our method is 0.19 mm, which is much smaller than that of Colmap and Kinect-based methods and has stronger geometric detail representation capabilities. As shown in Table 6, we also present the comparison of our tree seedling height measurement results with other literature.

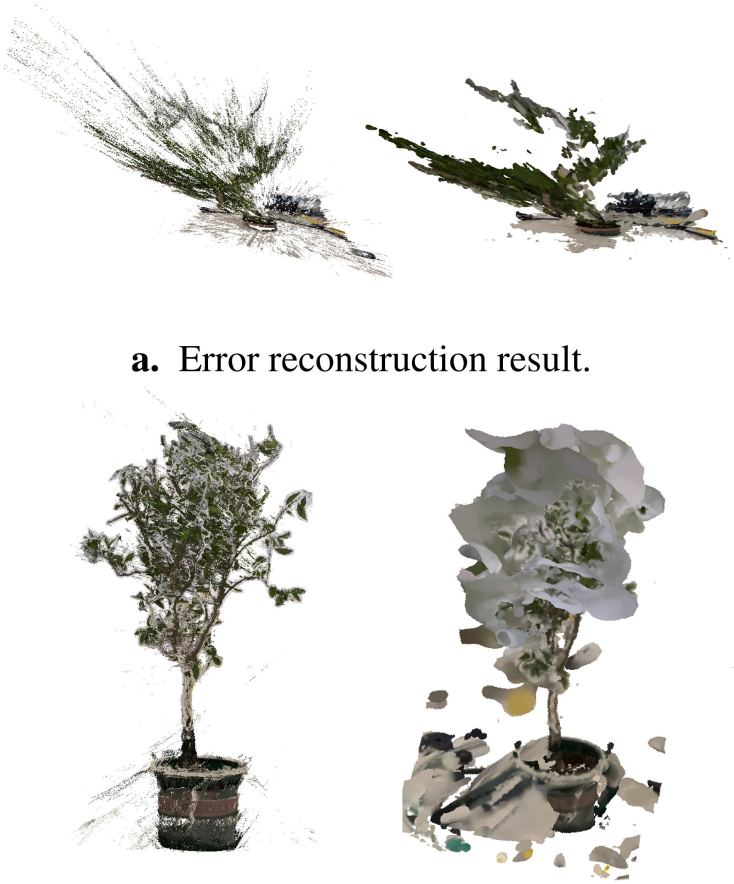


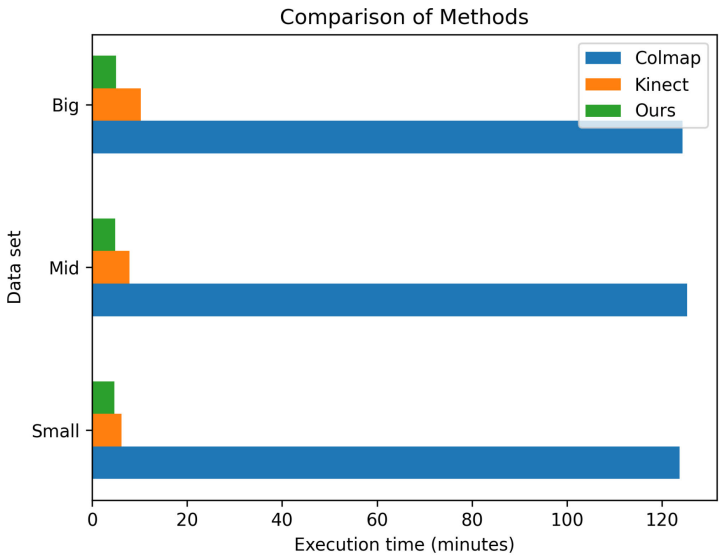
FIGURE 12
The reconstruction result of COLMAP. **(a)** Error reconstruction result. **(b)** True reconstruction result.

5 Conclusions

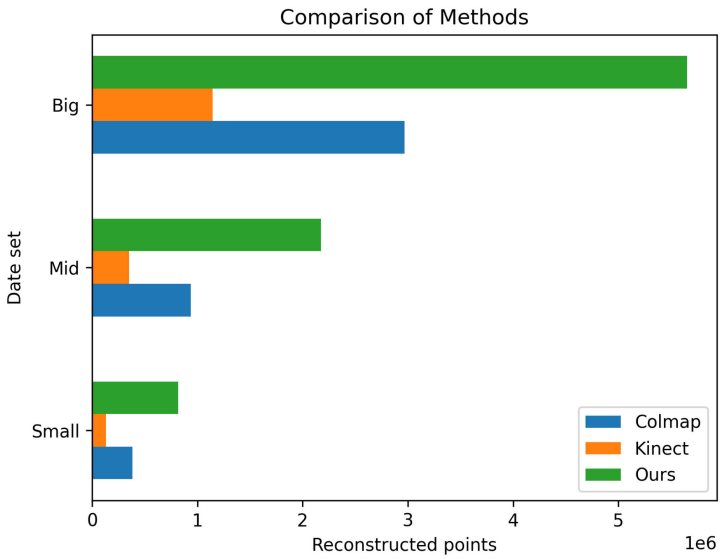
A 3D reconstruction platform based on multi-view images is proposed to reconstruct complex plants, comprising three steps: multi-view images acquisition, estimation, and calibration of camera parameters, and 3D reconstruction using OB-Nerf. In our study, to address the issue of model deformation during the reconstruction process, camera poses are automatically calibrated using the imaging trajectories as priori information, eliminating the need for additional calibrators. Furthermore, we propose OB-NeRF, an innovative Nerf-based algorithm for 3D reconstruction. The algorithm incorporates an optimizer of camera poses and an exposure adjustment mechanism. a new ray sampling strategy is introduced. It employs multi-resolution hash encoding techniques



FIGURE 13
The reconstruction result of Kinect.



a. Time efficiency of different methods.



b. The number of points by different methods.

FIGURE 14
Comparison of objective metrics for different methods on different data sets. **(a)** Time efficiency of different methods. **(b)** The number of points by different methods.

in conjunction with a shallow MLP network. Furthermore, the algorithm features an automated strategy for Mesh model extraction. OB-NeRF can reconstruct high-quality neural radiation fields of complex plants from images acquired under complex backgrounds and uneven illumination within 30 seconds. Subsequently, it accomplishes the synthesis of images and the automatic extraction of Mesh. From the perspectives of novel

TABLE 5 Spatial resolution of three methods.

Method	Spatial resolution
Colmap	0.0100mm
Kinect	0.0037mm
Ours	0.0019mm

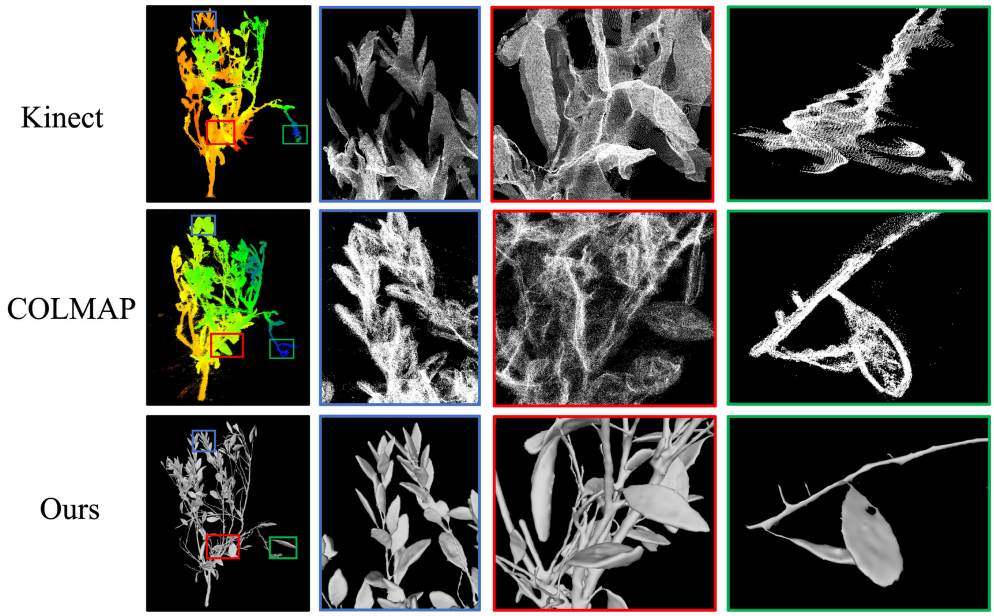


FIGURE 15
Comparison of the reconstruction results for Kinect, COLMAP, and Ours.

TABLE 6 H measurement results of our platform and other literature.

Plant	R^2	MRE(%)	MAE(cm)
Tree seedlings(Ours)	0.993	2.03	2.09
Seedlings (Yang et al., 2022)	0.991	2.79	2.72
lower plants (Wu et al., 2022)	0.991	–	–
Fruit tree (Yang et al., 2019)	0.960	2.50	–
Tree (Li and Tang, 2017)	0.982	–	–

viewpoint image synthesis and mesh modeling, our results have exhibited outstanding texture quality and geometric fidelity at the levels of both individual plants and their respective organs. The average PSNR of the synthesized images is 29.97dB, and the spatial resolution of the Mesh model is 0.19 mm. By comparing the three phenotypic parameters estimated from the model—tree height, leaf length, and leaf width—with manually measured values, the Mean Square Errors obtained were 2.0947 cm,0.1898 cm, and 0.1199 cm, respectively. The coefficients of determination were 0.9933,0.9881 and 0.9883, leading to a robust linear relationship between the extracted phenotype and measured traits.

In summary, the proposed low-cost reconstruction platform is capable of completing data acquisition for an individual plant in approximately 15 seconds and can perform high-quality reconstruction of the collected data within 250 seconds. The developed platform lays a solid foundation for the application of high-throughput phenotyping and digital twins in agriculture. It shows great potential in accelerating plant breeding, enabling precise crop management, and facilitating plant growth monitoring.

In the future, we plan to extend our work in the following directions:

1. Deploying the 3D reconstruction process to the cloud and achieving communication with the image acquisition system. This will improve the scalability, accessibility, and processing speed of our approach, making it more suitable for large-scale applications.
2. Leveraging the concept of collaborative design of software and hardware to optimize the 3D reconstruction pipeline further. By employing precise mechanical design and motor control, we aim to rotate and translate the camera under known poses accurately. This hardware-based approach will enable us to quickly obtain camera poses, replacing the need for the SFM method.
3. Our method is primarily designed for the 3D reconstruction of individual plants. However, when extended to large-scale field scenarios, experimental results revealed its performance limitations. To address this issue, future work could consider introducing new optimization strategies or adopting more advanced 3D reconstruction algorithms, such as 3DGS.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SW: Conceptualization, Data curation, Methodology, Visualization, Writing – original draft. CH: Data curation, Investigation, Software, Writing – original draft. BT: Data curation, Writing – original draft. YH: Funding acquisition,

Resources, Supervision, Writing – review & editing. SY: Formal Analysis, Project administration, Supervision, Writing – review & editing. SL: Funding acquisition, Supervision, Writing – review & editing. SX: Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was funded by the National Key Research and Development Program of China (grant number 2019YFD1001900), the Fundamental Research Funds for the Central Universities (2662022YLYJ010)Shannan City Local Science and Technology Plan Project(SNSBJKJJHXM2023004).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., and Srinivasan, P. P. (2021). Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. *Proceedings of the IEEE/CVF international conference on computer vision*, 5855–5864.

Cui, Z., Gu, L., Sun, X., Ma, X., Qiao, Y., and Harada, T. (2024). Aleth-nerf: Illumination adaptive nerf with concealing field assumption. *Proceedings of the AAAI Conference on Artificial Intelligence*. 38, 1435–1444.

Das Choudhury, S., Maturu, S., Samal, A., Stoerger, V., and Awada, T. (2020). Leveraging image analysis to compute 3d plant phenotypes based on voxel-grid plant reconstruction. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.521431

Feng, J., Saadati, M., Jubery, T., Jignasu, A., Balu, A., Li, Y., et al. (2023). 3d reconstruction of plants using probabilistic voxel carving. *Comput. Electron. Agric.* 213, 108248. doi: 10.1016/j.compag.2023.108248

Fu, L., Gao, F., Wu, J., Li, R., Karkee, M., and Zhang, Q. (2020). Application of consumer rgb-d cameras for fruit detection and localization in field: A critical review. *Comput. Electron. Agric.* 177, 105687. doi: 10.1016/j.compag.2020.105687

Gao, T., Zhu, F., Paul, P., Sandhu, J., Doku, H. A., Sun, J., et al. (2021). Novel 3D imaging systems for high-throughput phenotyping of plants. *Remote Sens.* 13 (11), 2113. doi: 10.3390/rs13112113

Golbach, F., Kootstra, G., Damjanovic, S., Otten, G., and van de Zedde, R. (2015). Validation of plant part measurements using a 3d reconstruction method suitabl for high-throughput seedling phenotyping. *Mach. Vision Appl.* 27, 663–680. doi: 10.1007/s00138-015-0727-5

Hu, K., Ying, W., Pan, Y., Kang, H., and Chen, C. (2024). High-fidelity 3d reconstruction of plants using neural radiance fields. *Comput. Electron. Agric.* 220, 108848. doi: 10.1016/j.compag

Kim, S., and Heo, S. (2024). An agricultural digital twin for mandarins demonstrates the potential for individualized agriculture. *Nat. Commun.* 15, 1561. doi: 10.1038/s41467-024-45725-x

Li, J., and Tang, L. (2017). Developing a low-cost 3d plant morphological traits characterization system. *Comput. Electron. Agric.* 143, 1–13. doi: 10.1016/j.compag.2017.09.025

Li, Y., Liu, J., Zhang, B., Wang, Y., Yao, J., Zhang, X., et al. (2022). Three-dimensional reconstruction and phenotype measurement of maize seedlings based on multi-view image sequences. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.974339

Lorensen, W. E., and Cline, H. E. (1987). "Marching cubes: A high resolution 3d surface construction algorithm," in *Proceedings of the 14th annual conference on computer graphics and interactive techniques*, vol. 87. (Association for Computing Machinery, New York, NY, USA), 163–169. doi: 10.1145/37401.37422

Lu, X., Ono, E., Lu, S., Zhang, Y., Teng, P., Aono, M., et al. (2020). Reconstruction method and optimum range of camera-shooting angle for 3d plant modeling using a multi-camera photography system. *Plant Methods* 16, 118. doi: 10.1186/s13007-020-00658-6

Luo, S., Liu, W., Zhang, Y., Wang, C., Xi, X., Nie, S., et al. (2021). Maize and soybean heights estimation from unmanned aerial vehicle (uav) lidar data. *Comput. Electron. Agric.* 182, 106005. doi: 10.1016/j.compag.2021.106005

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*. (New York, NY, USA: ACM) 55, 99–106.

Müller, T., Evans, A., Schied, C., and Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graphics* 41, 1–15. doi: 10.1145/3528223.3530127

Peladarinos, N., Piromalis, D., Cheimaras, V., Tserepas, E., Munteanu, R. A., and Papageorgas, P. (2023). Enhancing smart agriculture by implementing digital twins: A comprehensive review. *Sensors (Basel)* 23, 7128. doi: 10.3390/s23167128

Pylianidis, C., Osinga, S., and Athanasiadis, I. N. (2021). Introducing digital twins to agriculture. *Comput. Electron. Agric.* 184, 105942. doi: 10.1016/j.compag.2020

Scharr, H., Briese, C., Embgenbroich, P., Fischbach, A., Fiorani, F., and Muller-Linow, M. (2017). Fast high resolution volume carving for 3d plant shoot reconstruction. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01680

Schönberger, J. L., and Frahm, J.-M. (2016). "Structure-from-motion revisited," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Piscataway, NJ: IEEE), 4104–4113. doi: 10.1109/CVPR.2016.445

Seitz, S., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 1. (Piscataway, NJ: IEEE), 519–528. doi: 10.1109/CVPR.2006.19

Sishodia, R. P., Ray, R. L., and Singh, S. K. (2020). Applications of remote sensing in precision agriculture: A review. *Remote Sens.* 12, 3136. doi: 10.3390/rs12193136

Song, P., Li, Z., Yang, M., Shao, Y., Pu, Z., Yang, W., et al. (2023). Dynamic detection of threedimensional crop phenotypes based on a consumer-grade rgb-d camera. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1097725

Sun, Y., Shang, L., Zhu, Q. H., Fan, L., and Guo, L. (2022). Twenty years of plant genome sequencing: achievements and challenges. *Trends Plant Sci.* 27, 391–401. doi: 10.1016/j.tplants.2021.10.006

Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Wang, T., et al. (2023). "Nerfstudio: A modular framework for neural radiance field development," in *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Proceedings (ACM), SIGGRAPH '23*. (New York, NY, USA: Association for Computing Machinery), 1–12. doi: 10.1145/3588432.3591516

Teng, X., Zhou, G., Wu, Y., Huang, C., Dong, W., and Xu, S. (2021a). Three-dimensional reconstruction method of rapeseed plants in the whole growth period using rgb-d camera. *Sensors* 21, 4628. doi: 10.3390/s21144628

Teng, X., Zhou, G., Wu, Y., Huang, C., Dong, W., and Xu, S. (2021b). Three-dimensional reconstruction method of rapeseed plants in the whole growth period using rgb-d camera. *Sensors* 21. doi: 10.3390/s21144628

Tross, M. C., Gaillard, M., Zwiener, M., Miao, C., Grove, R. J., Li, B., et al. (2021). 3d reconstruction identifies loci linked to variation in angle of individual sorghum leaves. *PeerJ* 9, e12628. doi: 10.7717/peerj.12628

Tsoulias, N., Saha, K. K., and Zude-Sasse, M. (2023). *In-situ* fruit analysis by means of lidar 3d point cloud of normalized difference vegetation index (ndvi). *Comput. Electron. Agric.* 205, 107611. doi: 10.1016/j.compag.2022.107611

Walter, J. D. C., Edwards, J., McDonald, G., and Kuchel, H. (2019). Estimating biomass and canopy height with lidar for field crop breeding. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01145

Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., and Wang, W. (2023a). *Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction*.

Wang, Y., Han, Q., Habermann, M., Daniilidis, K., Theobalt, C., and Liu, L. (2023b). Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction.

Proceedings of the IEEE/CVF International Conference on Computer Vision 2023, 3295–3306.

Wu, S., Wen, W., Gou, W., Lu, X., Zhang, W., Zheng, C., et al. (2022). A miniaturized phenotyping platform for individual plants using multi-view stereo 3d reconstruction. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.897746

Xiao, Q., Bai, X., Zhang, C., and He, Y. (2022). Advanced high-throughput plant phenotyping techniques for genome-wide association studies: A review. *J. Adv. Res.* 35, 215–230. doi: 10.1016/j.jare.2021.05.002

Yang, H., Wang, X., and Sun, G. (2019). Three-dimensional morphological measurement method for a fruit tree canopy based on kinect sensor self-calibration. *Agronomy* 9, 741. doi: 10.3390/agronomy9110741

Yang, T., Ye, J., Zhou, S., Xu, A., and Yin, J. (2022). 3d reconstruction method for tree seedlings based on point cloud self-registration. *Comput. Electron. Agric.* 200, 107210. doi: 10.1016/j.compag.2022.107210

Zhang, Z. (1999). "Flexible camera calibration by viewing a plane from unknown orientations," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Vol. 1. (Piscataway, NJ: IEEE), 666–673. doi: 10.1109/ICCV.1999.791289

Zhang, J., Wang, X., Ni, X., Dong, F., Tang, L., Sun, J., et al. (2024). Neural radiance fields for multiscale constraint-free 3d reconstruction and rendering in orchard scenes. *Comput. Electron. Agric.* 217, 108629. doi: 10.1016/j.compag.2024.108629

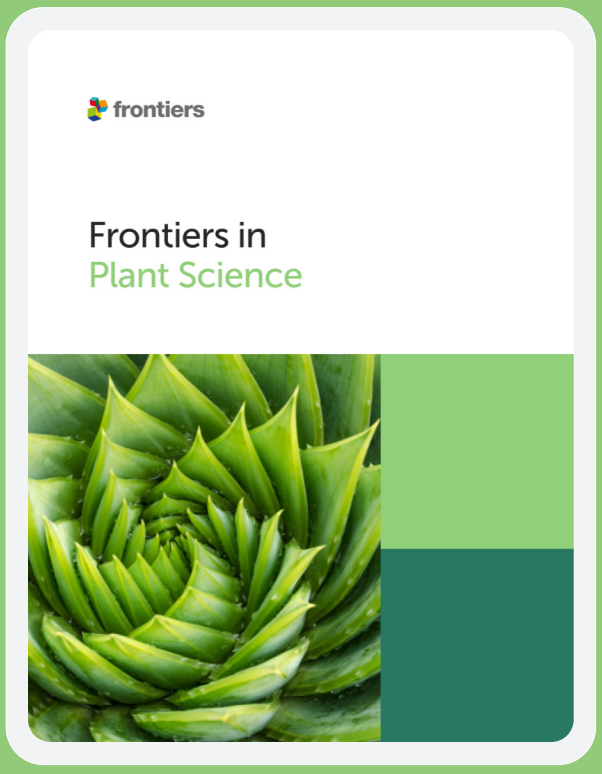
Frontiers in
Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact