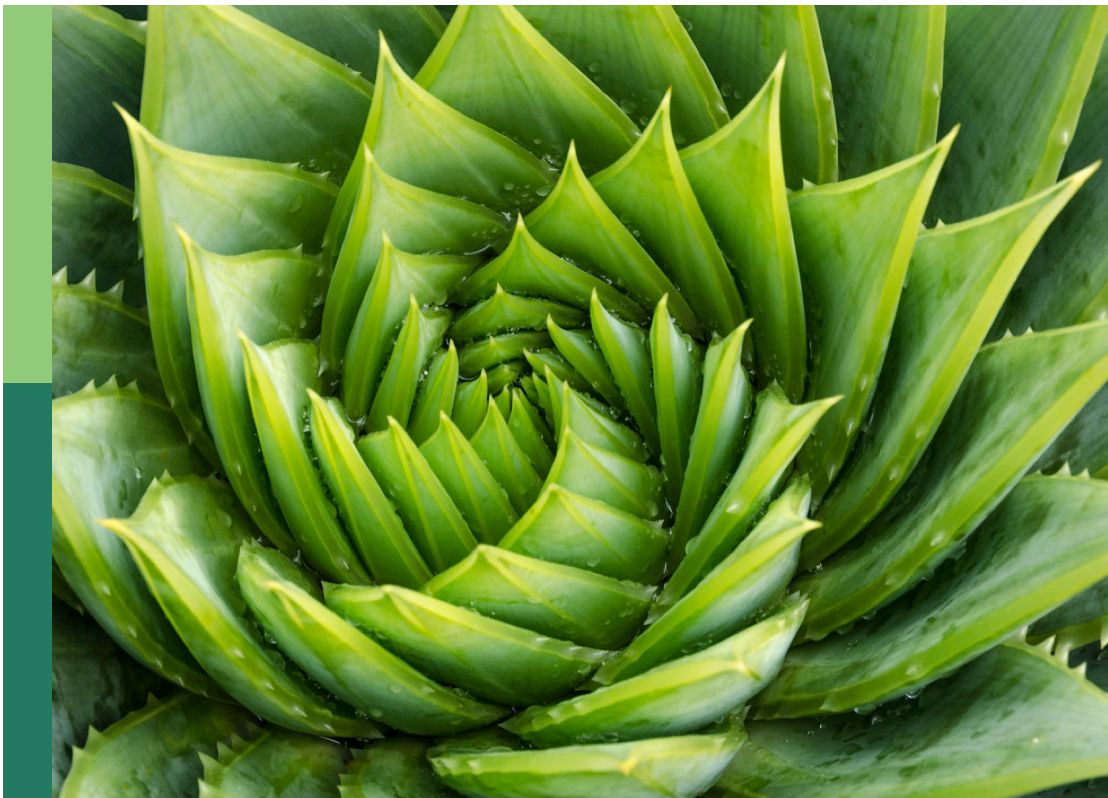
frontiers | Research Topics

Remote sensing for field-based crop phenotyping

Edited by
Jiangang Liu, Zhenjiang Zhou
and Bo Li

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 978-2-8325-4430-3
DOI 10.3389/978-2-8325-4430-3

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

Remote sensing for field-based crop phenotyping

Topic editors

Jiangang Liu — State Key Laboratory of Vegetable Biobreeding, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, China

Zhenjiang Zhou — Zhejiang University, China

Bo Li — Syngenta UK Limited, United Kingdom

Citation

Liu, J., Zhou, Z., Li, B., eds. (2024). *Remote sensing for field-based crop phenotyping*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4430-3

Table of contents

05 Editorial: Remote sensing for field-based crop phenotyping
Jiangang Liu, Zhenjiang Zhou and Bo Li

08 Remote sensing for field pea yield estimation: A study of multi-scale data fusion approaches in phenomics
Afef Marzougui, Rebecca J. McGee, Stephen Van Vleet and Sindhuja Sankaran

26 Significant increases in Donghong kiwifruit yield by a novel umbrella-shaped trellis system and identification of associated molecular mechanisms
Honghong Deng, Yao Li, Changqing Pang, Kun Zhang, Xinbo Tian, Tong Wang, Yan Liang, Zunzhen He, Yuxuan Lang, Jinbao Fang, Lijin Lin, Jin Wang, Xiulan Lv, Hui Xia and Dong Liang

38 Non-destructive monitoring of maize LAI by fusing UAV spectral and textural features
Xinkai Sun, Zhongyu Yang, Pengyan Su, Kunxi Wei, Zhigang Wang, Chenbo Yang, Chao Wang, Mingxing Qin, Lujie Xiao, Wude Yang, Meijun Zhang, Xiaoyan Song and Meichen Feng

49 Row selection in remote sensing from four-row plots of maize and sorghum based on repeatability and predictive modeling
Seth A. Tolley, Neal Carpenter, Melba M. Crawford, Edward J. Delp, Ayman Habib and Mitchell R. Tuinstra

61 A fast phenotype approach of 3D point clouds of Pinus massoniana seedlings
Honghao Zhou, Yang Zhou, Wei Long, Bin Wang, Zhichun Zhou and Yue Chen

77 Phenological stage and vegetation index for predicting corn yield under rainfed environments
Amrit Shrestha, Raju Bheemanahalli, Ardeshir Adeli, Sathishkumar Samiappan, Joby M. Prince Czarnecki, Cary Daniel McCraine, K. Raja Reddy and Robert Moorhead

89 Methodological evolution of potato yield prediction: a comprehensive review
Yongxin Lin, Shuang Li, Shaoguang Duan, Yanran Ye, Bo Li, Guangcun Li, Dianqiu Lyv, Liping Jin, Chunsong Bian and Jiangang Liu

114 Using crop intercepted solar radiation and vegetation index to estimate dry matter yield of Choy Sum
Yiyin He, Zhao Wang, Sashuang Sun, Lijun Zhu, Yu Li, Xiaoxiao Wang, Jiang Shi, Si Chen, Dunchang Qi, Junxiang Peng and Zhenjiang Zhou

122 Remote sensing continuity: a comparison of HTP platforms and potential challenges with field applications
Andrew W. Herr and Arron H. Carter

133 Estimating yield-contributing physiological parameters of cotton using UAV-based imagery
Amrit Pokhrel, Simerjeet Virk, John L. Snider, George Vellidis, Lavesta C. Hand, Henry Y. Sintim, Ved Parkash, Devendra P. Chalise, Joshua M. Lee and Coleman Byers

155 LAI estimation based on physical model combining airborne LiDAR waveform and Sentinel-2 imagery
Zixi Shi, Shuo Shi, Wei Gong, Lu Xu, Binhui Wang, Jia Sun, Bowen Chen and Qian Xu

171 Field phenotyping for African crops: overview and perspectives
Daniel K. Cudjoe, Nicolas Virlet, March Castle, Andrew B. Riche, Manal Mhada, Toby W. Waine, Fady Mohareb and Malcolm J. Hawkesford

194 Hyperspectral estimation of chlorophyll content in jujube leaves: integration of derivative processing techniques and dimensionality reduction algorithms
Nigela Tuerxun, Jianghua Zheng, Renjun Wang, Lei Wang and Liang Liu

205 Optimizing window size and directional parameters of GLCM texture features for estimating rice AGB based on UAVs multispectral imagery
Jikai Liu, Yongji Zhu, Lijuan Song, Xiangxiang Su, Jun Li, Jing Zheng, Xueqing Zhu, Lantian Ren, Wenhui Wang and Xinwei Li

224 Combining spectral and texture feature of UAV image with plant height to improve LAI estimation of winter wheat at jointing stage
Mengxi Zou, Yu Liu, Maodong Fu, Cunjun Li, Zixiang Zhou, Haoran Meng, Enguang Xing and Yanmin Ren

243 Estimating the frost damage index in lettuce using UAV-based RGB and multispectral images
Yiwen Liu, Songtao Ban, Shiwei Wei, Linyi Li, Minglu Tian, Dong Hu, Weizhen Liu and Tao Yuan

261 Enhancing estimation of cover crop biomass using field-based high-throughput phenotyping and machine learning models
Geng Bai, Katja Koehler-Cole, David Scoby, Vesh R. Thapa, Andrea Basche and Yufeng Ge

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 26 January 2024
DOI 10.3389/fpls.2024.1368694

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Peng Chen,
Anhui University, China

*CORRESPONDENCE
Bo Li
bo.li-1@syngenta.com

RECEIVED 11 January 2024
ACCEPTED 16 January 2024
PUBLISHED 26 January 2024

CITATION
Liu J, Zhou Z and Li B (2024)
Editorial: Remote sensing for field-based crop phenotyping.
Front. Plant Sci. 15:1368694.
doi: 10.3389/fpls.2024.1368694

COPYRIGHT
© 2024 Liu, Zhou 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.

Editorial: Remote sensing for field-based crop phenotyping

Jiangang Liu[1], Zhenjiang Zhou[2] and Bo Li[3]*

[1]State Key Laboratory of Vegetable Biobreeding, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China, [2]College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, Zhejiang, China, [3]Syngenta, Jealott's Hill International Research Centre, Bracknell, United Kingdom

KEYWORDS

smart agriculture, remote sensing, crop phenotyping, computer vision, machine learning, data fusion

Editorial on the Research Topic
Remote sensing for field-based crop phenotyping

With the population predicted to increase to over 9.6 billion by 2050, and food demand anticipated to increase by between 60 and 100%, sustainable and resilient agricultural production with a minimised impact on the environment is crucial particularly at the context of global climate change. Breeding and identifying crop varieties with high production and adopted specific environmental conditions requires considerable efforts to assess crop phenotypic traits (e.g., LAI, plant height, biomass, yield et al.), which contribute to the stable increased productivity and efficient use of resources. Traditional methods for determining crop phenotypic traits are mainly based on destructive field sampling, hand-held instrument measurement, which are characterized as time-consuming, limited representative. Remote sensing provides a novel solution to quantify crop structural and functional traits in a timely, rapid, non-invasive, and efficient manner for field crops. With the development of sensors and diversified algorithms, a range of crop phenotypic traits have been determined in a manner of high-throughput, including morphological parameters, spectral and textural characteristics, physiological traits, and responses to abiotic/biotic stresses under different environments.

This Research Topic presents 15 research articles and 2 reviews with an insight into the recent advances in crop phenotyping using remote sensing to address some high-priority challenges including comparing ground-based handheld and remote aerial system, plant biophysical parameters estimation with innovative traits, imagery data fusion and investigating the impact of experimental design on the performance of traits extraction.

The current state of art high-throughput phenotyping platform including ground-based and aerial platforms. The integrated imaging sensors such as visible light, 3D, hyperspectral and fluorescence sensors were reviewed by Cudjoe et al., and pointed out that the lack of appropriate field phenotyping infrastructures is impeding the development of new crop cultivars with improved traits and will eventually have a negative impact on the agricultural sector and African food security. Due to the operational complexity and limited funding, the deployment of efficient and low-cost high-throughput phenotyping methods are still highly demanded in Africa. In addition to ground-based and aerial-based remote sensing, Lin et al. also described satellite-based remote sensing in potato yield prediction. Furthermore, strategies for potato yield prediction including remote sensing, crop growth

models (CGM) and yield limiting factors were discussed in depth, and the application of different CGMs were analysed. As data from solely single sensor source often limits the performance of prediction model, Lin et al. proposed that multi-source data fusion and time-series data have enormous potential for future potato yield prediction. Marzougui et al. investigated two data fusion methods for unmanned aerial systems (UAS) multispectral imagery and high-resolution satellite imagery in field pea yield prediction, showing the improved model performance by fusing multiple time point and multiple sensor source information. With using only UAS, Liu et al. implemented feature fusion between texture features extracted from RGB imagery and spectral features from multispectral imagery for estimating the frost damage index in lettuce. Bai et al. applied more comprehensive sensor data fusion in an existing Field-based High-Throughput Plant Phenotyping (FHTPP) system and combined morphological, spectral, thermal, and environmental features as the inputs of the machine learning models to estimate the cover crop biomass. Shi et al. recognised the limitation of using single-source remote sensing spectral or LiDAR waveform data for Leaf Area Index (LAI) estimation and presented another example showing the advantage of data fusion. Rather than using empirical model that always requires sufficient amount of field measurements and does not provide a good explanation of the fusion mechanism, physical model geometric-optical and radiative transfer (GORT) integrating both spectral imagery and LiDAR waveform showed an enhancement in LAI estimation comparing with using spectral or LiDAR data alone. As a critical physiological and biochemical parameter indicating crop growth status and yield potential, LAI estimation was also investigated for winter wheat and maize in another two studies by Zou et al. and Sun et al. Although both studies use single multispectral imaging sensor equipped on UAS, both spectral and texture features were extracted and fused as the input variables of machine learning model. Multiple multivariate statistical regression models were constructed and compared, and all presented high performances. All the texture features mentioned above are based on the grey level co-occurrence matrix (GLCM) method, however, window size and direction texture parameters are highly sensitive to texture metrics and default parameter values are always applied in previous studies. As a result, Liu et al. conducted a detailed study to understand the optimum window size and directional parameters at different growing stages of rice. As most of the prediction models were developed for a specific growth stage in previous studies, Pokhrel et al. integrated growing degree days (GDD) with VIs extracted from UAV-based multispectral imagery for estimating intercepted photosynthetically active radiation (IPARf), radiation use efficiency (RUE) and harvest index (HI), which were three yield-contributing physiological parameters of cotton. The incorporation of GDD allowed the prediction to be reliably made throughout the whole growing season.

Many VIs were developed in the past decades showing correlation with crop physiological parameters, however, none of the previous studies investigated VIs that can be applied to predict the corn yield throughout the season. Shrestha et al. applied both correlation analysis and random forest model to identify the VIs with the most consistency and highest predictive power for corn yield prediction.

Crop traits evaluation on a row segment basis within plots is common in field phenotyping. Tolley et al. conducted UAS flights and extracted crop traits using RGB, LiDAR and VNIR (visible and near infrared) sensors and concluded that significant difference was observed between different row selections, and with large plot size, excluding outer rows could lead to more robust model performance. It is believed that this study can support long-standing principles of experiment design in both agronomy and crop breeding with remote sensing.

Apart from aerial platform, ground-based system is also popular in crop phenotyping due to the high resolution imaging across a wide range of wavelengths. Tuerxun et al. used a portable spectroradiometer to obtain the hyperspectral data, and multivariate regression model was developed for chlorophyll content estimation in jujube leaves after feature reduction. He et al. applied a handheld spectrometer to generate VIs and a light quantum sensor to measure photosynthetic active radiation, which showed fairly good relationships with the dry mater of Choy Sum. Although the same VIs can be derived from both ground-based and aerial-based systems, it was interesting that Herr and Carter found poor correlation across remote sensing platforms, which suggested that data collected from different systems should not be used interchangeable.

Crop phenotyping not only can assess the crop growth status, but also support the decision of the crop management strategy. Deng et al. collected phenotypic data to support the conclusion that the umbrella-shaped trellis system largely improved the production of Donghong Kiwifruit while maintaining the fruit quality.

Zhou et al. presented the only research of ground-based 3D model reconstruction and analysis in this topic. New algorithm was proposed for skeleton point search, which facilitated the stem and leaf segmentation, further leading to the five phenotypic parameters estimation including plant height, stem diameter, main stem length, regional leaf length and leaf number.

The studies published in this Research Topic present novel computer vision algorithms and provide new knowledge in the broad field of crop phenotyping. It is obvious that UAS integrating with imaging sensors has shown great advantages in crop phenotyping. Future studies should investigate more on multi-sensor data fusion and multiple time point data fusion for more reliable crop traits estimation under various environmental conditions. As data collected from different remote sensing systems cannot be used interchangeable, standardisation of remote sensing pipeline for crop traits estimation needs to be investigated in the future to provide re-usable data and reduced operational cost. We hope the studies presented in this Research Topic will help consolidate the integration between remote sensing and crop phenotyping, resulting in more reliable and affordable phenotyping tools under dynamic environments.

Author contributions

JL: Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing. ZZ: Writing – review & editing.

BL: Conceptualization, Resources, 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 is supported by National Key R&D Program of China (2023YFD 2302100), National Natural Science Foundation of China (32372232), Key scientific and technological projects of Heilongjiang province in China (2021ZXJ05A05-03) awarded to Jiangang Liu, also Science and Technology Department of Ningxia - The Key Research and Development Program (2023BCF01017).

Conflict of interest

Author BL was employed by company Syngenta.

The remaining 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.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 March 2023
DOI 10.3389/fpls.2023.1111575

Check for updates

OPEN ACCESS

EDITED BY
Zhenjiang Zhou,
Zhejiang University, China

REVIEWED BY
Haikuan Feng,
Beijing Research Center for Information Technology in Agriculture, China
Junxiang Peng,
Swedish University of Agricultural Sciences, Sweden
Yaohua Hu,
Zhejiang Agriculture and Forestry University, China

*CORRESPONDENCE
Sindhuja Sankaran
sindhuja.sankaran@wsu.edu

SPECIALTY SECTION
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

RECEIVED 29 November 2022
ACCEPTED 06 February 2023
PUBLISHED 03 March 2023

CITATION
Marzougui A, McGee RJ, Van Vleet S and Sankaran S (2023) Remote sensing for field pea yield estimation: A study of multi-scale data fusion approaches in phenomics. *Front. Plant Sci.* 14:1111575. doi: 10.3389/fpls.2023.1111575

COPYRIGHT
© 2023 Marzougui, McGee, Van Vleet and Sankaran. 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.

Remote sensing for field pea yield estimation: A study of multi-scale data fusion approaches in phenomics

Afef Marzougui[1], Rebecca J. McGee[2], Stephen Van Vleet[3] and Sindhuja Sankaran[1]*

[1]Department of Biological Systems Engineering, Washington State University, Pullman, WA, United States, [2]United States Department of Agriculture-Agricultural Research Service (USDA-ARS), Grain Legume Genetics and Physiology Research Unit, Washington State University, Pullman, WA, United States, [3]Agriculture and Natural Resources, Washington State University Extension, Colfax, WA, United States

Introduction: Remote sensing using unmanned aerial systems (UAS) are prevalent for phenomics and precision agricultural applications. The high-resolution data for these applications can provide useful spectral characteristics of crops associated with performance traits such as seed yield. With the recent availability of high-resolution satellite imagery, there has been growing interest in using this technology for plot-scale remote sensing applications, particularly those related to breeding programs. This study compared the features extracted from high-resolution satellite and UAS multispectral imagery (visible and near-infrared) to predict the seed yield from two diverse plot-scale field pea yield trials (advanced breeding and variety testing) using the random forest model.

Methods: The multi-modal (spectral and textural features) and multi-scale (satellite and UAS) data fusion approaches were evaluated to improve seed yield prediction accuracy across trials and time points. These approaches included both image fusion, such as pan-sharpening of satellite imagery with UAS imagery using intensity-hue-saturation transformation and additive wavelet luminance proportional approaches, and feature fusion, which involved integrating extracted spectral features. In addition, we also compared the image fusion approach to high-definition satellite data with a resolution of 0.15 m/pixel. The effectiveness of each approach was evaluated with data at both individual and combined time points.

Results and discussion: The major findings can be summarized as follows: (1) the inclusion of the texture features did not improve the model performance, (2) the performance of the model using spectral features from satellite imagery at its original resolution can provide similar results as UAS imagery, with variation depending on the field pea yield trial under study and the growth stage, (3) the model performance improved after applying multi-scale, multiple time point

feature fusion, (4) the features extracted from the pan-sharpened satellite imagery using intensity-hue-saturation transformation (image fusion) showed better model performance than those with original satellite imagery or high definition imagery, and (5) the green normalized difference vegetation index and transformed triangular vegetation index were identified as key features contributing to high model performance across trials and time points. These findings demonstrate the potential of high-resolution satellite imagery and data fusion approaches for plot-scale phenomics applications.

KEYWORDS

high-resolution satellite, unmanned aerial system, multispectral, yield prediction, pan-sharpening, high-throughput field phenotyping, plant breeding

1 Introduction

Crop improvement efforts focus on developing new cultivars with increased yield potential, stable agronomic traits, and better environmental adaptability. Genomic tools and technologies have played a major role in advancing plant breeding programs by enabling "accurate" quantitative trait loci mapping and genome-wide association studies (Varshney et al., 2021). However, locating associated genes depends on the accuracy of the phenotypic data (Pratap et al., 2019). The acquisition of phenotypic data needs accurate, rapid, and efficient tools to bridge the relationship between the genotype and phenotype and the interactions of the genotype with the environment and management practices. Phenomics or high-throughput plant phenotyping technologies have enabled accurate and rapid phenotyping at different scales and resolutions (Yang et al., 2020).

The rise of field phenomics is driven by the advances in sensing technologies that can be deployed using drones or unmanned aerial systems (UAS) (Sankaran et al., 2015; Pieruschka and Schurr, 2019; Zhao et al., 2019a; Yang et al., 2020; Guo et al., 2021). Such systems have been used for various phenomics applications – from spotting ideotypes (Roth et al., 2022) to studying latent heat flux (Tauro et al., 2022) in crops. The success of UAS-based sensing techniques for measuring crop phenotypes in breeding trials is contributed to the following: (i) low altitude flights enable acquisition of high spatial resolution data (Zhang et al., 2020; Jin et al., 2021), ideal for imaging the small-size plots; (ii) UAS equipment and cameras are small, light, and easily portable (Zhao et al., 2019a; Zhang et al., 2020); (iii) the flexibility of UAS in selecting the date and time for data acquisition (Zhang et al., 2020) or collecting data at different times within a day (Valencia-Ortiz et al., 2021); and (iv) the ability of the system to be integrated with single or multiple on-board sensors such as RGB, multispectral, hyperspectral, thermal, and/or light detection and ranging (LiDAR) sensing systems, which facilitates the measurements of a wide range of crop traits (Yang et al., 2017; Zhao et al., 2019a; Jin et al., 2021).

One of the key performance traits in crop breeding programs is yield and its components. As these traits are complex and influenced by the environment, there is a continuous and ongoing effort to identify novel approaches, including remote sensing applications, to predict yield and yield-associated traits. These applications rely on retrieving image-based features that can be indirectly associated with yield. A large body of literature reports the accuracy of vegetation indices (VIs) in predicting yield and other important agronomic and stress-tolerant traits at breeding plot level for multiple crops such as wheat (Kyratzis et al., 2017; Hassan et al., 2018; Hassan et al., 2019; Li et al., 2019; Fu et al., 2020; Shafiee et al., 2021; Zeng et al., 2021), soybean (Zhang et al., 2019; Maimaitijiang et al., 2020a; Maimaitijiang et al., 2020b; Roth et al., 2022; Santana et al., 2022), maize (Buchaillot et al., 2019; Adak et al., 2021; Sankaran et al., 2021), and pulse crops (Sankaran et al., 2018; Marzougui et al., 2019; Vargas et al., 2019; Valencia-Ortiz et al., 2021; Zhang et al., 2021; Tefera et al., 2022). While yield prediction using UAS-based sensing approaches has shown promising results, scaling up the application to cover large areas and/or multi-environment trials is still a major challenge (Zhang et al., 2020; Jin et al., 2021; Smith et al., 2021). To determine yield stability, crop breeders need to evaluate the breeding plant materials at different geographical locations and assess genotype to environment and management (G x E x M) interactions. In addition, it is important to note that the breeding cycle may require 10 or more years from the first genetic cross to commercial release of new varieties (Tracy et al., 2020; Kholová et al., 2021). This factor strongly depends on the crop type and environments (Watt et al., 2020). For instance, in the United States Department of Agriculture Agricultural Research Services' (USDA-ARS) pea breeding program in the Pacific Northwest USA, the total number of locations that each breeding plant material is evaluated in a given season can range from 4-15 (Supplementary Materials Figure 1) depending on whether it is entered in local, state-wide, or regional trials within the USA. Thus, scaling up the UAS-based phenomics is not only limited by the spatial coverage, and depends on several factors such as personnel, equipment, and travel time. In addition, other limitations could result from the airspace regulations, especially if trials are within a short distance of an airport (Yang et al., 2020; Zhang et al., 2020), limited access to remote locations (Zhang et al., 2020), and finite battery capacity and flight time (Guo et al., 2021). Therefore, to address some of these limitations, in this research, our major focus is towards exploring the potential of low-orbiting high-resolution satellite imagery

as a tool and its capacity for multi-scale, multi-sensor (with UAS data) and multi-modal (spectral and texture) data integration to improve the scalability of field phenomics applications.

The availability of low-orbiting high-resolution satellite imagery offers a great opportunity to measure phenotypic traits at the breeding plot scale. Such application was not previously possible due to spatial and temporal resolution constraints. For example, in the Pacific Northwest region of the USA, the size of breeding plots in spring wheat, pea, and chickpea breeding programs is about 1.5 x 6 m (Sangjan et al., 2021; Zhang et al., 2021). Segmenting each plot with coarse resolution imagery data was not possible. Nevertheless, as described in Zhang et al. (2020), there is a potential to leverage the very high-resolution satellite imagery for field phenomics applications, particularly utilizing the available commercial imagery with a spatial resolution of 0.30-0.50 m/pixel. In addition, these sources have a high temporal resolution, and the image acquisition time points can be tasked to the desired time window within a short period (less than a week), as long as cloud cover is not a limitation. The applications are particularly promising for crop breeding programs in the semi-arid agricultural production areas due to the preponderance of cloud-free days.

Combining multi-modal, multi-scale, multi-sensor data from satellite and UAS sources can also offer a more robust solution, especially for agricultural monitoring. Alvarez-Vanhard et al. (2021) reviewed the different scenarios of satellite-UAS information synergies for agricultural and non-agricultural applications and proposed four strategies that will benefit from satellite-UAS integration. These strategies include data comparison, multi-scale explanation, model calibration, and data fusion. In terms of data comparison, Sankaran et al. (2021) demonstrated that satellite-based vegetation indices were significantly correlated with those from UAS imagery and seed yield in a maize breeding trial. Multi-scale explanation in agriculture may include anomaly detection using satellite data, with deeper and precise assessment of underlying anomaly using UAS imagery. In Selvaraj et al. (2020), banana plantations were identified using machine learning approaches applied to satellite imagery (Planet and WorldView2), while UAS data were used to identify the major diseases in the canopy within those satellite imageries identified banana plantations. Examples of model calibration are the integration of satellite (with mixed pixels) and UAS data (with unmixed pixels) to develop a spectral un-mixing approach for improving the mapping of different plant communities (Alvarez-Vanhard et al., 2020), and improving the classification accuracies (irrigated area mapping) of the satellite-based data using the labeled dataset acquired with the UAS imagery (Nhamo et al., 2018). For the data fusion, both image and feature-based fusion can be explored. For example, Malamiri et al. (2021) enhanced the accuracy in classifying six pistachio cultivars (with precise weed and soil separation) using pan-sharpened Landsat 8 imagery with a UAS RGB image, which increased the spatial resolution of satellite imagery to 20 cm/pixel. Similarly, Zhao et al. (2019b) fused UAS RGB images with Sentinel-2A to improve the accuracy of crop mapping with an enhanced spatial resolution of 10 cm/pixel. Maimaitijiang et al. (2020a) found that the combination of canopy structure features extracted from UAS RGB data and canopy spectral features extracted from satellite multispectral data improved the estimation of soybean above ground biomass, leaf area index, and leaf nitrogen content compared to just a single sensor approach. Even if such studies in the literature that investigate the integration of satellite and UAS data are increasing, the studies exploring such synergies in the context of field phenomics involving small plot applications, especially in field pea, with low-orbiting high-resolution satellite imagery are limited.

Therefore, to help bridge this gap, in this study, we evaluated multiple approaches involving the integration of UAS and high-resolution satellite imagery to predict seed yield of field pea entries in breeding trials. We hypothesize that: (i) the quality of satellite-based features will be comparable to UAS-based features (to determine the suitability of satellite imagery for phenotyping applications in small plot research), (ii) the feature fusion from both sensing approaches will capture the temporal change of crop growth, and (iii) the fusion of satellite and UAS imagery will improve satellite image spatial resolution and thus the accuracy of the extracted features. The aspects mentioned above were evaluated by developing machine learning models for seed yield prediction and assessing information gains utilizing data from multiple sensors (versus a single sensor), and single and multiple time points. We explored the integration of satellite and UAS vegetation metrics extracted from multispectral imagery to enhance the seed yield estimation at both feature and image levels.

2 Materials, data acquisition, and pre-processing

2.1 Study site and experimental design

This study evaluated two field pea-breeding trials in two consecutive years (Figure 1). The advanced breeding trial, which we denote as 'Site 1' hereafter in this text, refers to the USDA-ARS replicated advanced yield trial at the end of the crop breeding cycle, prior to releasing a new cultivar. This trial was evaluated in 2019 near Pullman, Washington, USA. The variety testing trial (termed 'Site 2') was evaluated in 2020 at Johnson, Washington, USA. This trial comprises plant materials from both private and public breeding programs, including some commercial varieties. Both trials were planted using a randomized complete block design with three and four replications for Site 1 and Site 2, respectively. The advanced yield trial at Site 1 contained three adjacent trials – one in which all entries had green seeds, one in which all entries had yellow seeds, and one in which the entries had either yellow or green seeds. There were 65 entries total. The Site 2 trial included 33 entries with a 24% overlap with the Site 1 trial (8 common entries). The entries had either green or yellow seeds. In this trial, all seeds were inoculated with *Rhizobium* bacteria before sowing. For both trials, seeds of each entry were planted in separate plots. Each plot was 6.1 m long and 1.5 m width and had 6 rows. There was 0.75 m between plots.

2.2 Ground reference data

The seed yield (kg/ha) was collected from each plot at physiological maturity in August 2019 and 2020. The seed yield for both trials (Site 1 and Site 2) had a normal distribution (Supplementary Materials Figure 2). The summary statistics of seed yield collected across the two years are presented in Table 1. Plots without yield data points and

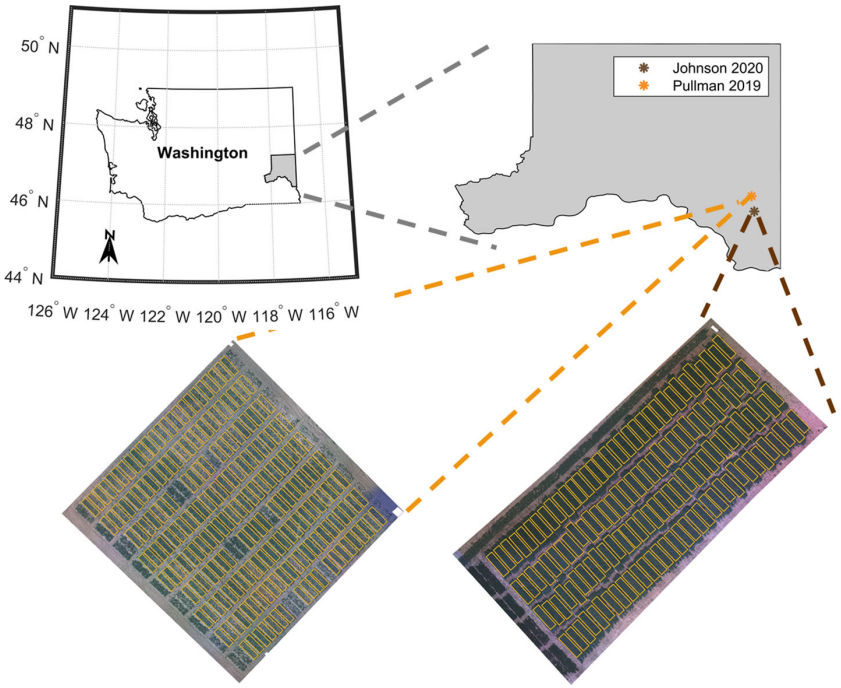


FIGURE 1
Location of the pea breeding field trials located in the Pacific Northwest region of USA in the state of Washington, USA: advanced yield trial grown in Pullman, Washington (including check varieties used as reference) evaluated in 2019 season (Site 1) and variety test trial grown in Johnson, Washington (including check varieties used as reference) evaluated in 2020 season (Site 2).

plots recognized as outliers (low yield) were excluded from further analysis. Moreover, some check varieties had more than three or four replicates. The 2020 field season was a longer season, with favorable weather conditions resulting in high yields and lower yield variance than the 2019 field season.

2.3 Satellite and UAS data acquisition

Multispectral images were collected using a quadcopter UAS (AgBot, ATI Inc., Oregon City, Oregon, USA), equipped with a five-band multispectral camera (RedEdge MX, Micasense Inc., Seattle, Washington, USA; Figure 2A and Supplementary Materials Tables 1, 2). Images were acquired with a resolution of 1.2 MP and dynamic range of 12-bit, at flying altitude of 25 m in 2019 (ground sampling distance (GSD) = 0.02 m/pixel) and 30 m in 2020 (GSD = 0.03 m/pixel). The flying speed was set to 2.5 m/s and the forward and side overlap to 80% in 2019 and 70% in 2020. Images covering the area of interest were stitched to generate an orthomosaic using Pix4Dmapper (Pix4D Inc., Lausanne, Switzerland). The description of the UAS image pre-processing technique can be found in Zhang et al. (2021). The UAS data collection was conducted twice at each season (17 June 2019 and 16 July 2019 at Site 1, 02 June 2020 and 06 July 2020 at Site 2). These time points coincided with two different growth stages of field pea (flowering and pod development at Site 1, vegetative stage and pod development at Site 2, Figure 2B). The orthomosaic bands [5 bands x 2 time points (TP 1 and TP 2) x 2 locations (Site 1 and Site 2)] were radiometrically calibrated by converting the raw imagery values (digital numbers) to surface reflectance using a Spectralon reflectance panel (99% reflectance; Spectralon, SRS-99-120, Labsphere Inc., North Sutton, NH, USA).

Satellite imagery from WorldView (-2 or -3) were obtained at the closest time to the UAS acquisition dates (11 June 2019 and 19 July 2019 at Site 1, 04 June 2020 and 24 July 2020 at Site 2, Figure 2B and Supplementary Materials Tables 1, 2). The satellite revisit frequency at a specific location is determined by factors such as the satellite's altitude, orbit, desired coverage area and location, as

TABLE 1 Summary statistics of field pea seed yield (kg/ha) from two locations.

Trials	No. of entries	No. of plots	Median	Mean	Max	Min	SD	CV (%)
Site 1	65	203	1964	1938	3195	662	498	26
Site 2	33	131	4476	4514	6395	2442	783	17

Site 1, advanced yield trial; Site 2, variety testing trial; SD, standard deviation; CV, coefficient of variation. One plot with low yield was considered as outlier and was removed from each trial.

10.3389/fpls.2023.1111575

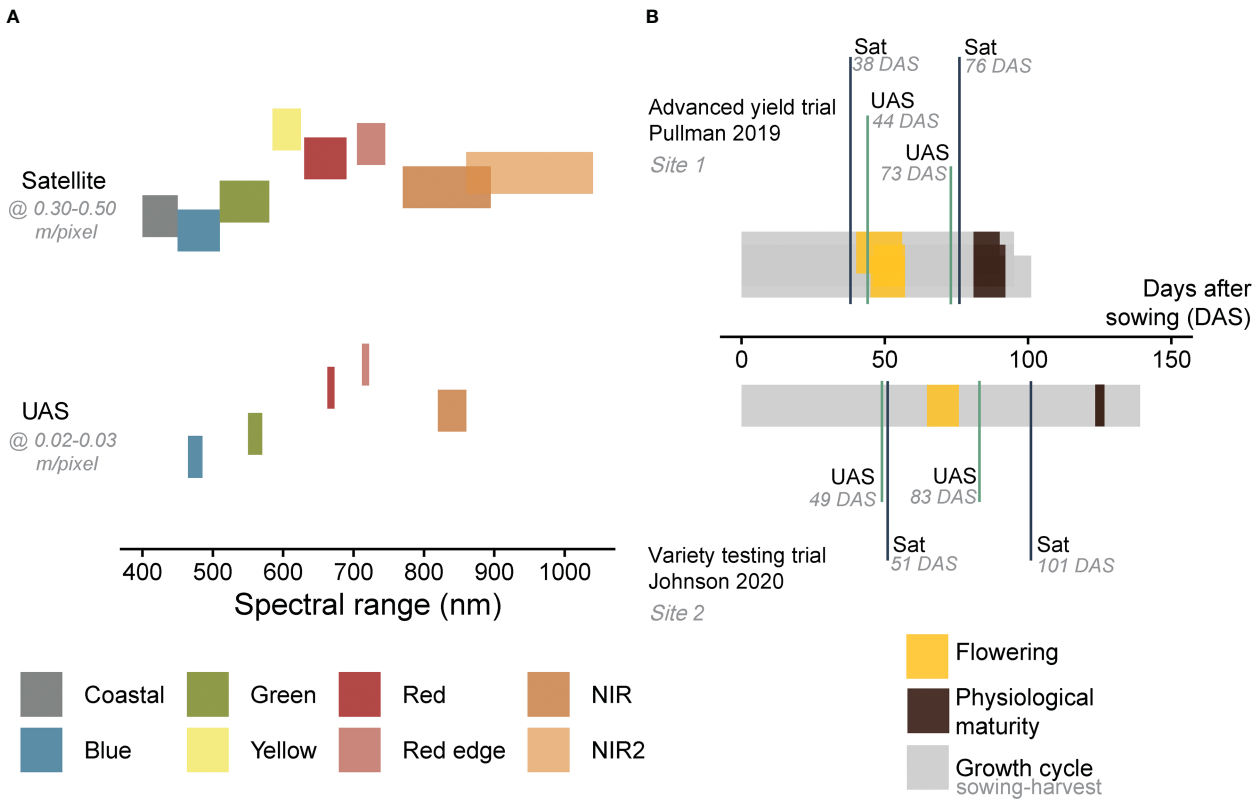


FIGURE 2
Spectral range of satellite and UAS multispectral imagery acquired in this study **(A)**. The data acquisition time points of remote sensing data **(B)**. Details about spectral resolution and acquisition dates are summarized in Supplementary Materials Table 1. DAS refers to days after sowing.

well as weather conditions. On average, the revisit time for WorldView satellites is approximately one day at 1 m GSD. The data were delivered by Maxar Technologies (Westminster, Colorado, USA) as standard level 2A with atmospheric compensation, which accounts for atmospheric scattering effects on the data. The imagery products were delivered as surface reflectance and comprised eight spectral bands in the visible near-infrared region (Figure 2A). The spatial resolution of the eight spectral bands in WorldView-2 or -3 ranges from ~1.8 to 1.2 m, while the panchromatic band has a resolution of ~ 0.50 or 0.30 m, respectively, depending on the sensor. The high-resolution panchromatic bands were then combined with the lower resolution multispectral bands to create a pan-sharpened image with improved spatial resolution and the final products were acquired with a GSD of about 0.30-0.50 m/pixel. In this study, the coastal bands were not included, resulting in [7 bands x 2 time points (TP 1 and TP 2) x 2 locations (Site 1 and Site 2)].

The images from both satellite and UAS sources were co-registered and aligned using the Georeferencer tool in QGIS (QGIS.org, 2021, version 3.10.16). The images were then cropped to the same region of interest covering the breeding trials. At each site, two shapefiles were manually created using the Digitizing tool in QGIS to delineate the boundaries of each plot. The alignment of satellite and UAS imagery facilitated the segmentation of plot boundaries.

3 Data processing and analysis

3.1 Overview of data analysis

The study explored the impact of different spatial resolutions and data fusion techniques on field pea seed yield estimation using remote sensing data. The analysis workflow was structured into three levels, as depicted in Figure 3.

The first level (Level 1) involved utilizing the surface reflectance imagery at its original spatial resolution. The GSD of satellite imagery was 0.30-0.50 m/pixel and the UAS imagery had a GSD of 0.02-0.03 m/pixel. The multispectral imagery was further processed to extract features that describe the spectral and texture information. Using these features as input, random forest models were trained to predict field pea seed yield. The performance of the models was evaluated using features extracted from multi-scale data (satellite and UAS), as well as features from single- and multi-time points. The assessment of the models incorporating features from multi-time points refers to the fusion of multi-modal (spectral and textural) and multi-scale vegetation features, extracted from both satellite and UAS imagery at combined time points.

The second level (Level 2) focused on evaluating the impact of image fusion on field pea seed yield estimation. This was achieved by enhancing the spatial resolution of satellite imagery using two

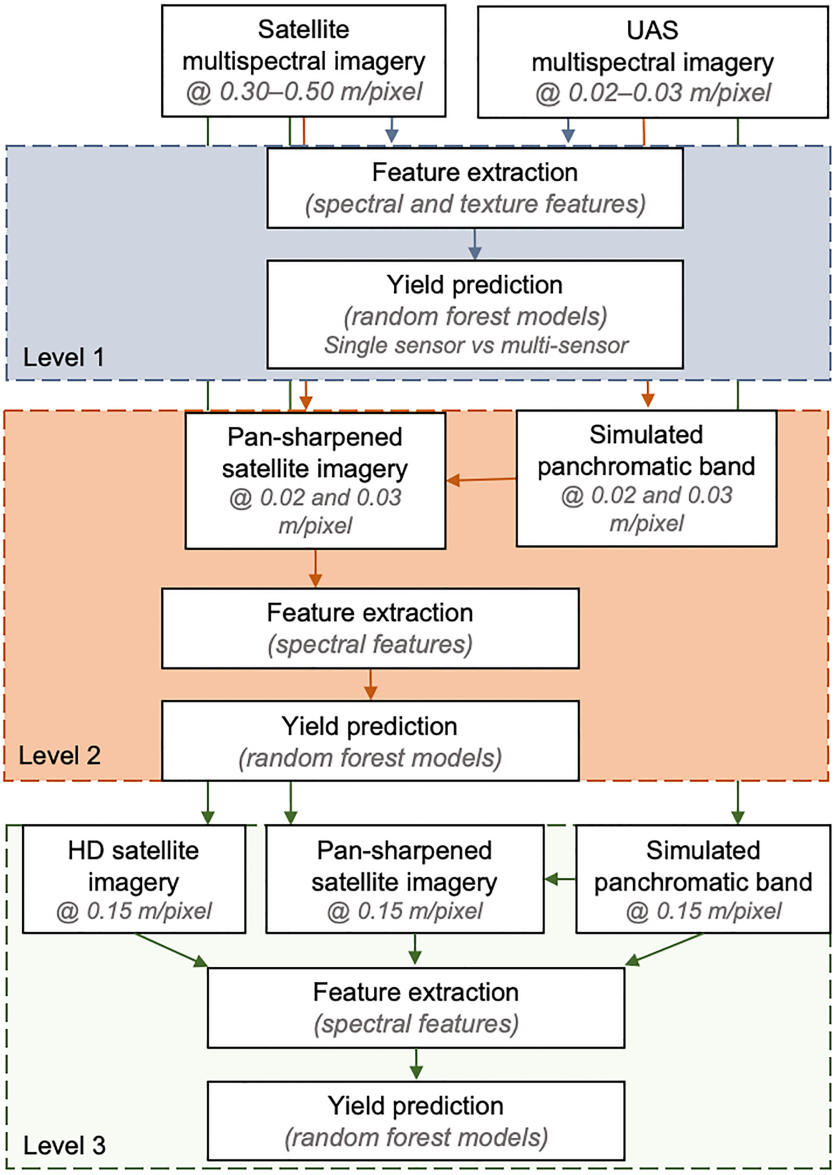


FIGURE 3
Image processing workflow for analyzing data extracted from satellite and UAS imagery. The top, middle and lower boxes refer to the evaluations in Level 1, Level 2, and Level 3, respectively. Level 1, 2, and 3 included feature extraction and yield prediction using extracted features from original resolution of the satellite and UAS imagery, pan-sharpening of satellite imagery using UAS imagery and feature extraction from the pan-sharpened imagery for yield prediction, and comparison of yield prediction using pan-sharpened image-based features and HD image features, respectively.

pan-sharpening techniques with UAS data to reach a satellite image GSD of 0.02-0.03 m/pixel. The panchromatic bands were simulated by averaging the five UAS spectral bands. The quality of the resulting synthesized pan-sharpened satellite imagery was evaluated using five image evaluation metrics, which are described in more detail in the next section. Additionally, the spectral quality of the pan-sharpened imagery was assessed by training random forest models to predict seed yield using spectral features (vegetation indices) derived from these pan-sharpened satellite imagery. These evaluations were conducted for both Site 1 and Site 2 data.

The final analysis level (Level 3) focused on comparing between two techniques of satellite image spatial resolution enhancement for seed yield estimation. The two techniques for enhancing satellite image spatial resolution were super-resolution and pan-sharpening. The super-resolution images were acquired as "High-Definition" products (HD) with a GSD of 0.15 m/pixel from Maxar Technologies. The HD products were generated using their proprietary super-resolution technique, which enhances the spatial details of the satellite image during post-processing using a machine learning approach. Since the HD satellite imagery was only available for Site 2 (two–time points), this aspect was assessed only for this location. These images were compared to the pan-sharpened satellite imagery (similar to those described in Level 2) but with a spatial resolution of 0.15 m/pixel and four spectral bands – RGB and NIR to match the HD image spatial resolution.

3.2 Satellite pan-sharpening

Image fusion is the process of combining multiple images taken at different times or with different sensors to create a single image that contains more information than the original imagery. Pan-sharpening is a type of image fusion approach, which combines the high spatial resolution of the panchromatic band (having a high spatial resolution) with a lower resolution multispectral imagery (RGB or multispectral images) to enhance the spatial resolution of the latter. The result is a single image that with usually a high spatial and spectral resolution. This technique is commonly used to enhance the quality and resolution of the satellite imagery. There are multiple approaches that have been developed for image fusion, each with varied performance efficiencies (Ehlers et al., 2010; Yokoya et al., 2017; Gašparović et al., 2019; Ghamisi et al., 2019; Dadrass Javan et al., 2021). In general, image fusion can be broadly characterized into component substitution, multi-resolution analysis, variational optimization-based techniques, and machine learning-based approaches (Vivone et al., 2019; Vivone et al., 2021). This study selected a technique representing the component substitution and the multi-resolution approaches. These techniques were adopted from the MATLAB Pan Sharpening toolbox (Vivone et al., 2015).

The intensity-hue-saturation (IHS) transformation is a component substitution approach, where the bands of the low-resolution multispectral imagery are converted to IHS components. The intensity component is replaced by the panchromatic band after histogram matching (Al-Wassai et al., 2011; Johnson et al., 2014). In general, the extraction of the intensity component is performed by averaging the bands in the visible region (red, green, blue). In this study, we computed the intensity component by averaging all the spectral bands from the low-resolution satellite imagery (Yilmaz et al., 2019), excluding the coastal band. Similarly, additive wavelet luminance proportional (AWLP) pan-sharpening, which is a multi-resolution analysis approach, was utilized for image spatial enhancement. In the AWLP approach, a low-spatial resolution image is decomposed into scale levels while injecting the panchromatic band matched by each decomposed layer and applying an inverse transformation (Dadrass Javan et al., 2021).

The ideal pan-sharpened imagery should have the same spatial properties as the high-resolution panchromatic band and the same spectral properties as the multispectral input bands, though the process can lead to spectral and/or spatial distortions (Siok et al., 2020). The assessment of the quality of the resulting pan-sharpened imagery can be conducted using Wald's protocol, which states that the pan-sharpening process is reversible, and that the original multispectral imagery can be obtained by degrading the pan-sharpened imagery (Wald et al., 1997; Vivone et al., 2015). In this study, the spectral and spatial qualities of the pan-sharpened imagery were assessed using five statistical metrics (correlation coefficient CC, structural similarity index measure SSIM, spectral angle mapper SAM, erreur relative globale adimensionnelle de synthese ERGAS, and peak signal to noise ratio PSNR), after degrading their spatial resolution to match that of the original satellite multispectral imagery (Borra-Serrano et al., 2015; Luo et al., 2018; Li et al., 2020; Siok et al., 2020).

3.3 Background removal and feature extraction

Background segmentation and soil removal were conducted prior to feature extraction. A threshold based on the soil adjusted vegetation index (SAVI) pixel values was utilized for UAS imagery. However, for satellite imagery, segmentation was based on histogram distribution of SAVI pixel intensity. Thresholds were set as 15% for satellite imagery at spatial resolution of 0.15, 0.30, and 0.50 m/pixel and 25% for the pan-sharpened satellite imagery at 0.02 and 0.03 m/pixel. These thresholds were selected based on visual observations of the SAVI intensity distribution and implemented to eliminate spectrally mixed pixels. Median reflectance values from each vegetation index (normalized difference vegetation index NDVI, green normalized difference vegetation index GNDVI, normalized difference red-edge index NDREI, soil adjusted vegetation index SAVI, atmospherically resistant vegetation index ARVI, transformed triangular vegetation index TVI, infrared percentage vegetation index IPVI, renormalized difference vegetation index RDVI, two-band enhanced vegetation index EVI2, normalized difference red-edge index – with yellow band NDRE2, and normalized difference vegetation index – with NIR2 band NDVI2) were extracted as canopy spectral features. Additionally, texture features were extracted from the individual spectral bands (5 bands for UAS data and 7 bands for satellite data) using the GLCM (grey level co-occurrence matrix) approach. These features included contrast (CO), homogeneity (HO), correlation (CR), and energy (EN). A detailed summary of extracted features is provided in Supplementary Materials Table 3. All image-related analyses were conducted using a customized script in MATLAB (Matlab, 2021b; MathWorks Inc., Natick, Massachusetts, USA).

3.4 Statistical analysis, yield prediction, and feature importance

The statistical analysis, model development and validation, feature assessment, and visualization were performed in R (http://www.r-project.org/; release 4.0.5). The correlation analysis was performed to evaluate the similarity between the extracted imagery-based features and harvested seed yield at each time point and for each type of imagery data (satellite and UAS). Moreover, machine learning algorithms were applied to estimate harvested seed yield and evaluate the importance of the extracted features. Random forest models were constructed using *randomForest* implementation in caret (Kuhn, 2008). The models were trained using non-scaled data. The coefficients of determination (R^2) and root mean square error (RMSE) were computed to evaluate the performance of the yield prediction model. The mean ± standard deviation computed from multiple runs was reported for both R^2 and RMSE.

Each type of data was divided into training and testing sets (80/20). At the beginning of our analysis, two training frameworks were tested: random holdout plot and random holdout entry. For the random holdout plot (plots were held back for testing irrespective of entry), 80% of the data were randomly chosen. For the random holdout entry (all replicates from an entry were held back for testing), 80% of the entries

were randomly chosen. The models trained with random holdout entry showed stable performance with lower variation in the testing set than those trained with random holdout plot (Supplementary Materials Tables 4, 5). Therefore, we selected the random holdout entry as a data resampling technique.

The random forest hyperparameters were kept at the default level, and only the number of variables at each split was tuned using 3-fold internal cross-validation repeated 15 times. The data split was repeated 10 times by setting 10 random seeds in each run (referred to as 10 random runs henceforth) to assess the variability of model performance with different data splits. The input spectral features used to train the models are summarized in Supplementary Materials Table 6. In addition to spectral features extracted from individual sensor UAS and satellite at original spatial resolution at each time point and combined time points, integrating spectral and texture features was also evaluated. Before training the model for each random run, feature selection was conducted on the training dataset to remove highly correlated features with a correlation coefficient threshold of 0.99.

Feature importance was evaluated using a permutation-based method (increased mean squared error IncMSE after removing one feature). Three ranks were created based on three factors to create a new metric reflecting feature importance across all experiments. This new metric was denoted as the adjusted rank (%) hereafter in the text. The first factor (R^1, ranking of the numeric score) was the minimum value of feature importance computed by extracting the minimum values of IncMSE from 10 random runs, which was ranked such that a lower rank would indicate high importance of the feature across multiple runs. The second factor was based on the coefficient of variation (CV) of the feature importance across 10 random runs, where the factor was also ranked (R^2, ranking of the CV percentage). A high CV indicates that the variability of feature importance score is high. The third factor (R^3, ranking of the frequency) was the occurrence of the feature during 10 random runs after feature selection using a correlation filter. The third factor was ranked so that the lowest rank indicates consistent selection of the feature (for example, rank 1 would indicate the presence of the feature in all 10 random runs). In summary, a lower rank of the first factor would indicate a higher feature importance, a lower rank of the second factor would indicate a lower CV of feature importance across multiple random runs, and a lower rank of the third factor would indicate a higher frequency of feature occurrence used to train the models. The ranks were converted to adjusted rank by dividing the sum of the ranks (all factors) for each feature by the sum of maximum ranks within each category (all factors) and subtracting this ratio from one (Eq. 1).

$$R_F^A = 1 - \frac{(R_F^1 + R_F^2 + R_F^3)}{(maxR_F^1 + maxR_F^2 + maxR_F^3)} \quad (1)$$

where, R_F^A is the adjusted rank for each feature; R_F^1, R_F^2, and R_F^3 are the ranks of each factors; and *max* refers to maximum rank within each category. This equation holds only when the specific feature was selected more than once within multiple runs. The adjusted rank was normalized ($norm_R F^A$) and presented as a percent (Eq. 2).

$$norm_R F^A(\%) = \frac{R_F^A}{(max\ R_F^A)} \times 100 \quad (2)$$

Feature comparison was performed by observing the normalized adjusted rank data.

4 Results

4.1 Spectral features from remote sensing data and seed yield

The satellite-based features were comparable to UAS-based features in both trials, where the vegetation indices from both sources were significantly and positively correlated (Pearson correlation coefficient, $0.14 \leq |r| \leq 0.78$, $p < 0.05$) (Figure 4A). This relationship (Pearson correlation coefficient = $0.50 \leq r \leq 0.78$ and $p < 0.0001$) was consistent across time points and sites for eight VIs (except NDREI for Site 1).

The seed yield varied between the two sites as a function of the environment, location, and the evaluated entries. The yield recorded at Site 2 in 2020 (4514 ± 783 kg/ha, n = 131) was higher than at Site 1 in 2019 (1938 ± 489 kg/ha, n = 203). It was encouraging to note that the correlation between the extracted spectral features from satellite and UAS imagery and seed yield data showed similar patterns (Figure 4B). Satellite- and UAS-based VIs with the combination of red and NIR bands (e.g., EVI2, RDVI, IPVI, ARVI, SAVI, and NDVI) were significantly and positively correlated with yield data (satellite: $0.38 \leq r \leq 0.80$, UAS: $0.20 \leq r \leq 0.72$, and $p < 0.05$). Correlation between other spectral features with yield data varied depending on time point (~ crop growth stage) and location. On the other hand, the correlation analysis between texture features extracted from both sensing platforms and seed yield data showed a weak to no correlation (Supplementary Materials Figure 3). The relation between these features and seed yield were further assessed in training random forest models to predict the final seed yield in the following sections.

4.2 Yield estimation using remote sensing data (Level 1)

4.2.1 Yield estimation using multi-modal data at separate time points

During the analysis of individual time points, the satellite- and UAS-based features were evaluated based on the performance of random forest models, trained with these features, to predict harvested seed yield (Figure 5; training results and spatial distribution of yield differences in Supplementary Materials Table 7 and Figure 4, respectively). On average, UAS data-based models performed better (higher R^2, lower RMSE) compared to satellite data-based models (R^2 = 0.36 ± 0.23; RMSE=523 ± 197 kg/ha for satellite, R^2 = 0.46 ± 0.16; RMSE=486 ± 173 kg/ha for UAS). However, the difference between the results from the two sensing approaches (two image scales) varied with the crop growth stage during data

10.3389/fpls.2023.1111575

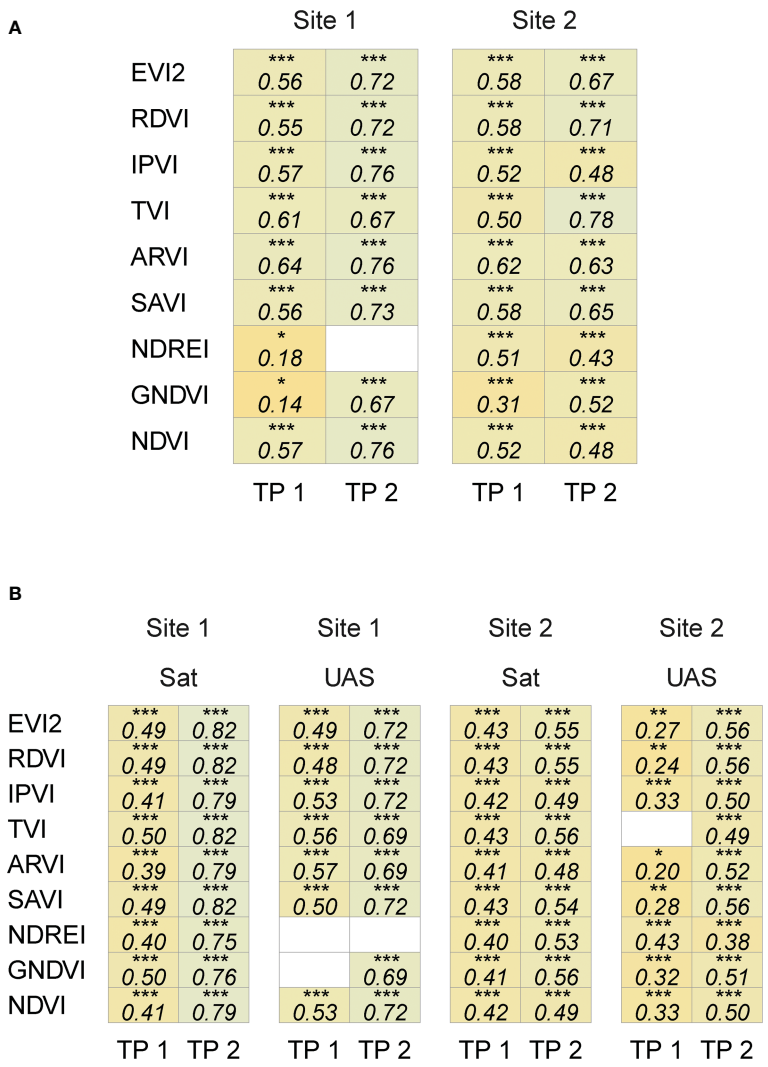


FIGURE 4
Correlation coefficient between spectral features extracted from satellite and UAS imagery **(A)**, and with seed yield **(B)**. White cells indicate non-significant correlation ($p \geq 0.05$), *($0.01 < p \leq 0.05$), **($0.001 < p \leq 0.01$), and ***($p \leq 0.001$).

acquisition (individual time point or combined time points), the breeding trial and its yield variability (Site 1 and Site 2), and the type of features (spectral or spectral + texture) used as input in the models.

Models trained on features extracted from imagery acquired at TP 1 (crop growth stage around flowering at Site 1 and early vegetative growth at Site 2) had weaker performance than TP 2 (pod development). For instance, models developed with features extracted from UAS imagery at TP 2 showed better performance compared to those extracted at TP 1 ($R^2 = 0.25 \pm 0.15$; RMSE=570 ± 180 kg/ha for TP 1, $R^2 = 0.45 \pm 0.17$; RMSE=489 ± 189 kg/ha for TP 2). At this growth stage at Site 1 (TP 2), satellite data-based models showed an improved performance compared to UAS data-based models ($R^2 = 0.63 \pm 0.09$; RMSE=295 ± 37 kg/ha for satellite, $R^2 = 0.53 \pm 0.11$; RMSE=335 ± 58 kg/ha for UAS). However, at Site 2, satellite data-based models did not outperform UAS data-based models.

Satellite and UAS data-based models performed similarly even when texture features or the two unique bands, yellow (~ 585 – 625 nm) and NIR2 (~ 860 – 1040 nm), were added to the common nine VIs. In both cases (satellite and UAS), the model performances did not improve significantly, although small increases or decreases were occasionally observed (Figure 5 and Supplementary Materials Table 7). Overall, for satellite data-based models, adding texture information reduced the accuracy of models, particularly for the dataset at TP 1. This could be due to the spectral mixing and additional noise. The sample size with respect to the increasing feature space could also contribute to these results.

When the time points were combined, the accuracy of random forest models in predicting seed yield increased in most cases at both sites and for both types of datasets (satellite and UAS). At Site 1, the accuracy of satellite data-based models was slightly better than UAS data-based models ($R^2 = 0.65 \pm 0.10$; RMSE=284 ± 45 kg/ha for satellite, $R^2 = 0.64 \pm 0.06$; RMSE=296 ± 55 kg/ha for UAS) when using the combined dataset (TP 1 + TP 2). In fact, combining features from the two time points extracted from satellite or UAS imagery improved the model performance regardless of the image scale (Figure 5 and Supplementary Materials Table 7).

10.3389/fpls.2023.1111575

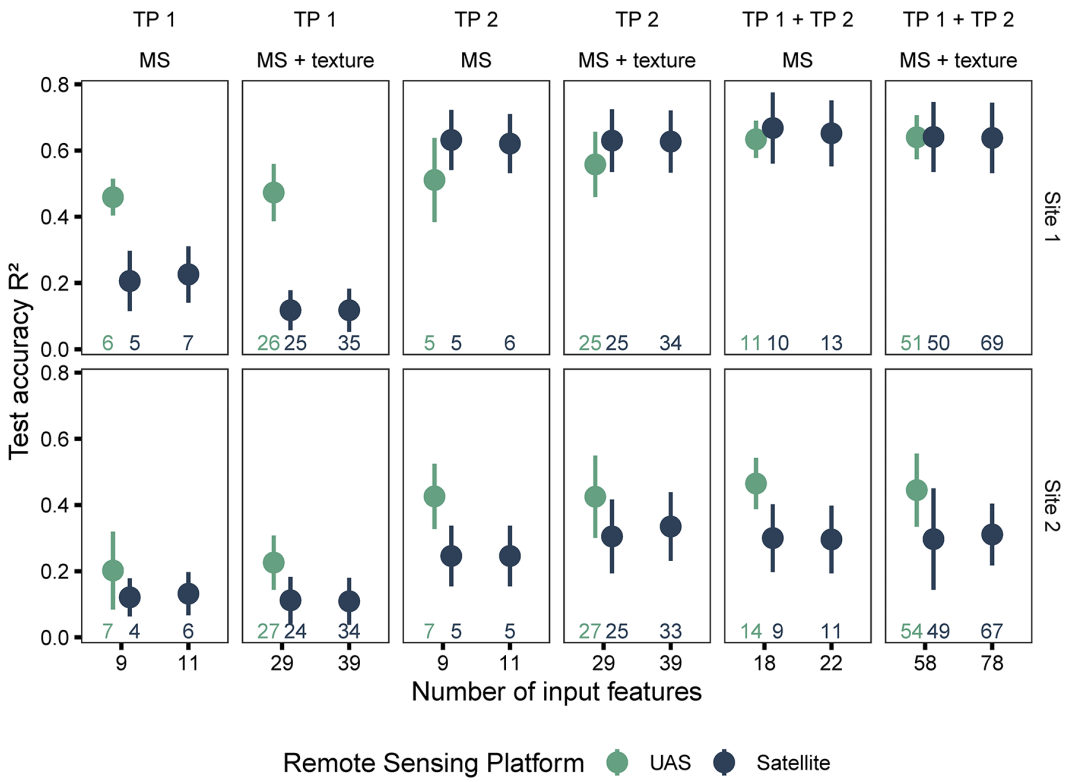


FIGURE 5
Model performance based on data source (satellite and UAS), type of features used in model (spectral and spectral + texture), and time points (individual and combined time points). The number of input features represented as labels of the x-axis indicates the number of features before the correlation filter. The colored numbers above the x-axis represent the total number of features retained after the correlation filter, which were used to build the random forest models.

In summary, the satellite data-based models performed poorly compared to UAS-based models, especially at early pea growth stages (prior to canopy closure). The gap in performance was reduced at the later TP 2 and was further improved with combining the data from two time points (especially for Site 1, which could be associated with the crop growth stage). Moreover, the texture information did not improve the performance of the model. Since texture data did not improve the model performance, only spectral data was used for evaluating image and feature fusion henceforth.

4.2.2 Yield estimation using fused features datasets

Table 2 summarizes the performance of the random forest models resulting from data fusion (three scenarios) of spectral features extracted from satellite and UAS imagery (spatial distribution of yield differences in Supplementary Materials Figure 5). The overall trend depicts that combining information from the two sensing approaches improved the prediction accuracy compared to individual time points and/or individual sensors. Particularly, at Site 1, combining

TABLE 2 Performance evaluation of models using multi-scale/sensor input features.

Site	Scenario	Input/Selected features	Train		Test	
			R^2	RMSE (kg/ha)	R^2	RMSE (kg/ha)
1	UAS_{TP1} + $Satellite_{TP2}$	**18/11**	**0.69 ± 0.03**	**284 ± 19**	**0.69 ± 0.09**	**272 ± 53**
	$Satellite_{TP1}$ + UAS_{TP2}	18/10	0.57 ± 0.04	331 ± 21	0.58 ± 0.11	316 ± 62
	$UAS_{TP1+TP2}$ + $Satellite_{TP1+TP2}$	**36/21**	**0.72 ± 0.02**	**272 ± 13**	**0.72 ± 0.08**	**259 ± 53**
2	UAS_{TP1} + $Satellite_{TP2}$	18/12	0.35 ± 0.07	661 ± 44	0.41 ± 0.14	621 ± 75
	$Satellite_{TP1}$ + UAS_{TP2}	18/11	0.32 ± 0.04	673 ± 24	0.44 ± 0.07	603 ± 55
	$UAS_{TP1+TP2}$ + $Satellite_{TP1+TP2}$	**36/23**	**0.37 ± 0.05**	**638 ± 33**	**0.50 ± 0.11**	**575 ± 70**

The values reported are mean±standard deviation.
Bold values indicate the best-performing models.

features extracted from UAS_{TP1} and $Satellite_{TP2}$ improved the performance of random forest models by a mean increase of R^2 ranging between 3–50% (decrease in RMSE by 1–26%) compared to the single sensor approach. However, for Site 2, the performance of the random forest models did not improve with integrated features from UAS_{TP1} and $Satellite_{TP2}$ imagery. When compared with the model performance of $UAS_{TP1+TP2}$ dataset, combining the features extracted from the scenario of UAS_{TP1} and $Satellite_{TP2}$ decreased the prediction accuracy (R^2 = 0.47 ± 0.08; RMSE=587 ± 57 kg/ha for $UAS_{TP1+TP2}$, R^2 = 0.41 ± 0.14; RMSE=621 ± 75 kg/ha for UAS_{TP1} + $Satellite_{TP2}$). Nevertheless, when all the features extracted from both sources (satellite and UAS) and both time points were integrated, the prediction accuracy surpassed the best result acquired from the individual sensor ($UAS_{TP1+TP2}$) with an increase in R^2 and a decrease in RMSE. This increase in performance could be attributed to the addition of features extracted from UAS_{TP2}.

4.3 Evaluation of multi-scale image fusion (Level 2)

4.3.1 Qualitative and quantitative spectral evaluation of pan-sharpened imagery

Qualitative evaluation of the pan-sharpening methods was based on visual inspection of the resulting images. Figures 6 and 7 illustrate image fusion results using UAS as a panchromatic band at their original resolution for each time point at Site 1 and Site 2, respectively (correlation with yield presented in Supplementary Materials Figure 6).

Spectral distortion (based on color) can be noted in images obtained using the IHS technique, while the AWLP technique preserved the spectral quality of the satellite imagery in most cases. On the other hand, the spatial quality and resolution of the IHS technique were visually better than the AWLP technique.

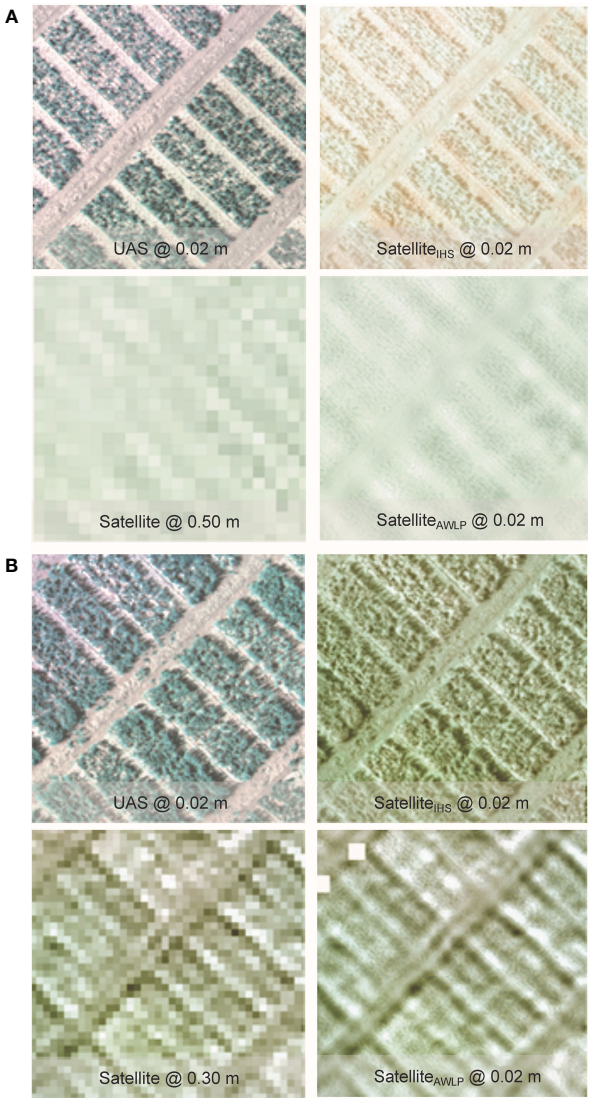


FIGURE 6
Visual comparison of RGB from the original satellite and UAS imagery, and pan-sharpened satellite imagery generated using IHS and AWLP approaches from Site 1 at TP 1 **(A)** and TP 2 **(B)**.

10.3389/fpls.2023.1111575

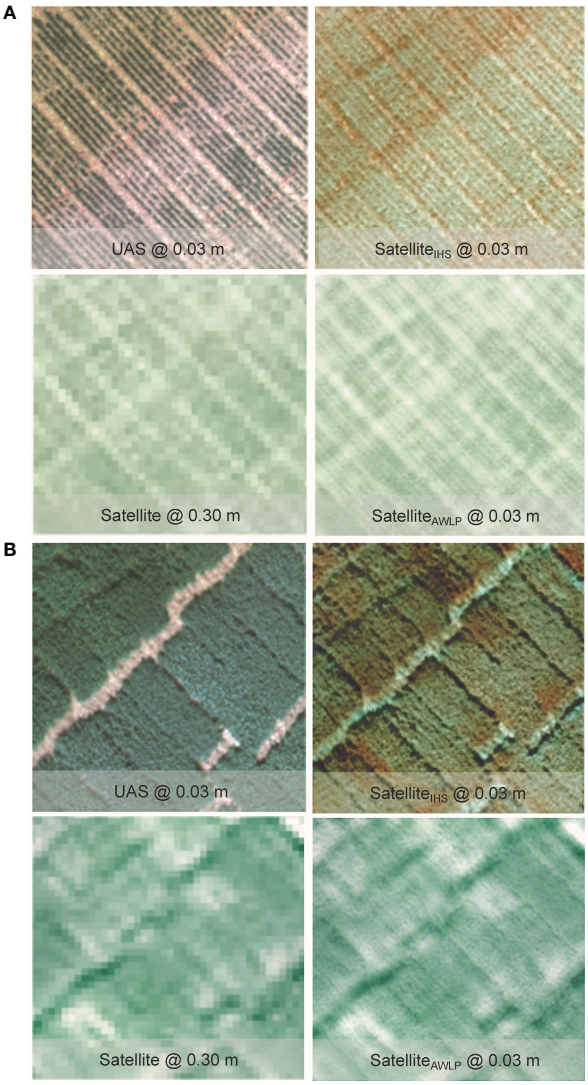


FIGURE 7
Visual comparison of RGB from the original satellite and UAS imagery, and pan-sharpened satellite imagery generated using intensity-hue-saturation (IHS) and additive wavelet luminance proportional (AWLP) approaches from Site 2 at TP 1 **(A)** and TP 2 **(B)**.

TABLE 3 Image comparison metrics correlation coefficient (CC), the structural similarity index measure (SSIM), the peak signal to noise ratio (PSNR), the erreur relative globale adimensionnelle de synthese (ERGAS), and the spectral angle mapper (SAM) comparing different pan-sharpened imagery (GSD = 0.02-0.03 cm/pixel) generated using intensity-hue-saturation (IHS) and additive wavelet luminance proportional (AWLP) approaches with original satellite imagery.

Site	Original imagery	Technique	CC	SSIM	PSNR	ERGAS	SAM
1	UAS_{TP1} + $Satellite_{TP1}$	IHS	0.82	0.77	23.47	2.75	0.87
		AWLP	0.81	0.77	23.21	2.83	0.21
	UAS_{TP2} + $Satellite_{TP2}$	IHS	0.87	0.81	28.27	3.84	1.96
		AWLP	0.88	0.85	28.43	3.72	0.88
2	UAS_{TP1} + $Satellite_{TP1}$	IHS	0.84	0.84	25.46	6.92	1.80
		AWLP	0.84	0.85	25.57	6.81	0.43
	UAS_{TP2} + $Satellite_{TP2}$	IHS	0.83	0.77	25.53	7.20	2.90
		AWLP	0.89	0.85	26.72	5.94	0.70

Overall, we notice that the IHS technique generated images that were very similar to the panchromatic bands (UAS imagery).

Table 3 summarizes the results of the evaluation metrics of the pan-sharpened imagery. The AWLP method showed comparable image evaluation metrics (CC, SSIM, PSNR, and ERGAS) to the IHS method. The major difference was observed in SAM values, where images pan-sharpened with the AWLP approach had 1.33 times lower SAM values compared to the images pan-sharpened with the IHS approach (averaged over all case studies). This indicates better spectral quality comparison between the pan-sharpened image and reference image (original satellite imagery).

4.3.2 Yield estimation using satellite pan-sharpened imagery

In terms of image evaluation metrics, especially SAM, the AWLP approach was better than the IHS approach. However, models trained with features extracted from satellite imagery pan-sharpened with the IHS approach were more accurate in predicting seed yield than models trained with features extracted from satellite imagery pan-sharpened with the AWLP approach (Figure 8 and Supplementary Materials Table 8).

The potential reasons for this observation could include the following: (i) image spatial quality was critical in addition to spectral quality for yield prediction (IHS approach displayed higher spatial quality than AWLP approach), (ii) the image evaluation metrics do not necessarily indicate statistical quantitative assessment (Yokoya et al., 2017, specified in reference to component substitution approach such as IHS transformation), which in this case is yield prediction, (iii) the image evaluation metrics were mainly developed for satellite to satellite image comparisons, and (iv) it was observed that the AWLP approach produced small anomalies during image pan-sharpening that may have affected the results. It should be noted that at Site 1 and with TP 2 and TP 1 + TP 2 datasets, the performance of models built with the features extracted from pan-sharpened imagery with the two approaches did not differ.

When the models were trained with satellite-based features extracted at original resolutions, model performance with features extracted from $Satellite_{IHS}$ increased, particularly at TP 1. For example, the mean increase in R^2 ranged between 67% at Site 1 (R^2 = 0.21 ± 0.09; RMSE=442 ± 62 kg/ha for $Satellite_{TP1}$ at GSD = 0.50 m/pixel, R^2 = 0.35 ± 0.07; RMSE=397 ± 49 kg/ha for $Satellite_{TP1}$ at GSD = 0.02 m/pixel) to 92% at Site 2 (R^2 = 0.12 ± 0.06; RMSE=747 ± 47 kg/ha for $Satellite_{TP1}$ at GSD = 0.30 m/pixel, R^2 = 0.23 ± 0.08; RMSE=702 ± 43 kg/ha for $Satellite_{TP1}$ at GSD = 0.03 m/pixel). Our results show that although the models with pan-sharpened imagery had weaker performance than those obtained from UAS-based models at original resolution, the performance of satellite-derived features, in general, improved after pan-sharpening (spatial enhancement).

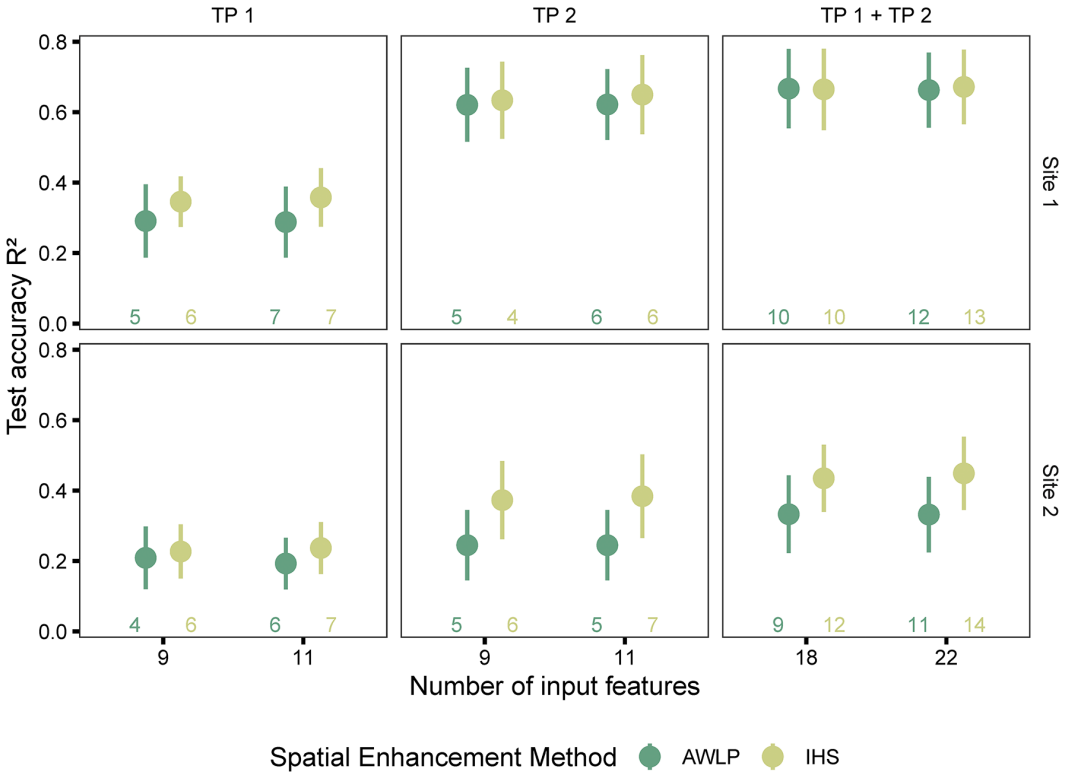


FIGURE 8
Model performance based on spectral data (pan-sharpened imagery developed using integration of satellite and UAS imagery using AWLP and IHS approaches), and time points (individual and combined time points). The number of input features represented as labels of the x-axis indicates the number of features before the correlation filter. The colored numbers above the x-axis represent the total number of features retained after the correlation filter, which were used to build the random forest models.

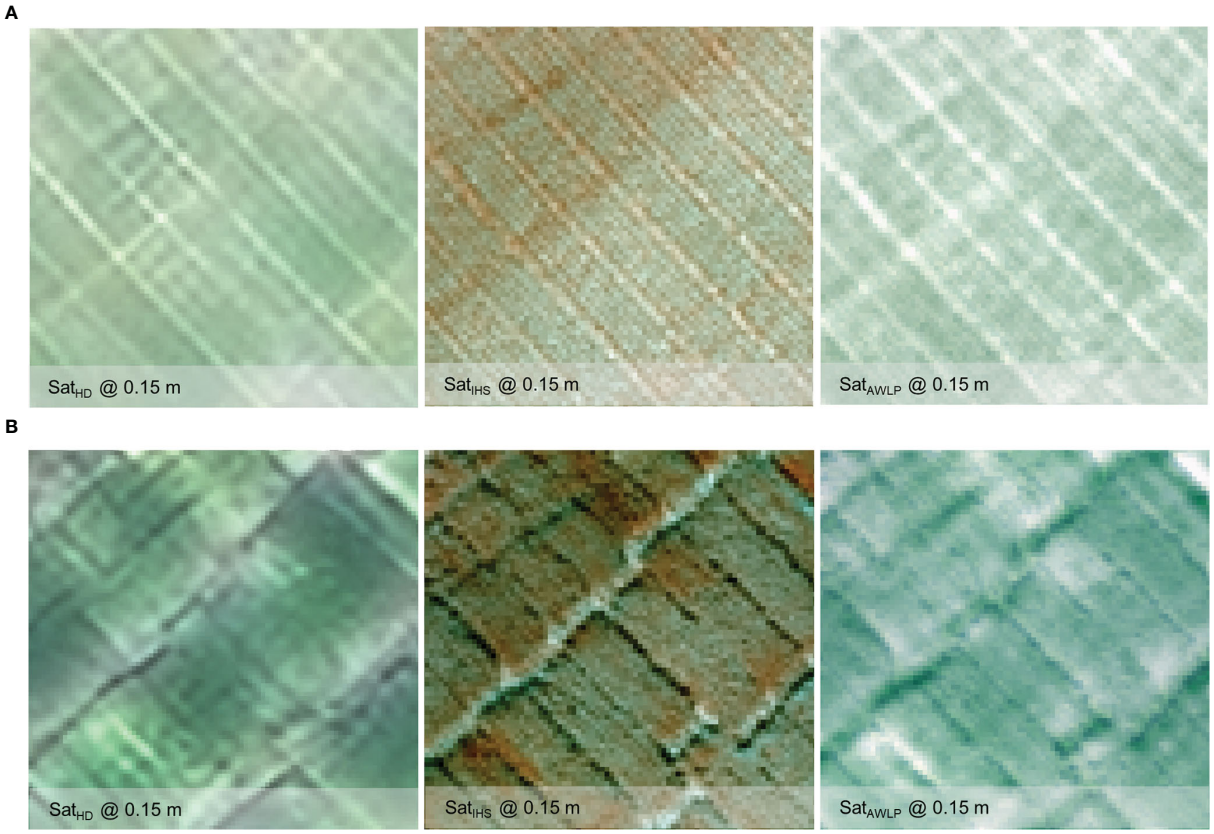


FIGURE 9
Visual comparison of the high definition (HD) and pan-sharpened satellite imagery generated using intensity-hue-saturation (IHS) and additive wavelet luminance proportional (AWLP) approaches from Site 2 at TP 1 **(A)** and TP 2 **(B)**.

4.4 Comparison of high definition and pan-sharpened imagery (Level 3)

The pan-sharpened imagery was developed by fusing the satellite and UAS images to a resolution (GSD = 15 cm) similar to HD imagery (correlation with yield presented in Supplementary Materials Figure 7) for Site 2 datasets. Figure 9 and Supplementary Materials Table 9 present the visual representation and spectral and spatial quality metrics variation upon image fusion, respectively. On average, all spectral metrics were comparable between IHS and AWLP approaches, although IHS or AWLP approach may be better in some cases depending on the time points. Nevertheless, in terms of SAM values, the AWLP approach resulted in better results than the IHS approach, similar to the previous section (Table 3).

Comparing image evaluation metrics (with original satellite imagery as reference image) of HD images with pan-sharpened images developed using IHS and AWLP approaches, the CC, SSIM, PSNR, and ERGAS values were better, while SAM values were higher. The yield prediction accuracy (Figure 10) varied based on the image fusion approach. The random forest model developed with features extracted from Satellite$_{IHS}$ showed the highest mean accuracy (especially TP 2 and TP 1 + TP 2) in comparison to the model developed using features extracted from Satellite$_{AWLP}$ and Satellite$_{HD}$ images.

As previously discussed, all models evaluated with features extracted from TP 1 imagery performed poorly and are probably associated with the early growth stage. For the combined time points dataset, the model developed with features extracted from the Satellite$_{HD}$ image performed poorly (Supplementary Materials Table 10). The models trained with features extracted from the Satellite$_{IHS}$ approach gave a comparable performance to those trained with UAS-features (original resolution retrained using the same number of features); with an average increase in R^2 of only 3% for UAS data-based models. These models outperformed the satellite-based features (original resolution retrained with the same number of features) by 57%.

4.5 Feature importance

The evaluation of feature importance based on the adjusted rank takes into account different sources of variability that might influence the assessment instead of relying solely on the feature importance scores (permutation feature importance). These ranks (presented as percentages) reflected feature importance scores, consistency across 10

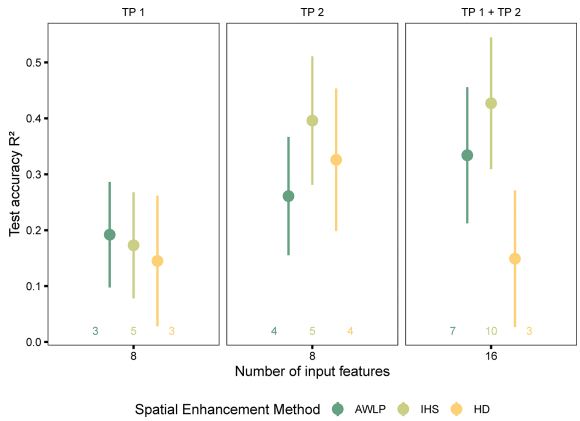


FIGURE 10
Model performance based on spectral data (pan-sharpened imagery developed using integration of satellite and UAS imagery using AWLP and IHS approaches with GSD = 0.15 m), and time points (individual and combined time points). The number of input features represented as labels of the x-axis indicates the number of features before the correlation filter. The colored numbers above the x-axis represent the total number of features retained after the correlation filter, which were used to build the random forest models.

random runs, and the correlation between features (which was translated as the frequency of occurrence). Therefore, in this section, we focus on reporting the variation of adjusted rank as a metric to capture the stability of these features (Figure 11).

Both satellite- and UAS-based models indicate that VIs with NIR and G spectral bands (e.g., GNDVI and TVI; Supplementary Materials Figure 8) are the most relevant and consistent features in this study. We found that VIs that are computed with the NIR and R spectral bands are more likely to be collinear (highly correlated with each other) in at least one of the random runs and, as a result, were eliminated before the random forest model training step. With a correlation coefficient threshold of 0.99, SAVI, ARVI, NDVI, RDVI, and EVI2 were found to be highly correlated with each other. When combining features extracted from imagery acquired at both time points, features from TP 2 were selected more frequently during model training, which indicates the importance of early pod development stages for capturing differences in crop performance between entries (Zhang et al., 2021).

5 Discussion

UAS-based phenomics is an accurate and efficient tool for providing quality features that supplement traditional field phenotyping. This study demonstrates that high-resolution satellite

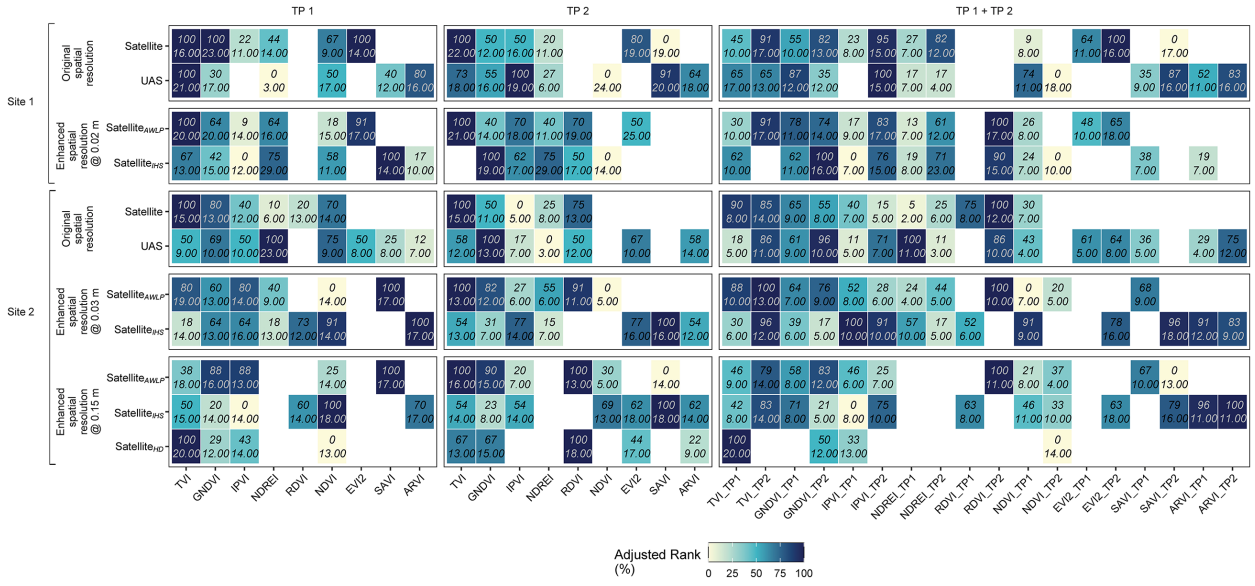


FIGURE 11
Adjusted rank (%) extracted to evaluate the feature importance from random forest models developed using different datasets. The comparison was made with respect to original (satellite and UAS) and pan-sharpened imagery across multiple time points and trials. In each square, there are two numbers, the top one indicates adjusted rank (%) and bottom one in, *italics*, is indicative of mean importance (%IncMSE).

imagery with a spatial resolution of 0.30 m/pixel can provide quality features as good as UAS-based imagery when evaluating the harvested seed yield of field pea genotypes in small (~9 m^2) breeding plots at later growth stage. The vegetation indices extracted from satellite imagery can be associated with seed yield at the breeding plot level. These results are in agreement with other studies (Tattaris et al., 2016; Sankaran et al., 2021). However, at earlier growth stages, there was no association between canopy reflectance extracted using satellite imagery and the harvested seed yield, which could be explained by the spatial resolution (pixel size) and the problem of spectral mixing. During this stage, ~50 DAS at Site 2, the pea plants are still in vegetative stage. Spectral mixing with soil adds noise and may reduce the quality of vegetation indices extracted from satellite imagery. Dalla Marta et al. (2015) found similar results when evaluating satellite-based features to estimate the nitrogen concentration of durum wheat at an early stage. In our study, the accuracy of predicting harvested seed yield can be related to the data acquisition time and the crop's growth stage for breeding plots of 9 m^2 (25 to 50 pixels per plot) compared to >1000 pixels per plot, as is the case for the UAS imagery. However, another factor that might have played a role in the low model accuracies at Site 2 is the data size. We hypothesized that the features extracted from images are proxies of the plant traits, which follow a non-linear trend with field pea seed yield. The choice of random forest models stems from the fact that this model was extensively used for seed yield prediction in remote sensing applications. The R^2 during prediction at Site 1 was up to 0.60, with a total number of observations or plots of 203 (data from ~164 plots used to train the models). The lower accuracy at Site 2 can be attributed to small data size (data from ~103 plots used to train the models). Mkhabela et al. (2011) found that the R^2 between NDVI (extracted from satellite MODIS) and field pea grain yield on the Canadian Prairies was ranging between 0.53 and 0.89, depending on the agro-climatic zone.

When we enhanced the spatial resolution of satellite imagery through pan-sharpening techniques, the accuracy of seed yield prediction increased at both sites compared to the results obtained from models trained with satellite based-features at original resolution. However, it is important to note that it was challenging to separate the soil from the canopy, even with enhanced spatial resolution. Thus, the satellite may be a better option after canopy closure than early growth stages. In future work, we can also explore better approaches to segment the pan-sharpened satellite imagery, such as spectral un-mixing with the assistance of UAS imagery, as reported in Alvarez-Vanhard et al. (2020).

The assessment of feature importance revealed that not all features showed consistency across time points and field pea breeding trials. Sankaran et al. (2021) reported similar results with different features selected for maize yield prediction using UAS and satellite imagery. This change in feature selection between time points can be associated with changes in canopy structure and the resulting changes in reflectance properties. For example, at advanced growth stages, some pea lines develop a taller and denser canopy and, as a result, become more susceptible to lodging, creating more shadowed areas or intertwining with adjacent plots. Many studies have reported that the association of VIs with yield depends on the growth stage (Kyratzis et al., 2017; Fu et al., 2020; Adak et al., 2021; Sankaran et al., 2021; Shafiee et al., 2021). This finding on the importance of the growth stage was further demonstrated by our experiment while investigating the feature fusion approach from multi-scale sensing sources. Even if the integrated features were extracted from different spatial resolution datasets, adding satellite-based features at later growth stages provided more information to capture the change of crop growth dynamics compared to single time point and growth stage models. Fu et al. (2020) and Adak et al. (2021) also found that temporal phenotyping using UAS-based features was more accurate in estimating wheat grain yield.

These insights, in turn, may help to scale-up field phenomics applications. Therefore, to enhance field pea breeding with remote sensing-assisted procedures, high-resolution satellite and UAS can be used separately to derive spectral features associated with yield performance at critical growth stages (flowering and pod filling) or integrated (feature fusion) to provide additional temporal features.

6 Summary

High-resolution satellite and UAS-based-multispectral features were evaluated to estimate seed yield using a random forest model in different breeding lines from two diverse (advanced yield and variety testing) trials. Satellite and UAS image features and image fusion approaches were explored in this study. The major potential implications from the study can be described as: (i) High-resolution satellite imagery can be used to estimate seed yield at breeding plot level depending on the growth stage (after canopy closure). (ii) Multi-time points data fusion can be explored to capture crop growth patterns with the temporal features. (iii) And pan-sharpening (multi-source image fusion) is another tool to improve satellite spatial resolution, which could help plant breeders to study historical performance with archived satellite imagery and/or explore satellite hyperspectral imagery (hyper-sharpening) for field-based 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

Conceptualization: AM and SS; Data collection: AM, RM, and SV; Data processing and analysis: AM; Resources: SS, RM, and SV; Writing- Original Draft Preparation: AM; Writing – Review & Editing: SS, RM, and SV; Visualization: AM; Supervision: SS; Project Administration: SS, RM, and SV; Funding Acquisition: SS, RM, and SV. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by the US Department of Agriculture-National Institute of Food and Agriculture (USDA-NIFA) competitive projects (accession numbers 1011741, 1022033, 1028108) and hatch project (accession number 1014919). We would like to also thank Microsoft AI for Earth Grant for opportunities of cloud computing.

Acknowledgments

We greatly appreciate the assistance from Dr. Kumar Navulur and Mr. Mark Andel from Maxar Technologies for providing the satellite imageries for our study. We extend our sincere thanks to Mr. Jarrod Pfaff and Mr. Jeff Haines for the Site 1 ground truth data, Ms. Sarah Hallyburton for the Site 2 ground truth data, and Dr. Chongyuan Zhang and Mr. Worasit Sangjan for their support during the UAS data collection.

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.2023.1111575/full#supplementary-material

References

Adak, A., Murray, S. C., Božinović, S., Lindsey, R., Nakasagga, S., Chatterjee, S., et al. (2021). Temporal vegetation indices and plant height from remotely sensed imagery can predict grain yield and flowering time breeding value in maize *via* machine learning regression. *Remote Sens.* 13, 2141. doi: 10.3390/RS13112141/S1

Alvarez-Vanhard, E., Houet, T., Mony, C., Lecoq, L., and Corpetti, T. (2020). Can UAVs fill the gap between *in situ* surveys and satellites for habitat mapping? *Remote Sens. Environ.* 243, 111780. doi: 10.1016/J.RSE.2020.111780

Alvarez-Vanhard, E., Thomas, C., and Thomas, H. (2021). UAV & satellite synergies for optical remote sensing applications: A literature review. *Sci. Remote Sens.* 3, 100019. doi: 10.1016/j.srs.2021.100019ï

Al-Wassai, F. A., Kalyankar, N. V., and Al-Zuky, A. A. (2011). The IHS transformations based image fusion. arXiv preprint doi: arXiv:1107.4396

Borra-Serrano, I., Peña, J. M., Torres-Sánchez, J., Mesas-Carrascosa, F. J., and López-Granados, F. (2015). Spatial quality evaluation of resampled unmanned aerial vehicle-imagery for weed mapping. *Sens* 5 (8), 19688–19708. doi: 10.3390/s150819688

Buchaillot, M. L., Gracia-Romero, A., Vergara-Diaz, O., Zaman-Allah, M. A., Tarekegne, A., Cairns, J. E., et al. (2019). Evaluating maize genotype performance under low nitrogen conditions using RGB UAV phenotyping techniques. *Sensors* 19, 1815. doi: 10.3390/S19081815

Dadrass Javan, F., Samadzadegan, F., Mehravar, S., Toosi, A., Khatami, R., and Stein, A. (2021). A review of image fusion techniques for pan-sharpening of high-resolution satellite imagery. *ISPRS J. Photogramm. Remote Sens.* 171, 101–117. doi: 10.1016/j.isprsjprs.2020.11.001

Dalla Marta, A., Grifoni, D., Mancini, M., Orlando, F., Guasconi, F., and Orlandini, S. (2015). Durum wheat in-field monitoring and early-yield prediction: Assessment of potential use of high resolution satellite imagery in a hilly area of Tuscany, central Italy. *J. Agric. Sci.* 153, 68–77. doi: 10.1017/S0021859613000877

Ehlers, M., Klonus, S., Johan Åstrand, P., and Rosso, P. (2010). International journal of image and data fusion multi-sensor image fusion for pansharpening in remote sensing multi-sensor image fusion for pansharpening in remote sensing. *Int. J. Image Data Fusion* 1, 25–45. doi: 10.1080/19479830903561985

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

Gašparović, M., Rumora, L., Miler, M., and Medak, D. (2019). Effect of fusing sentinel-2 and WorldView-4 imagery on the various vegetation indices. *J. Appl. Remote Sens.* 13, 036503. doi: 10.1117/1.JRS.13.036503

Ghamisi, P., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., et al. (2019). Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. *IEEE Geosci. Remote Sens. Mag.* 7, 6–39. doi: 10.1109/MGRS.2018.2890023

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 Phenom.* doi: 10.34133/2021/9840192

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

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

Jin, X., Zarco-Tejada, P. J., Schmidhalter, U., Reynolds, M. P., Hawkesford, M. J., Varshney, R. K., et al. (2021). High-throughput estimation of crop traits: A review of ground and aerial phenotyping platforms. *IEEE Geosci. Remote Sens. Mag.* 9, 200–231. doi: 10.1109/MGRS.2020.2998816

Johnson, B. A., Scheyvens, H., and Shivakoti, B. R. (2014). An ensemble pansharpening approach for finer-scale mapping of sugarcane with landsat 8 imagery. *Int. J. Appl. Earth Obs. Geoinf.* 33, 218–225. doi: 10.1016/j.jag.2014.06.003

Kholová, J., Urban, M. O., Cock, J., Arcos, J., Arnaud, E., Aytekin, D., et al. (2021). In pursuit of a better world: crop improvement and the CGIAR. *J. Exp. Bot.* 72 (14), 5158–5179. doi: 10.1093/jxb/erab226

Kuhn, M. (2008). Building predictive models in r using the caret package. *J. Stat. Softw.* 28 (5), 1–26. doi: 10.18637/jss.v028.i05

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. doi: 10.3389/FPLS.2017.01114/BIBTEX

Li, Y., Li, J., He, L., Chen, J., and Plaza, A. (2020). A new sensor bias-driven spatio-temporal fusion model based on convolutional neural networks. *Sci. China Inform. Sci.* 63 (4), 1–16. doi: 10.1007/s11432-019-2805-y

Li, J., Veeranampalayam-Sivakumar, A. N., Bhatta, M., Garst, N. D., Stoll, H., Stephen Baenziger, P., et al. (2019). Principal variable selection to explain grain yield variation in winter wheat from features extracted from UAV imagery. *Plant Methods* 15:1–13. doi: 10.1186/s13007-019-0508-7

Luo, X., Zhou, L., Shi, X., Wan, H., and Yin, J. (2018). Spectral mapping based on panchromatic block structure analysis. *IEEE Access 6*, 40354–40364. doi: 10.1109/ACCESS.2018.2853150

Maimaitijiang, M., Sagan, V., Sidike, P., Daloye, A. M., Erkbol, H., and Fritschi, F. B. (2020a). Crop monitoring using satellite/UAV data fusion and machine learning. *Remote Sens* 12:1357. doi: 10.3390/RS12091357

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., and Fritschi, F. B. (2020b). Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sens. Environ.* 237, 111599. doi: 10.1016/J.RSE.2019.111599

Malamiri, H. R. G., Aliabad, F. A., Shojaei, S., Morad, M., and Band, S. S. (2021). A study on the use of UAV images to improve the separation accuracy of agricultural land areas. *Comput. Electron. Agric.* 184, 106079. doi: 10.1016/J.COMPAG.2021.106079

Marzougui, A., Ma, Y., Zhang, C., McGee, R. J., Coyne, C. J., Main, D., et al. (2019). Advanced imaging for quantitative evaluation of aphanomyces root rot resistance in lentil. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00383

MathWorks. (2021). MATLAB R2021b (Version 9.11) [Software]. Available at: https://www.mathworks.com/products/matlab.html (Accessed on 15 February 2023).

Mkhabela, M. S., Bullock, P., Raj, S., Wang, S., and Yang, Y. (2011). Crop yield forecasting on the Canadian prairies using MODIS NDVI data. *Agric. For. Meteorol.* 151 (3), 385–393. doi: 10.1016/j.agrformet.2010.11.012

Nhamo, L., Van Dijk, R., Magidi, J., Wiberg, D., and Tshikolomo, K. (2018). Improving the accuracy of remotely sensed irrigated areas using post-classification enhancement through UAV capability. *Remote Sens.* 10 (5), 712. doi: 10.3390/rs10050712

Pieruschka, R., and Schurr, U. (2019). Plant phenotyping: Past, present, and future. *Plant Phenom.* 2019, 1–6. doi: 10.34133/2019/7507131

Pratap, A., Gupta, S., Nair, R. M., Gupta, S. K., Schafleitner, R., Basu, P. S., et al. (2019). Using plant phenomics to exploit the gains of genomics. *Agronomy* 9, 1–25. doi: 10.3390/agronomy9030126

QGIS Development Team. (2021). Development Team. (2021). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available at: https://qgis.org/en/site/forusers/download.html (Accessed on 15 February 2023).

Roth, L., Barendregt, C., Bétrix, C. A., Hund, A., and Walter, A. (2022). High-throughput field phenotyping of soybean: Spotting an ideotype. *Remote Sens. Environ.* 269, 112797. doi: 10.1016/J.RSE.2021.112797

Sangjan, W., Carter, A. H., Pumphrey, M. O., Jitkov, V., and Sankaran, S. (2021). Development of a raspberry pi-based sensor system for automated in-field monitoring to support crop breeding programs. *Invent* 6 (2), 42. doi: 10.3390/inventions6020042

Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., et al. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *Eur. J. Agron.* 70, 112–123. doi: 10.1016/J.EJA.2015.07.004

Sankaran, S., Marzougui, A., Hurst, J. P., Zhang, C., Schnable, J. C., and Shi, Y. (2021). Can high-resolution satellite multispectral imagery be used to phenotype canopy traits and yield potential in field conditions? *Trans. ASABE* 64, 879–891. doi: 10.13031/trans

Sankaran, S., Zhou, J., Khot, L. R., Trapp, J. J., Mndolwa, E., and Miklas, P. N. (2018). High-throughput field phenotyping in dry bean using small unmanned aerial vehicle based multispectral imagery. *Comput. Electron. Agric.* 151, 84–92. doi: 10.1016/j.compag.2018.05.034

Santana, D. C., Paulo De Oliveira Cunha, M., Garcia Dos Santos, R., Cotrim, M. F., Pereira, L., Teodoro, R., et al. (2022). High-throughput phenotyping allows the selection of soybean genotypes for earliness and high grain yield. *Plant Methods* 18, 1–11. doi: 10.1186/S13007-022-00848-4

Selvaraj, M. G., Vergara, A., Montenegro, F., Ruiz, H. A., Safari, N., Raymaekers, D., et al. (2020). Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and republic of Benin. *ISPRS J. Photogramm. Remote Sens.* 169, 110–124. doi: 10.1016/j.isprsjprs.2020.08.025

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

Siok, K., Ewiak, I., and Jenerowicz, A. (2020). Multi-sensor fusion: A simulation approach to pansharpening aerial and satellite images. *Sens* 20 (24), 7100. doi: 10.3390/s20247100

Smith, D. T., Potgieter, A. B., and Chapman, S. C. (2021). Scaling up high-throughput phenotyping for abiotic stress selection in the field. *Theor. Appl. Genet.* 134 (6), 1845–1866. doi: 10.1007/S00122-021-03864-5

Tattaris, M., Reynolds, M. P., and Chapman, S. C. (2016). A direct comparison of remote sensing approaches for high-throughput phenotyping in plant breeding. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01131

Tauro, F., Maltese, A., Giannini, R., and Harfouche, A. (2022). Latent heat flux variability and response to drought stress of black poplar: A multi-platform multi-sensor remote and proximal sensing approach to relieve the data scarcity bottleneck. *Remote Sens. Environ.* 268, 112771. doi: 10.1016/j.rse.2021.112771

Tefera, A. T., Banerjee, B. P., Pandey, B. R., James, L., Puri, R. R., Cooray, O. O., et al. (2022). Estimating early season growth and biomass of field pea for selection of divergent ideotypes using proximal sensing. *F. Crop Res.* 277, 108407. doi: 10.1016/J.FCR.2021.108407

Tracy, S. R., Nagel, K. A., Postma, J. A., Fassbender, H., Wasson, A., and Watt, M. (2020). Crop improvement from phenotyping roots: Highlights reveal expanding opportunities. *Trends Plant Sci.* 25, 105–118. doi: 10.1016/j.tplants.2019.10.015

Valencia-Ortiz, M., Sangjan, W., Selvaraj, M. G., McGee, R. J., and Sankaran, S. (2021). Effect of the solar zenith angles at different latitudes on estimated crop vegetation indices. *Drones* 5, 80. doi: 10.3390/DRONES5030080

Vargas, J. J. Q., Zhang, C., Smitchger, J. A., McGee, R. J., and Sankaran, S. (2019). Phenotyping of plant biomass and performance traits using remote sensing techniques in pea (*Pisum sativum*, l.). *Sens. (Switzerland)* 19:2031. doi: 10.3390/s19092031

Varshney, R. K., Bohra, A., Yu, J., Graner, A., Zhang, Q., and Sorrells, M. E. (2021). Designing future crops: genomics-assisted breeding comes of age. *Trends Plant Sci.* 26, 631–649. doi: 10.1016/J.TPLANTS.2021.03.010

Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G. A., et al. (2015). A critical comparison among pansharpening algorithms. *IEEE Trans. Geosci. Remote Sens.* 53, 2565–2586. doi: 10.1109/TGRS.2014.2361734

Vivone, G., Alparone, L., Garzelli, A., and Lolli, S. (2019). Fast reproducible pansharpening based on instrument and acquisition modeling: AWLP revisited. *Remote Sens.* 11, 2315. doi: 10.3390/RS11192315

Vivone, G., Dalla Mura, M., Garzelli, A., Restaino, R., Scarpa, G., Ulfarsson, M. O., et al. (2021). A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods. *IEEE Geosci. Remote Sens. Mag.* 9, 53–81. doi: 10.1109/MGRS.2020.3019315

Wald, L., Ranchin, T., and Mangolini, M. (1997). Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogramm. Eng. Remote Sens.* 63 (6), 691–699.

Watt, M., Fiorani, F., Usadel, B., Rascher, U., Muller, O., and Schurr, U. (2020). Phenotyping: New windows into the plant for breeders. *Annu. Rev. Plant Biol.* 71, 689–712. doi: 10.1146/annurev-arplant-042916-041124

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, 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. doi: 10.3389/fpls.2017.01111

Yilmaz, C. S., Yilmaz, V., Gungor, O., and Shan, J. (2019)Metaheuristic pansharpening based on symbiotic organisms search optimization (Accessed 2.22.22).

Yokoya, N., Grohnfeldt, C., and Chanussot, J. (2017). Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geosci. Remote Sens. Mag.* 5, 29–56. doi: 10.1109/MGRS.2016.2637824

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, C., Marzougui, A., and Sankaran, S. (2020). High-resolution satellite imagery applications in crop phenotyping: An overview. *Comput. Electron. Agric.* 175, 105584. doi: 10.1016/j.compag.2020.105584

Zhang, C., McGee, R. J., Vandemark, G. J., and Sankaran, S. (2021). Crop performance evaluation of chickpea and dry pea breeding lines across seasons and locations using phenomics data. *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.640259/BIBTEX

Zhang, X., Zhao, J., Yang, G., Liu, J., Cao, J., Li, C., et al. (2019). Establishment of plot-yield prediction models in soybean breeding programs using UAV-based hyperspectral remote sensing. *Remote Sens.* 11, 2752. doi: 10.3390/RS11232752

Zhao, L., Shi, Y., Liu, B., Hovis, C., Duan, Y., and Shi, Z. (2019b). Finer classification of crops by fusing UAV images and sentinel-2A data. *Remote Sens.* 11 (24), 3012. doi: 10.3390/rs11243012

Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., et al. (2019a). Crop phenomics: Current status and perspectives. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00714

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 13 March 2023
DOI 10.3389/fpls.2023.1143525

Check for updates

OPEN ACCESS

EDITED BY
Bo Li,
Syngenta, United Kingdom

REVIEWED BY
Fei Shen,
Beijing Academy of Agricultural and Forestry Sciences, China
Jun He,
Southwest University, China

*CORRESPONDENCE
Hui Xia
xiahui@sicau.edu.cn
Dong Liang
liangeast@sicau.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

RECEIVED 13 January 2023
ACCEPTED 27 February 2023
PUBLISHED 13 March 2023

CITATION
Deng H, Li Y, Pang C, Zhang K, Tian X, Wang T, Liang Y, He Z, Lang Y, Fang J, Lin L, Wang J, Lv X, Xia H and Liang D (2023) Significant increases in Donghong kiwifruit yield by a novel umbrella-shaped trellis system and identification of associated molecular mechanisms. *Front. Plant Sci.* 14:1143525. doi: 10.3389/fpls.2023.1143525

COPYRIGHT
© 2023 Deng, Li, Pang, Zhang, Tian, Wang, Liang, He, Lang, Fang, Lin, Wang, Lv, Xia and Liang. 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.

Significant increases in Donghong kiwifruit yield by a novel umbrella-shaped trellis system and identification of associated molecular mechanisms

Honghong Deng[1†], Yao Li[1†], Changqing Pang[1], Kun Zhang[1], Xinbo Tian[1], Tong Wang[1], Yan Liang[1], Zunzhen He[1], Yuxuan Lang[1], Jinbao Fang[2], Lijin Lin[1], Jin Wang[1], Xiulan Lv[1], Hui Xia[1*] and Dong Liang[1*]

[1]Institute of Pomology and Olericulture, College of Horticulture, Sichuan Agricultural University, Chengdu, China, [2]Zhengzhou Fruit Research Institute, Chinese Academy of Agricultural Sciences, Zhengzhou, China

China is the largest kiwifruit producer in the world, accounting for more than half of the total. However, in terms of yield per unit area, China is much lower than the global average and lags behind that of other countries. Yield improvement is of critical importance for the current kiwifruit industry in China. In this study, an improved overhead pergolas trellis (OPT) system, namely, the umbrella-shaped trellis (UST) system, was developed for Donghong kiwifruit, which is now the second most popular and widely cultivated red-fleshed kiwifruit in China. Surprisingly, the estimated yield on the UST system was more than two times higher than that with a traditional OPT, while the external fruit quality was maintained and the internal fruit quality was improved. One of the mechanisms contributing to the yield improvement was the significant promotion of the vegetative growth of canes at 6 ~ 10 mm in diameter by the UST system. The upper canopy of the UST treatment served as a natural shading condition for the lower fruiting canopy and thus had positive effects on the accumulation of chlorophylls and total carotenoids in the fruiting canopy. The most productive zones on the fruiting canes (6 ~ 10 mm in diameter) contained significantly higher ($P < 0.05$) levels of zeatin riboside (ZR) and auxin (IAA) and ratios of ZR/gibberellin (GA), ZR/abscisic acid (ABA), and ABA/GA. A relatively high carbon/nitrogen ratio may promote the flower bud differentiation process of Donghong kiwifruit. The outcomes of this study provide a scientific basis for manifold increase in production of kiwifruit and contribute to the sustainability of the kiwifruit industry.

KEYWORDS

budbreak, carbon to nitrogen ratio, endogenous hormones, fruiting cane, gas exchange parameters

Introduction

Kiwi fruit is the berry of a woody deciduous climbing vine species belonging to the genus *Actinidia* Lindl. of the family Actinidiaceae (Ferguson, 2016). It has gained enormous popularity in recent decades due to its exceptionally high levels of vitamin C and a balanced nutritional profile of carbohydrates, proteins, lipids, minerals, vitamins, organic acids, chlorophylls, phenolic compounds, and carotenoids (Wang et al., 2021). Because of its nutritional qualities, kiwifruit has been considered as the "king of fruits" for centuries (Huang et al., 2013).

Kiwifruit is believed to have originated in the Yangzi and Pearl Rivers and mountainous ranges in China (Huang et al., 2013). Historical records of kiwifruit vine cultivars trace back to the 7th century A.D. during the Tang Dynasty. In the early 20th century, kiwifruit cultivation spread from China to New Zealand, where the commercial planting of this fruit was initiated (Huang and Ferguson, 2007). Its commercial cultivation has since spread to other significant producing countries. China is now the world's largest kiwifruit producer and presented an annual production of approximately 2.20 million tons in 2019, followed by New Zealand (0.56 million tons), Italy (0.52 million tons), Iran (0.34 million tons), and Greece (0.29 million tons), with these countries accounting for 50.52, 12.84, 12.06, 7.92, and 6.57% of the global annual production (4.35 million tons), respectively. However, in terms of yield per unit area, China (120,325 hg/ha) is much lower than the world average (161,764 hg/ha) and ranks at approximately the 20th among the kiwifruit producing countries, and it lags behind that of New Zealand (374,073 hg/ha) and Greece (277,804 hg/ha) (FAOSTAT, 2019). This phenomenon has substantially restricted the economic output. Thus, improving the yield per unit area is of critical importance for the current kiwifruit industry in China (Fang and Zhong, 2019; Qi et al., 2020).

Yield formation can be controlled by the integration of various factors, such as genotypes, environmental conditions, and cultural practices. Kiwifruit vines are best trained upward on trellises because they can become unruly within one season and require sturdy support to protect the vines from tangling and shadowing one another. The tangled and shaded vines yield less fruit. T-bar and overhead pergola structures are two main types of support trellis systems commonly used in commercial kiwifruit production (Costa and Biasi, 1991). T-bar trellis is easier to build, less labor-intensive, and better for bee pollination, and it reduces the risk of contracting Botrytis compared with the pergola trellis (Strik and Cahn, 2000). However, the pergola trellis may increase productivity and reduce the vulnerability of resulting fruit to wind damage. Additionally, once the complete pergola canopy is in place, the shade slows the growth of weeds (Hartmann, 2018).

Kiwifruit has become a mainstream fruit crop and advantageous characteristic industry in Sichuan Province, where the traditional overhead pergolas trellis (OPT) is used. In our previous study, we developed an improved overhead pergolas trellis, the umbrella-shaped trellis (UST) system, for Sichuan kiwifruit production. Surprisingly, UST has been found to be effective at enhancing kiwifruit yield based on more than 5 years of observations by local growers. However, the mechanisms underlying the high yield of the UST are still unclear.

To study the mechanisms of high yield of the UST, a two-consecutive-years comparative study of the effects of the UST and OPT on kiwifruit yield was performed in this work. The newly developed cultivar, 'Donghong' (DH) kiwifruit (*A. chinensis*), which descended from the F1 offspring of open-pollinated 'Hongyang' (*A. chinensis*) kiwifruit (Zhong et al., 2016), was selected as experimental material. DH is now the second most popular and widely cultivated red-fleshed kiwifruit in China (Huang et al., 2022). To understand mechanisms underlying the high yield of the UST, the shoot growth, photosynthetic capacity, flower bud differentiation, and fruit quality and yield were evaluated. Elucidation of the regulatory mechanism will provide a scientific basis for the efficient increase in kiwifruit production.

Materials and methods

Plant materials and experimental site

An experimental trial was conducted over two consecutive years in 2020 and 2021 at a commercial kiwifruit orchard of the Xian Nong Fen Xiang Organic Agriculture Development Co. Ltd., which is located at Pujiang County (30°27′N, 103°42′E) of Sichuan Province, China. The climate was a subtropical monsoon climate with a wild summer and mild winter, and annual average rainfall, sunshine and temperature of 1,117.3 mm, 1,107.9 h and 16.3°C, respectively. Five-year-old plants (grafted onto wild *A. deliciosa*) bearing kiwifruit with uniform health, size, and vigor were used. The planting density was 667 vines/ha, with 3.0 m spacing between plants and 6.0 m spacing between rows. The row orientation was from north to south. All plants were subjected to identical standard orchard management practices for DH kiwifruit in this area throughout the experiment, including winter pruning, pest and pathogen control, basal fertilization, and irrigation.

Experimental design and treatment

The OPT was constructed with a number of vertical posts with a horizontal beam attached at the top that runs continuously between and perpendicular to the rows (Figure 1A). As an improved OPT system, the UST included a bamboo pole stands at the top center of four vertical posts with wires attached to the center wires of the OPT and the top of the bamboo pole (Figure 1B). The detailed parameters are shown in the schematic diagrams (Figures 1A, B) drawn using SketchUp software (https://www.sketchup.com). Their phenotypic characteristics in the field are shown in Figures 1C–F. The trees were trained with a straight vertical trunk approximately 1.5 m high and two central leader vines. The horizontal central leader produced the

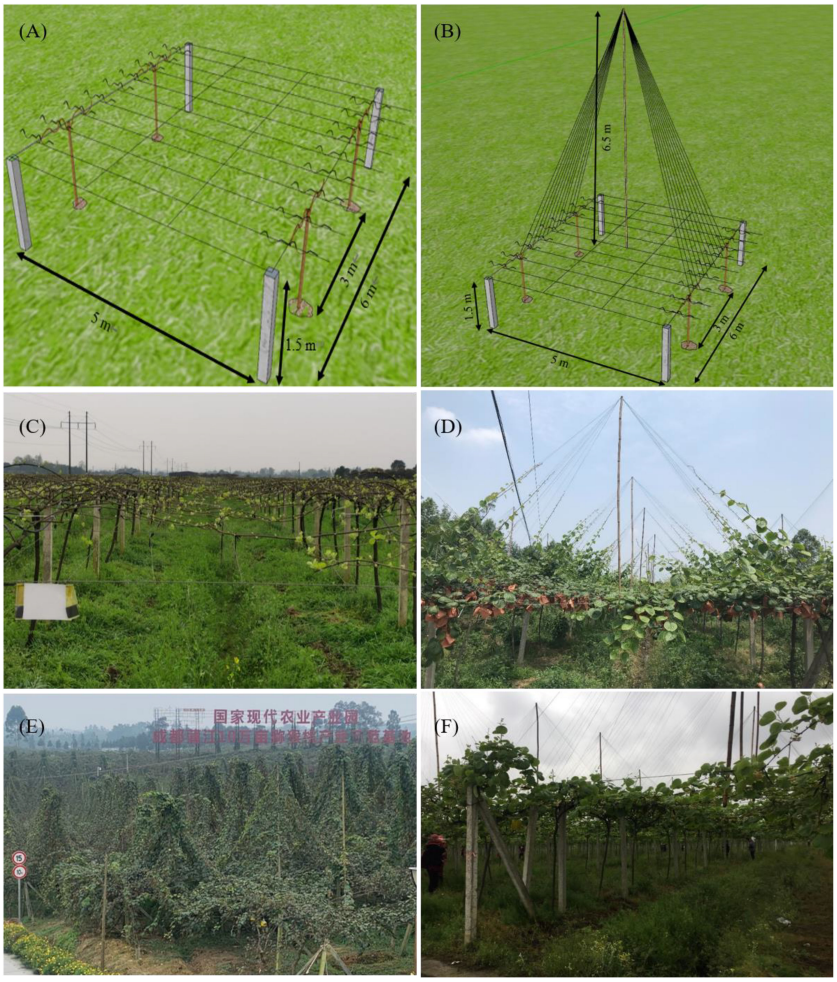


FIGURE 1
Schematic representations and field phenotypic characteristics of the traditional overhead pergolas trellis (OPT) and umbrella-shaped trellis (UST). **(A)** Schematic diagram illustrating the cane configuration (tree architecture) of the OPT, **(B)** schematic diagram illustrating cane configuration (tree architecture) of the UST, **(C)** field performance of the OPT, and **(D–F)** field performances of the UST.

fruiting framework, which was made up primarily of one-year-old canes (Figure 1C). The vegetative canes (lateral branches) ran along the wires to the top, thus forming a canopy that resembled an "umbrella". The fruited canes during the previous season as well as any dead, diseased, or twisted canes were pruned during winter, while those canes on the "umbrella" canopy were pulled down to horizontal in the following year to continue growing into flowering and fruiting canes (Figures 1D–F). Each treatment had three biological replicates, which each consisted of at least ten canes that were randomly chosen and tagged. Each flower on the tagged canes received supplementary pollination during blooming by being dusted with commercial bee-collected male pollen.

Chemicals and reagents

High performance liquid chromatography (HPLC)-grade regents, including zeatin riboside (ZR), auxin (IAA), abscisic acid (ABA), and gibberellin (GA), were purchased from Sigma–Aldrich (St Louis, MO, USA). Aqueous solutions were prepared using ultra-purified water (18.2 MΩ cm) from a Milli-Q gradient water purification system (Millipore Corporation, Bedford, MA).

Determination of cane length and diameter, bud number, and percentages of budbreak and flower bud

The length of current-year canes with diameters of 6 ~ 8 mm (slender), 8 ~ 10 mm (medium), > 10 mm (thick) were recorded on June 28, July 11, July 24, August 10, and August 24, 2020, while those with diameters less than 6 mm were not measured because they were heavily curled and too thin. The cane thickness was measured using calipers. The numbers of buds, budbreak, and flower buds on one-year-old slender, medium, and thick canes were counted at the flowering season (April 9, 2021). The corresponding proportion of budbreak and flowering shoots were calculated.

Determination of leaf gas exchange parameters and photosynthesis pigment contents

The gas-exchange parameters included the net photosynthetic rate (P_n), stomatal conductance (*Gs*), intercellular CO_2 concentration (*Ci*), and transpiration rate (*Tr*). The measurements performed on the completely expanded leaves of the third to fifth node of canes on the UST upper canopy, fruiting canopy, and OPT fruiting canopy at the same time (09:00 am ~ 11:00 am at the end of June and the beginning of July of 2021) using a Li-6400XT portable photosynthesis system (Li-COR Inc., Lincoln, NE, USA). The upper canopy doesn't bear fruit in UST system. The photosynthetic pigments, including chlorophyll a (Chl a), chlorophyll b (Chl b), and total carotenoids (TC), were extracted using fresh mature leaves (30 mg) in 95% ethanol (v/v, 3 ml). The photosynthetic pigment contents were determined using equations developed in a previous report (Molnárová and Fargašová, 2009).

Determination of endogenous hormones using HPLC coupled with ultraviolet (UV) detection

Extraction and quantification of the endogenous hormones, including ZR, IAA, ABA, and GA, in the buds were performed using HPLC coupled with UV detection following our previously published protocol (Liang et al., 2019) with minor modifications. Briefly, samples were ground to a fine powder in liquid nitrogen. Then, 300 mg of powder was transferred into 2.0 mL conical-bottom plastic tubes and 1.5 mL of ice-cold extraction solvent containing 80% aqueous methanol and 0.5% formic acid was added. The samples were extracted by sonication for 5 min in an ultrasonic bath (Branson Ultrasonic Corporation, Connecticut, USA), cooled with ice water, and then agitated for 15 min at 4°C in a refrigerator using a rotator at 50 rpm. The homogenates were centrifuged (12,000 rpm, 10 min, 4°C) and the resultant supernatant was carefully collected into a clean amber vial. This extraction procedure was performed in triplicate. Thereafter, the combined supernatants were evaporated to water phase under vacuum at 37°C using an N-EVAP analytical evaporator (Organomation Associates Inc., South Berlin, MA, USA). The residue was dissolved in 200 µL of 30% (v/v) acetonitrile. Each extract was purified *via* a 0.22 µm porosity polytetrafluoroethylene membrane prior to injection in the HPLC system.

The HPLC system consisted of an Agilent 1260 HPLC instrument and a diode array detector (Agilent, Santa Clara, CA, USA). Chromatographic separation was achieved using a ZORBAX StableBond C_{18} column (inner diameter × length: 4.6 mm × 150 mm; particle size: 5 µm) (Agilent). Water containing 45% methanol and 0.6% acetic acid was used as the mobile phase, with a flow velocity of 1.0 mL min^{-1}. The column temperature over was maintained at 35°C, and the injection volume was 20 µL. Peaks were detected using a UV-visible detector at 270, 218, 200, and 270 nm for ZR, IAA, ABA, and GA_3, respectively. The peak area of each endogenous hormone was quantified by comparing with the corresponding peak area of authentic standard containing known concentration.

Determination of carbon (C) to nitrogen (N) ratio of bud

The anthrone colorimetric method was employed to measure the content of the soluble sugar and starch (Dubois et al., 1956). A Braford assay utilizing Coomassie brilliant blue G-250 was used to quantify the content of soluble protein (Bradford, 1976). The C/N ratio was defined as the sum contents of soluble sugar and starch to the sum content of soluble protein.

Determination of the physicochemical parameters of fruit quality and yield

One hundred fruit samples from each treatment were randomly harvested at the end of August, which corresponds to the commercial ripening stage based on the total solid soluble (TSS) content > 7.5°Brix. These samples were randomly divided into five biological replicates. The single fruit weight was obtained from an average of 20 fruits of each biological replicate. The fruit shape index was calculated by dividing the fruit vertical length with transverse diameter. Dry weight was defined as the weight recorded after drying fruity slices at 60°C. TSS was determined using a digital refractometer (Model PAL1 0~53%; Atago Co., Ltd., Tokyo, Japan), and the corresponding result was expressed as °Brix. The soluble sugar content (SSC) was determined using the anthrone colorimetric method. Titratable acidity (TA, also called total acidity) was determined by standard acid-base titration methods following (Horwitz and Latimer, 2006), and the result was expressed as percentage of citric acid.

Data processing and statistical analysis

All data are shown as the means ± SD of at least three replicates. One-way analysis of variance (ANOVA) and Tukey's honestly significant difference (HSD) test ($P < 0.05$) were applied to examine the statistical significance of responses of kiwifruit to the OPT and UST systems. All statistical analyses were conducted with Minitab software v19.0 (Minitab, State College, PA, USA).

Results

Effect of the OPT and UST on the yield and quality of DH kiwifruit

The schematic representations and field phenotypic characteristics of the traditional OPT and UST systems were shown in Figure 1. The effects of the OPT and UST systems on DH kiwifruit yield and its

components (fruit weight, fruit number per fruiting cane, and fruiting cane number per plant) were evaluated (Table 1). The estimated yield (3528.6 kg/667 m^2) in the UST-treated group was 264.81% greater than that in the OPT control group (1332.5 kg/667 m^2). The single kiwifruit weight or fruiting cane number per plant was not affected ($P > 0.05$) by the treatments. The average kiwifruit number per fruiting cane in the OPT control group was 36.7 ± 1.15, while it significantly increased ($P < 0.05$) by 70.03% in the UST-treated group (62.4 ± 6.80). Changes in the estimated yield paralleled those of the average kiwifruit number per fruiting cane, although even greater differences were observed due to the minor increment in fruiting cane number per plant under the UST treatment (Table 1).

The extrinsic quality (fruit shape index) was similar ($P > 0.05$) in DH kiwifruit from OPT- (control) and UST-treated trees at both the harvest and ripening stages. At harvest, DH kiwifruit grown on the UST had significantly lower ($P < 0.05$) TA content (1.36 ± 0.03c g 100 g^{-1}) than that in the control treatment (1.58 ± 0.02b g 100 g^{-1}). The increases in TSS and SSC observed in all samples were due to fruit ripening after storage, although significant differences were not observed ($P > 0.05$) between those of the OPT- and UST-treated DH kiwifruit at both the harvest and ripening stages. The TSS/TA levels increased after storage in all treatments. At the ripening stage, the UST-treated DH kiwifruit showed significantly higher ($P < 0.05$) TSS/TA (7.54 ± 0.06a) compared to the control (6.40 ± 0.01b). Dry matter slightly decreased in all groups during storage, although significant differences were not observed between the different samples (Table 2).

Effect of OPT and UST on cane and bud growth of DH kiwifruit

The total one-year-old cane length of the UST-treated trees during the whole period was significantly higher ($P < 0.05$) than that of the OPT control trees. Similar trends in the slender and medium canes were observed, and the maximum values of 145.11 ± 2.17 cm and 104.69 ± 3.55 cm were found on June 28 and August 24 in the UST-treated trees and controls, respectively, which accounted for 62% and 38% of their total one-year-old cane length, respectively. However, an opposite trend was observed in the thick canes throughout the whole period, with a significantly lower ($P < 0.05$) maximum value (52.35 ± 1.57 mm, 19%, recorded on August 24) of the UST-treated trees compared with the control (115.08 ± 2.72 mm, 56%) (Table 3).

The bud number per meter on canes with different diameters of the UST-treated group was significantly higher ($P < 0.05$) than that of the control. The highest amounts were recorded at August 24 (18.29 ± 1.92), July 24 (19.95 ± 2.07), and August 10 (18.53 ± 1.65) in the slender, medium, and thick canes of the UST-treated trees, respectively, which were 1.25, 1.29, and 1.21 times higher than those of the OPT control group (Table 4).

UST-treated trees had significantly higher ($P < 0.05$) total lengths of slender canes (156.33 ± 2.31 mm) and medium canes (115.72 ± 2.69 mm), bud number per fruiting cane (60.4), budbreak number (31.83), and flower bud number (31.32) on the fruiting cane, which were 2.39, 2.34, 1.78, 1.94 and 2.00 times higher than that of the OPT control group. No significant differences in these parameters for the thick canes were observed between trees treated with the OPT or UST system (Table 5).

Effect of the OPT and UST on the leaf gas exchange parameters and photosynthesis pigment content of DH kiwifruit

The UST system significantly ($P < 0.05$) increased the *Pn*, *Gs*, *Ci*, and *Tr* of the upper canopy by 64.25, 141.80, 23.47, and 51.93% compared with that observed for the OPT fruiting canopy. There was no significant ($P > 0.05$) change in the leaf gas exchange parameters of the UST fruiting canopy compared to the OPT fruiting canopy except for *Pn* (Figure 2).

As shown in Figure 3, the UST system enhanced the Chl a, Chl b, and TC contents of the fruiting canopy by 39.38, 48.11, and 44.63%, respectively, compared to those of the OPT fruiting canopy, while the Chl a, Chl b, and TC contents of the UST upper canopy was lower than those of the OPT fruiting canopy (Figure 3).

Effect of the UST on bud endogenous hormones of DH kiwifruit

Different endogenous hormones underwent a series of changes throughout the measurement period. As shown in Figure 4, the endogenous ZR and IAA contents showed a declining trend while the GA content increased over time. The ABA content changes were irregular and initially increased, peaking on July 11 before sharply decreasing and further increasing. The ZR and IAA contents of the slender and medium canes were significantly higher ($P < 0.05$) than those of the thick canes at most time points, whereas the GA and ABA contents of the slender and medium canes were significantly lower ($P < 0.05$) than those of the thick canes (Figure 4).

The variation tendencies of the ratios of different endogenous hormones are shown in Figure 5. Overall, ZA/GA, ZA/ABA, and IAA/

TABLE 1 Effects of the OPT and UST on the estimated yield of Donghong kiwifruit.

Treatment	Single fruit weight (g)	Fruit number per fruiting cane	Fruiting cane number per plant	Plant number per 667 m^2	Estimated yield (kg/667 m^2)
OPT	79.09 ± 1.47a	36.65 ± 1.15b	10.33 ± 0.57a	44.5	1332.5
UST	80.91 ± 1.52a	77.35 ± 6.80a	12.67 ± 0.58a	44.5	3528.6

Data were expressed as means ± SD. Different letters within each column indicate significant differences (P < 0.05).

TABLE 2 Extrinsic and intrinsic quality parameters in the OPT and UST-treated Donghong kiwifruit at harvest and after storage (ripening).

Parameter	Harvest		Ripening	
	OPT	UST	OPT	UST
fruit shape index	1.30 ± 0.07a	1.32 ± 0.09a	1.30 ± 0.07a	1.32 ± 0.09a
total soluble solids (°Brix)	7.23 ± 0.06b	7.57 ± 0.12b	13.25 ± 0.01a	13.73 ± 0.06a
soluble sugar content (%)	5.26 ± 0.17b	5.84 ± 0.15b	6.37 ± 0.10a	6.89 ± 0.07a
titratable acidity (%)	1.58 ± 0.02b	1.36 ± 0.03c	2.07 ± 0.06a	1.82 ± 0.12b
total soluble solids/titratable acidity	5.24 ± 0.03c	5.57 ± 0.08c	6.40 ± 0.01b	7.54 ± 0.06a
dry matter (%)	16.24 ± 0.004a	16.35 ± 0.002a	15.37 ± 0.006a	15.23 ± 0.001a

Data were expressed as means± SD. Different letters within each row indicate significant differences (P < 0.05).

IBA roughly decreased with a similar pattern and timing but to different extents, while ZR/IAA and ABA/GA initially increased and then decreased. Significantly higher ($P < 0.05$) levels of ZR/GA, ZR/ABA, ABA/GA, and IAA/ABA were found on the slender and medium canes compared to the thick canes. For ZR/IAA, the opposite behavior was observed, with thick canes containing significantly higher ($P < 0.05$) levels of ZR/IAA compared to the slender and medium canes (Figure 5).

Effect of the UST on the C/N ratio of DH kiwifruit

The C/N ratio showed decreasing trends similar to that of the soluble sugar, soluble carbon, and soluble nitrogen contents. The soluble starch content peaked transiently at the initial measurement period, declined rapidly to undetectable levels, then increased over the time course, and gradually decreased thereafter. On July 11, the C/N ratio of slender canes reached the maximum value of 0.74 at July 11, which was 1.38 and 1.80 times higher than that of the medium and thick canes, respectively (Figure 6).

Discussion

Kiwifruit has been renowned as the "king of fruits" due to its nutritional characteristics (Huang et al., 2013). The present global production of kiwifruit stands at 4.35 million tons, with China accounting for more than half of the total. However, in terms of yield per unit area, China is much lower than the global average,

TABLE 3 Effects of the OPT and UST on one-year-old cane length of Donghong kiwifruit.

Date	Treatment	Cane length (cm)				Ratio
		Total length	Slender canes with 6 ~ 8 mm in diameter	Medium canes with 8 ~ 10 mm in diameter	Thick canes with > 10 mm in diameter	
June 28	OPT	171.05 ± 4.18f	51.32 ± 1.35cd	51.32 ± 1.55c	68.42 ± 1.28c	30:30:40
	UST	245.75 ± 5.49c	145.11 ± 2.17a	88.94 ± 2.73b	11.70 ± 0.59g	59:36:05
July 11	OPT	194.73 ± 5.56e	59.50 ± 1.98c	59.50 ± 1.69c	75.73 ± 1.89bc	31:31:38
	UST	236.50 ± 4.80c	134.81 ± 2.33ab	80.41 ± 1.92b	21.29 ± 0.55f	57:34:09
July 24	OPT	182.43 ± 4.11e	49.26 ± 1.56d	47.43 ± 1.08d	85.74 ± 1.47b	27:26:47
	UST	246.30 ± 4.28b	135.47 ± 2.25ab	83.74 ± 1.52d	27.09 ± 0.51ef	55:34:11
August 10	OPT	203.40 ± 6.14d	46.78 ± 1.61de	48.84 ± 1.77d	107.80 ± 2.76a	23:24:53
	UST	251.77 ± 7.25b	118.33 ± 3.95b	103.23 ± 2.61a	30.21 ± 0.69e	47:41:12
August 24	OPT	207.56 ± 5.28d	42.16 ± 1.59e	49.32 ± 0.97d	115.08 ± 2.72a	21:24:55
	UST	281.02 ± 10.33a	123.98 ± 5.21b	104.69 ± 3.55a	52.35 ± 1.57d	44:37:19

Data were expressed as means ± SD. Different letters within each column indicate significant differences (P < 0.05).

TABLE 4 Effects of OPT and UST on bud number per meter of Donghong kiwifruit.

Date	Treatment	Slender canes with 6 ~ 8 mm in diameter	Medium canes with 8 ~ 10 mm in diameter	Thick canes with > 10 mm in diameter
June 28	OPT	11.21 ± 0.79f	12.53 ± 0.83f	13.23 ± 1.71e
	UST	14.01 ± 1.05d	15.31 ± 1.51c	15.85 ± 1.85bc
July 11	OPT	11.52 ± 0.83f	13.82 ± 1.27de	13.03 ± 1.62e
	UST	13.52 ± 0.97e	14.21 ± 1.35d	16.59 ± 1.56b
July 24	OPT	15.13 ± 1.25c	15.52 ± 1.69c	14.26 ± 1.72d
	UST	16.73 ± 1.92b	19.95 ± 2.07a	18.38 ± 1.93a
August 10	OPT	15.33 ± 1.37c	13.25 ± 1.17e	15.32 ± 1.77c
	UST	15.41 ± 1.29c	15.11 ± 1.39c	18.53 ± 1.65a
August 24	OPT	14.58 ± 1.43d	15.18 ± 1.83c	15.91 ± 1.27bc
	UST	18.29 ± 1.92a	17.39 ± 1.92b	18.37 ± 1.95a

Data were expressed as means ± SD. Different letters within each column indicate significant differences (P < 0.05).

placing roughly 20th among kiwifruit producing countries (FAOSTAT, 2019). Improving the yield per unit area is a more profitable approach for the whole kiwifruit industry. Here, we consider the potential of the UST system to address yield challenges in our kiwifruit industry. Surprisingly, the estimated yield of the UST treatment is more than two times higher than that of the traditional OPT. In addition, significant differences were observed in the internal fruit quality. These outcomes of this study may contribute to the sustainability of the kiwifruit industry.

Vegetative growth of one-year-old canes

In its natural habitat, the kiwifruit plant is a fast-growing climbing vine that needs to be properly trellised under for commercial production (Costa and Biasi, 1991; Cieslak et al., 2011). In this study, UST system significantly ($P < 0.05$) promoted the vegetative growth of slender and medium canes with diameters of 6 ~ 10 mm, but inhibited the vegetative growth of thick canes with diameters > 10 mm (Table 3). The kiwifruit plant presents rapid growth and large biomass and is not self-supporting. Its cane weight increases as it grows, resulting in cane bending, which hinders apical dominance by stimulating the growth of lateral and axillary buds. A plausible explanation for the observed significantly vigorous growth of one-year-old canes is that when the canes were treated upward on the UST system, the cane weight was supported, the scrambling vegetative growth was managed, and the preferred apical dominance of one-year-old fruiting canes was maintained (Barbier et al., 2017).

Leaf photosynthesis

The upper umbrella-shaped canopy of the UST served as a natural shading condition for the lower fruiting canopy (Figure 1).

TABLE 5 Effects of the OPT and UST on fruiting canes of Donghong kiwifruit.

Fruiting canes with different diameters	Treatment	Total fruiting cane length (cm)	Bud number	Budbreak number	Budbreak rate (%)	Flower bud number	Flower bud rate (%)
slender canes with 6 ~ 8 mm in diameter	OPT	65.37 ± 1.27e	9.67 ± 0.13c	5.78 ± 0.13c	59.73 ± 0.31a	5.74 ± 0.15c	59.38 ± 0.27a
	UST	156.33 ± 2.31a	23.50 ± 0.27a	14.31 ± 0.10a	60.89 ± 0.35a	14.21 ± 0.13c	60.47 ± 0.22a
medium canes with 8 ~ 10 mm in diameter	OPT	49.32 ± 1.15f	7.67 ± 0.15d	4.10 ± 0.09d	53.48 ± 0.27b	4.04 ± 0.02d	52.69 ± 0.19b
	UST	115.72 ± 2.69b	20.67 ± 0.21b	11.30 ± 0.17b	54.67 ± 0.41b	11.16 ± 0.05b	53.99 ± 0.21b
thick canes with > 10 mm in diameter	OPT	103.27 ± 1.58c	16.57 ± 0.11d	6.51 ± 0.03e	39.26 ± 0.37c	5.89 ± 0.01e	35.56 ± 0.11c
	UST	82.51 ± 1.30d	16.23 ± 0.11d	6.22 ± 0.11e	38.33 ± 0.15c	5.95 ± 0.01e	36.63 ± 0.13c

Data were expressed as means ± SD. Different letters within each column indicate significant differences (P < 0.05).

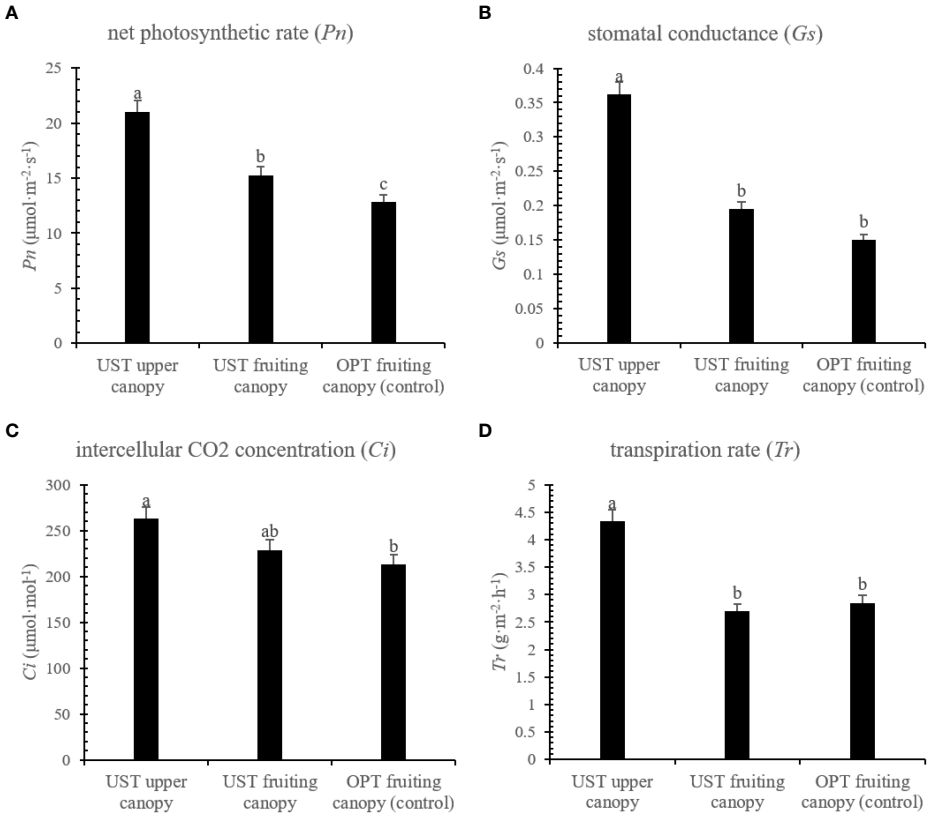


FIGURE 2
Effects of the OPT and UST on **(A)** net photosynthetic rate (*Pn*), **(B)** stomatal conductance (*Gs*), **(C)** intercellular CO_2 (*Ci*), and **(D)** transpiration rate (*Tr*) of Donghong kiwifruit. The lower-case letters on the bars indicate significant differences ($P < 0.05$).

In previous studies, proper shading has been shown to positively affect the kiwifruit yield and quality and alleviate the high temperature and intense sunlight stresses (Biasi et al., 1995; He et al., 2007). In this study, the UST system significantly ($P < 0.05$) increased the leaf gas exchange parameters of the upper canopy (Figure 2). An intriguing interpretation is that the UST system allows the canopy to better exploit both the upper and lower spaces with sufficient light and ventilation. Significant ($P > 0.05$) changes were not observed between the UST and OPT fruiting canopies (Figure 2), which is consistent with previous studies (Rosenqvist et al., 1991; He et al., 2007). One potential explanation is that the shading function of the upper canopy to the lower fruiting canopy is still within the range of proper shading and the fruiting canopy is still capable of capturing sufficient sunlight for photosynthesis under summer conditions (Rosenqvist et al., 1991).

Photosynthesis is a main process in plant development and physiology that is closely linked to crop yield. Leaf chlorophyll molecules are essential pigments that determine photosynthesis rates (Croft et al., 2017). In this study, the UST treatment had positive effects on the accumulation of Chl a, Cl b, and TC in the fruiting canopy, while negative effects on these parameters were observed in the upper canopy (Figure 3). The excessive sunlight stress can be especially severe in summer, when a lack of available water combined with high temperatures severely limits the foliar ability to use radiant energy for photosynthesis. The intense sunlight in summer might promote photosynthetic pigment degradation (Miller et al., 2010; Pinheiro and Chaves, 2011; Flexas et al., 2014). With help of the upper canopy architecture, a relatively weak light condition was generated in the lower fruiting canopy, which triggered the plants to synthesize more photosynthetic pigments to maintain efficient light harvesting.

Endogenous hormones, carbon-nitrogen ratio, and flower bud differentiation

The UST had a significant effect on one-year-old fruiting cane length (Table 3), bud number per meter (Table 4), budbreak number, and flower bud number (Table 5). The most productive zone was on the fruiting canes with a diameter of 6~10 mm (Tables 1-5). Flower bud differentiation is a critical step in the plant fruiting process, and endogenous hormones are key elements that contribute to flower bud differentiation (Sandoval-Oliveros et al., 2017; He et al., 2018); however, their effects on flower differentiation vary depending on the plant species. For example, GA has a promoting effect on flower formation of long-day and biennial plants (Mutasa-Göttgens and Hedden, 2009), but an inhibitory effect on *Arabidopsis thaliana* (Yamaguchi et al., 2014).

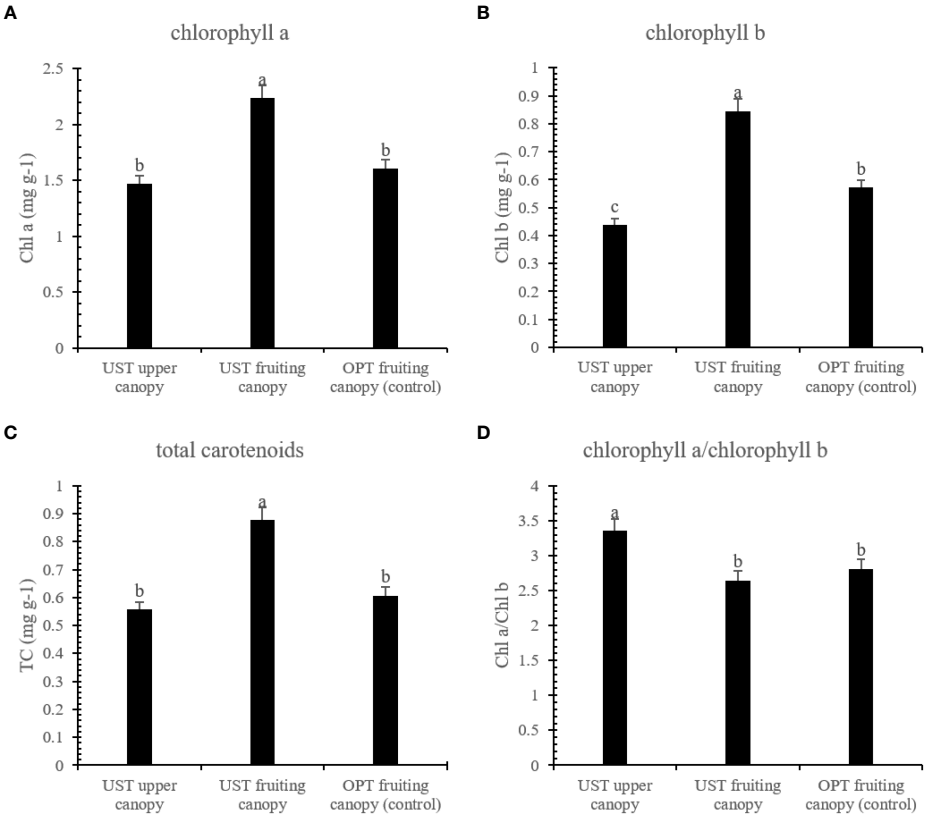


FIGURE 3
Effects of the OPT and UST on **(A)** chlorophyll a (Chl a), **(B)** chlorophyll b (Chl b), **(C)** total carotenoid (TC), and **(D)** chlorophyll a/chlorophyll b of Donghong kiwifruit. The lower-case letters on the bars indicate significant differences ($P < 0.05$).

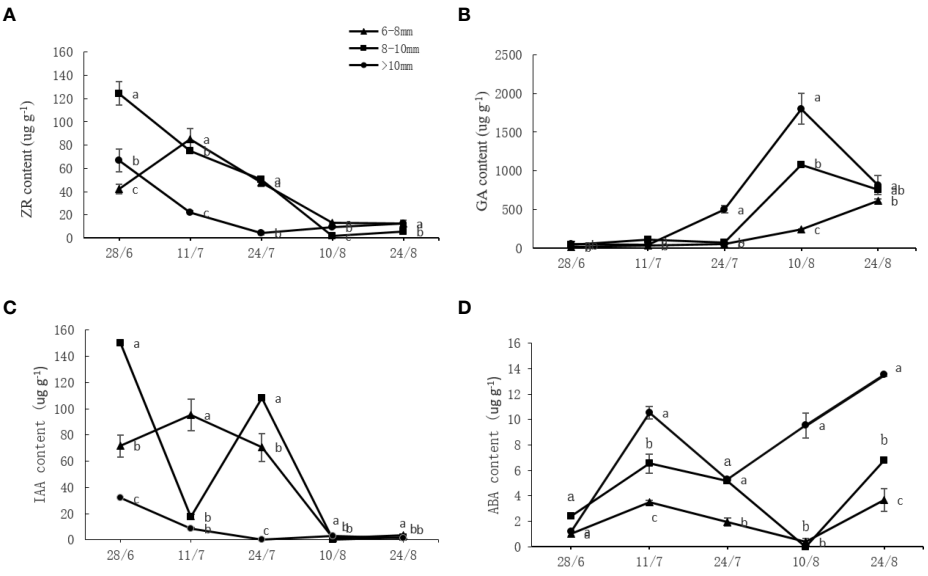


FIGURE 4
Effect of the UST on **(A)** zeatin riboside (ZR), **(B)** auxin (IAA), **(C)** abscisic acid (ABA), and **(D)** gibberellin (GA) content in buds of the upper canopy. The lower-case letters indicate significant differences ($P < 0.05$).

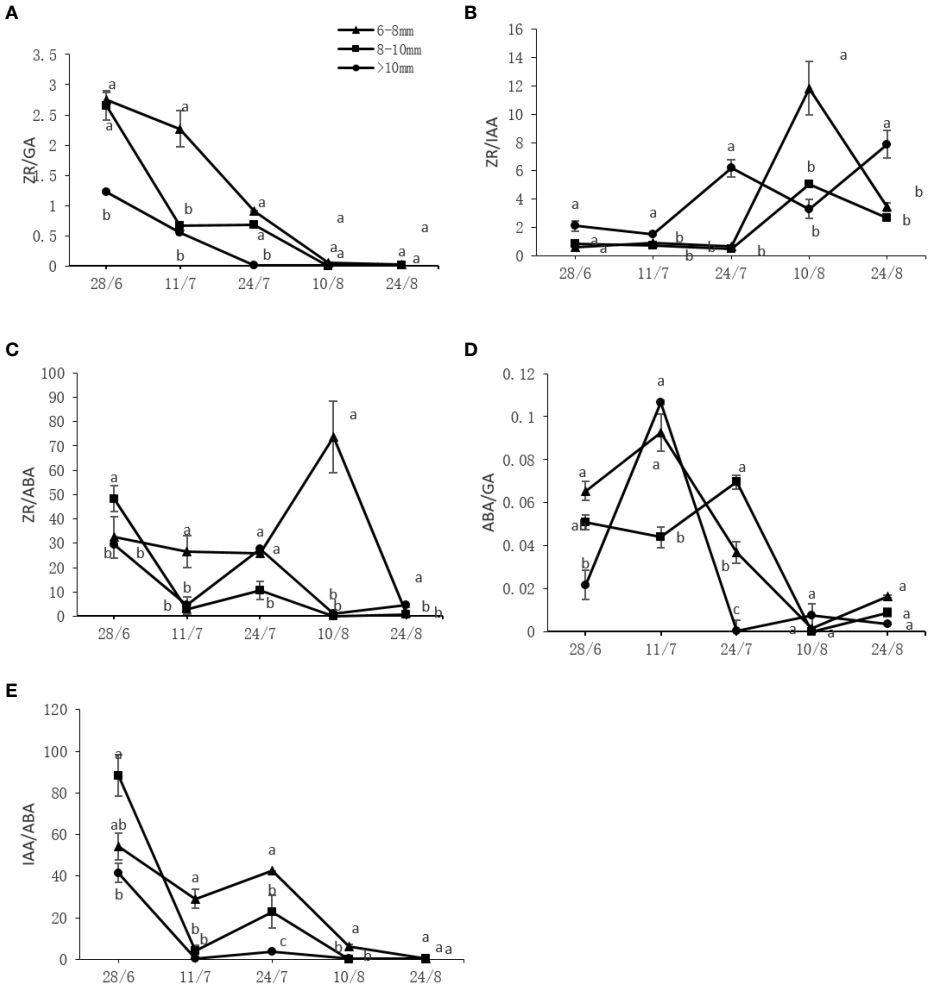


FIGURE 5
Effect of the UST on the **(A)** zeatin riboside (ZR)/gibberellin (GA), **(B)** ZR/auxin (IAA), **(C)** ZR/abscisic acid (ABA), **(D)** ABA/GA, and **(E)** IAA/ABA ratios in the bud on upper canopy. The lower-case letters indicate significant differences ($P < 0.05$).

In this study, the slender and medium fruiting canes on UST-treated trees with significantly higher ($P < 0.05$) budbreak and flower bud numbers (Table 5) contained significantly higher ($P < 0.05$) levels of ZR and IAA (Figure 4). The thick canes with significantly lower ($P < 0.05$) budbreak and flower bud numbers (Table 5), however, contained significantly higher ($P < 0.05$) levels of GA and ABA (Figure 4).

A single hormone cannot exert a substantial impact on plant flower bud differentiation. As a complex physiological and biochemical process, the bud differentiation is generally coordinated by various hormones and is greatly influenced by the content and ratio of hormones (Domagalska et al., 2010). In this study, the slender and medium fruiting canes on UST-treated trees with significantly higher ($P < 0.05$) budbreak and flower bud numbers (Table 5) contained significantly higher ($P < 0.05$) levels of ZR/GA, ZR/ABA, and ABA/GA (Figure 5). This suggests that high levels of ZR/GA, ZR/ABA, and ABA/GA ratios might be beneficial to flower bud differentiation of DH kiwifruit.

The C/N ratio is considered an indicator of plant growth and flower quality characteristics. High C/N ratios are capable of promoting reproductive growth, whereas low C/N ratios are thought to promote early vegetative growth or even inhibit flowering (Tsai and Chang, 2022). The results here (Figure 6) showed that a relatively high C/N ratio may promote the flower bud differentiation process of DH kiwifruit.

Conclusions

The current kiwifruit yield per unit area in China is not keeping pace with its global production ranking. In this study, an UST system was developed with the aim of increasing kiwifruit yield. Surprisingly, UST-treated trees were twice as productive than traditional OPT-treated system. UST-treated trees also presented improved fruit quality. The UST system significantly ($P < 0.05$) promoted the vegetative growth of fruiting canes with diameters of 6 ~ 10 mm and had positive effects on the accumulation of chlorophyll and total

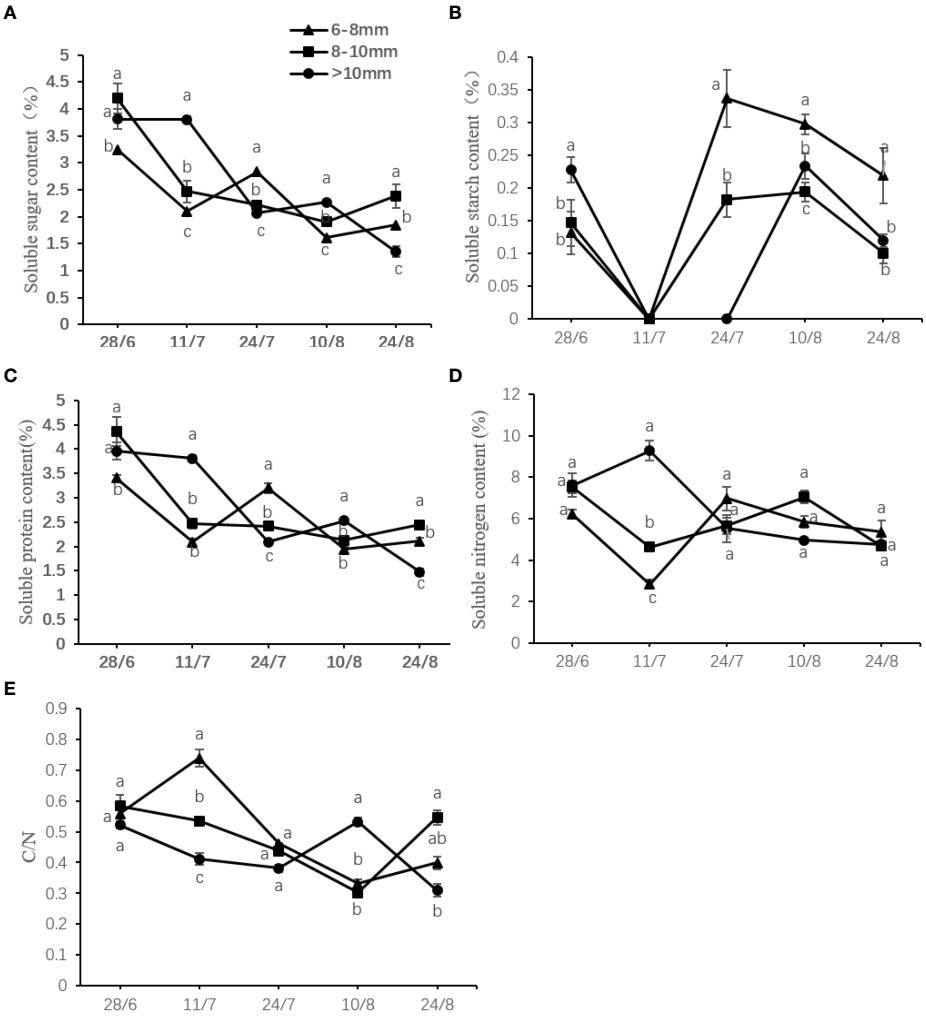


FIGURE 6
Effect of the UST on the **(A)** soluble sugar content, **(B)** soluble starch content, **(C)** soluble carbon content, **(D)** soluble protein content, and **(E)** C/N ratio in the bud on upper canopy. The lower-case letters represented significant difference ($P < 0.05$).

carotenoid contents in the fruiting canopy. The high yield may also be due to the improvements in ZR and IAA contents and ZR/GA, ZR/ABA, and ABA/GA ratios. A high ratio of C/N might be beneficial to promoting the flower bud differentiation process.

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

HX and DL conceived and designed the research. HD, YLi, KZ, CP, XT, TW, YLiang, ZH, and YLang performed the experiments. HD, YLi, JF, LL, JW, and XL analyzed the data. HD prepared and wrote the manuscript. HD, HX, and DL checked and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was partially supported by grants from Science & Technology Bureau of Chengdu City (2022-YF05-00655-SN), Science & Technology Department of Sichuan Province (2022YFH0049 and 2023YFH0003), and 'Xi Bu Zhi Guang' Visiting Scholar Program to HX.

Acknowledgment

The authors are grateful to Dr. Erika Varkonyi-Gasic and Dr. Zexi Cai (Aarhus University in Denmark) for help with data analysis and figure improvement.

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

Barbier, F. F., Dun, E. A., and Beveridge, C. A. (2017). Apical dominance. *Curr. Biol.* 27, R864–R865.

Biasi, R., P.J, M., and Costa, G. (1995). Light influence kiwifruit(*Actinidia deliciosa*) quality. *Acta Hortic.* 379, 245–251.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem.* 72, 248–254.

Cieslak, M., Seleznyova, A. N., and Hanan, J. (2011). A functionalstructural kiwifruit vine model integrating architecture, carbon dynamics and effects of the environment. *Ann. Bot.* 107, 747–764.

Costa, G., and Biasi, R. (1991). Comparison of kiwifruit traininng systems. *Acta Hortic.* 297, 427–434.

Croft, H., Chen, J. M., Luo, X., Bartlett, P., Chen, B., and Staebler, R. M. (2017). Leaf chlorophyll content as a proxy for leaf photosynthetic capacity. *Glob. Change Biol.* 23, 3513–3524.

Domagalska, M. A., Sarnowska, E., Nagy, F., and Davis, S. J. (2010). Genetic analyses of interactions among gibberellin, abscisic acid, and brassinosteroids in the control of flowering time in arabidopsis thaliana. *PloS One* 5, e14012.

Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A., and Smith, F. (1956). Colorimetric method for determination of sugars and related substances. *Anal. Chem.* 28, 350–356.

Fang, J., and Zhong, C. (2019). Fruit scientific research in new China in the past 70 years: kiwifruit (In Chinese with English abtract). *J. Fruit Sci.* 36, 1352–1359.

FAOSTAT (2019) *Top 10 kiwi fruit producing countries in the world - production and area under cultivation*. Available at: https://numerical.co.in/numerons/collection/611a67171785141f7892d4af (Accessed September 26, 2022).

Ferguson, A. R. (2016). "Botanical description," in *The kiwifruit genome, compendium of plant genomes*. Eds. R. Testolin, H. Huang and A. R. Ferguson (Springer Switzerland), 1–14.

Flexas, J., Diaz-Espejo, A., Gago, J., Gallé, A., Galmés, J., Gulías, J., et al. (2014). Photosynthetic limitations in Mediterranean plants: A review. *Environ. Exp. Bot.* 103, 12–23.

Hartmann, T. (2018) *Establishing kiwifruit orchards in Texas*. Available at: https://aggie-horticulture.tamu.edu/kiwifruit/content/fact_sheets/kiwifruit_orchards_in_texas.pdf (Accessed September 26, 2022).

He, K., Wang, Z., and Wang, R. (2007). Effects of overhead shading in summer on ecological factors and photosynthesis of kiwifruit orchard. *J. Fruit Sci.* 24, 616–619.

Horwitz W. and Latimer G. W. (Eds.) (2006). *Official methods of analysis of AOAC international. 18th edition* (Gaithersburg, MD: AOAC International Gaithersburg).

He, W., Chen, Y., Gao, M., Zhao, Y., Xu, Z., Cao, P., et al. (2018). Transcriptome analysis of *Litsea cubeba* floral buds reveals the role of hormones and transcription factors in the differentiation process. *G3 Genes Genomes Genet.* 8, 1103–1114.

Huang, H. W., and Ferguson, A. R. (2007). *Actinidia* in China: Natural diversity, phylogeographical evolution, interspecific gene flow and kiwifruit cultivar improvement. *Acta Hortic.* 753, 31–40.

Huang, W., Wang, Z., Zhang, Q., Feng, S., and Burdon, J. (2022). Maturity , ripening and quality of ' donghong ' kiwifruit evaluated by the kiwi-meter ^{TM}. *Horticulture* 8, 852.

Huang, S., Ding, J., Deng, D., Tang, W., Sun, H., Liu, D., et al. (2013). Draft genome of the kiwifruit *Actinidia* chinensis. *Nat. Commun.* 4, 2640.

Liang, D., Huang, X., Shen, Y., Shen, T., Zhang, H., Lin, L., et al. (2019). Hydrogen cyanamide induces grape bud endodormancy release through carbohydrate metabolism and plant hormone signaling. *BMC Genomics* 20, 1034.

Miller, G., Suzuki, N., Ciftci-Yilmaz, S., and Mittler, R. (2010). Reactive oxygen species homeostasis and signalling during drought and salinity stresses. *Plant Cell Environ.* 33, 453–467.

Molnárová, M., and Fargašová, A. (2009). Se(IV) phytotoxicity for monocotyledonae cereals (*Hordeum vulgare* l., *Triticum aestivum* l.) and dicotyledonae crops (*Sinapis alba* l., *Brassica napus* l.). *J. Hazard Mater.* 172, 854–861.

Mutasa-Göttgens, E., and Hedden, P. (2009). Gibberellin as a factor in floral regulatory networks. *J. Exp. Bot.* 60, 1979–1989.

Pinheiro, C., and Chaves, M. M. (2011). Photosynthesis and drought: Can we make metabolic connections from available data? *J. Exp. Bot.* 62, 869–882.

Qi, X., Guo, D., Wang, R., Zhong, Y., and Fang, J. (2020). Development status and suggestions on Chinese kiwifruit industry (In Chinese with English abtract). *J. Fruit Sci.* 37, 754–763.

Rosenqvist, E., Wingsle, G., and Ögren, E. (1991). Photoinhibition of photosynthesis in intact willow leaves in response to moderate changes in light and temperature. *Physiol. Plant* 83, 390–396.

Sandoval-Oliveros, R., Guevara-Olvera, L., Beltrán, J. P., Gómez-Mena, C., and Acosta-Garcia, G. (2017). Developmental landmarks during floral ontogeny of jalapeño chili pepper (*Capsicum annuum* l.) and the effect of gibberellin on ovary growth. *Plant Reprod.* 30, 119–129.

Strik, B., and Cahn, H. (2000) *Growing kiwifruit*. Available at: http://content.libraries.wsu.edu/index.php/utils/getfile/collection/cahnrs-arch/id/499/filename/96136182432004_pnw507.pdf (Accessed September 26, 2022).

Tsai, S. S., and Chang, Y. C. A. (2022). Plant maturity affects flowering ability and flower quality in phalaenopsis, focusing on their relationship to carbon-to-nitrogen ratio. *HortScience* 57, 191–196.

Wang, S., Qiu, Y., and Zhu, F. (2021). Kiwifruit (*Actinidia* spp.): A review of chemical diversity and biological activities. *Food Chem.* 350, 128469.

Yamaguchi, N., Winter, C. M., Wu, M.-F., Kanno, Y., Yamaguchi, A., Seo, M., et al. (2014). Gibberellin acts positively then negatively to control onset of flower formation in *Arabidopsis*. *Science* 344, 638–641.

Zhong, C., Han, F., Li, D., Liu, Xi., Zhang, Q., Jiang, Z., et al. (2016). Breeding of red-fleshed kiwifruit cultivar. "Donghong." *J. Fruit Sci.* 32, 1596–1599.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 31 March 2023
DOI 10.3389/fpls.2023.1158837

Check for updates

OPEN ACCESS

EDITED BY
Jiangang Liu,
Chinese Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Wenjuan Li,
Institute of Agricultural Resources and Regional Planning (CAAS), China
Haikuan Feng,
Beijing Research Center for Information Technology in Agriculture, China

*CORRESPONDENCE
Meichen Feng
fmc101@163.com

SPECIALTY SECTION
This article was submitted to Sustainable and Intelligent Phytoprotection, a section of the journal Frontiers in Plant Science

RECEIVED 04 February 2023
ACCEPTED 20 March 2023
PUBLISHED 31 March 2023

CITATION
Sun X, Yang Z, Su P, Wei K, Wang Z, Yang C, Wang C, Qin M, Xiao L, Yang W, Zhang M, Song X and Feng M (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

COPYRIGHT
© 2023 Sun, Yang, Su, Wei, Wang, Yang, Wang, Qin, Xiao, Yang, Zhang, Song 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.

Non-destructive monitoring of maize LAI by fusing UAV spectral and textural features

Xinkai Sun[1], Zhongyu Yang[1], Pengyan Su[1], Kunxi Wei[1], Zhigang Wang[1], Chenbo Yang[1], Chao Wang[1], Mingxing Qin[2], Lujie Xiao[1], Wude Yang[1], Meijun Zhang[1], Xiaoyan Song[1] and Meichen Feng[1*]

[1]College of Agriculture, Shanxi Agricultural University, Jinzhong, Taigu, Shanxi, China, [2]College of Resources and Environment, Shanxi Agricultural University, Jinzhong, Taigu, Shanxi, China

Leaf area index (LAI) is an essential indicator for crop growth monitoring and yield prediction. Real-time, non-destructive, and accurate monitoring of crop LAI is of great significance for intelligent decision-making on crop fertilization, irrigation, as well as for predicting and warning grain productivity. This study aims to investigate the feasibility of using spectral and texture features from unmanned aerial vehicle (UAV) multispectral imagery combined with machine learning modeling methods to achieve maize LAI estimation. In this study, remote sensing monitoring of maize LAI was carried out based on a UAV high-throughput phenotyping platform using different varieties of maize as the research target. Firstly, the spectral parameters and texture features were extracted from the UAV multispectral images, and the Normalized Difference Texture Index (NDTI), Difference Texture Index (DTI) and Ratio Texture Index (RTI) were constructed by linear calculation of texture features. Then, the correlation between LAI and spectral parameters, texture features and texture indices were analyzed, and the image features with strong correlation were screened out. Finally, combined with machine learning method, LAI estimation models of different types of input variables were constructed, and the effect of image features combination on LAI estimation was evaluated. The results revealed that the vegetation indices based on the red (650 nm), red-edge (705 nm) and NIR (842 nm) bands had high correlation coefficients with LAI. The correlation between the linearly transformed texture features and LAI was significantly improved. Besides, machine learning models combining spectral and texture features have the best performance. Support Vector Machine (SVM) models of vegetation and texture indices are the best in terms of fit, stability and estimation accuracy (R^2 = 0.813, RMSE = 0.297, RPD = 2.084). The results of this study were conducive to improving the efficiency of maize variety selection and provide some reference for UAV high-throughput phenotyping technology for fine crop management at the field plot scale. The results give evidence of the breeding efficiency of maize varieties and provide a certain reference for UAV high-throughput phenotypic technology in crop management at the field scale.

KEYWORDS

UAV, multispectral images, leaf area index, spectral feature, texture feature, maize

1 Introduction

Maize is one of the essential food crops in the world, and its production mode and planting area are related to world food security (Tester and Langridge, 2010). Leaf area index (LAI) refers to the ratio of total green leaf area per unit land area to the unit land area (Ke et al., 2016), which is one of the important indexes for evaluating crop growth and ecological environment research (Richardson et al., 2011). It is not only closely related to crop photosynthesis (Duncan, 1971) and transpiration (Kang et al., 2003), but also often used as one of the basis for yield estimation (Duan et al., 2019; Fu et al., 2020). Therefore, efficient and precise monitoring of maize LAI is crucial for gaining insights into maize growth dynamics and optimizing maize breeding strategies.

However, the traditional methods to obtain LAI are mostly destructive sampling methods, using leaf area meters to measure isolated blades and calculate them (Yang et al., 2021). This method is time-consuming, laborious and inefficient. Although non-destructive monitoring of LAI can be achieved using handheld instruments such as the SunScan Canopy Analyser (Oguntunde et al., 2012) and the LAI-2200 (Wilhelm et al., 2000), the data obtained using handheld instruments only represent LAI at a small scale, making it difficult to realize rapid and nondestructive monitoring at field scale (Liu et al., 2016). Satellite remote sensing enables rapid and non-destructive monitoring of crop LAI at a regional scale. (Chen et al., 2010). However, its susceptibility to adverse weather conditions, low temporal and spatial resolution limits its ability to meet the quantitative monitoring requirements at the field and plot scales. The ground platform is mainly suitable for small-scale crop growth monitoring, which is affected by the scope of data acquisition and the cost of use, and cannot achieve rapid monitoring at spatial scales.

In recent years, the continuous development of UAV flight platforms and airborne sensor technology has promoted the application of UAV remote sensing technology in agricultural and forestry information monitoring. UAV remote sensing platforms have the advantages of low cost, simple structure, high mobility and high spatial and temporal resolution to make up for the shortage of satellite and ground-based remote sensing platforms (Xie and Yang, 2020). The UAV platform, equipped with visible, multispectral, and hyperspectral cameras, is used to acquire image data. Image processing techniques are applied to extract essential information, such as spectral features, texture, and point clouds, which are subsequently used to build models for crop growth parameter monitoring and yield estimation. (Liu et al., 2018; Sarkar et al., 2021; Jiang et al., 2022). The model construction process often involves a combination of non-linear relationships that affect the model's universality. Machine learning methods can effectively solve the modelling problem of non-linear relational combinations and have been widely used in remote sensing monitoring. Wang and Jiang (2021) used UAV multispectral remote sensing data to achieve the monitoring of soybean leaf area index, and the support vector machine (SVM) had better predictions compared to the linear model (R^2 = 0.688, RMSE = 0.016). Qi et al. (2020) achieved accurate estimation of peanut LAI (R^2 = 0.968, RMSE = 0.165) using Back Propagation neural network algorithm (BPNN) combined with UAV spectral features. Kanning et al. (2018) extracted wheat canopy spectra based on UAV hyperspectral images and constructed a model for monitoring wheat LAI and chlorophyll using Partial Least Squares Regression (PLSR), demonstrating the feasibility of UAV hyperspectral imaging technology for monitoring crop growth parameters at the field scale. Although spectral features combined with machine learning algorithm can better estimate crop LAI, when LAI is high, the estimation model constructed by various vegetation indexes will appear "over-fitting" phenomenon (Hang et al., 2021). In addition, UAV multispectral images offer limited spectral information, and relying solely on spectral parameters such as reflectance or vegetation index may result in "same spectrum and different things" or "same thing and different spectrum" scenarios (Liu et al., 2018). Thus, crop growth monitoring should incorporate multiple data dimensions, such as time and space, to account for spectral, temporal, and spatial resolutions.

Texture features are also one of the image features of UAV remote sensing imagery (Yang et al., 2021), and are widely used for image classification and monitoring of crop growth physiological indicators (Haralick et al., 1973; Coburn and Roberts, 2004; Li et al., 2019). Chen et al. (2019) used spectral information and texture information to estimate chlorophyll content of potatoes and found that the fusion of vegetation index and texture features could significantly improve the estimation accuracy of the model. Zhu et al. (2022) developed a machine learning model for remote sensing monitoring of Wheat Scab using multispectral and texture features. Their results demonstrated that the fusion of vegetation indices and texture parameters improved the accuracy of Wheat Scab detection. However, little is known about field-scale remote sensing monitoring of maize LAI using spectral and texture features extracted from UAV multispectral imagery.

Based on the above problems, this study attempts to extract and optimize the spectral and texture features from UAV multispectral images. It combines the optimized features with SVM, Random Forest (RF), BPNN and PLSR to build a field-scale corn LAI remote sensing monitoring model. We compared the effects of spectral features and texture features on LAI estimation. Furthermore, we explored the influence of machine learning method synergistic spectral and texture features on LAI estimation potential.

2 Materials and methods

2.1 Experimental design

This experiment was conducted in the maize experimental field of Shanxi Agricultural University, Jinzhong City, Shanxi Province (37°25′ N, 112°29′ E) (Figure 1).

The experimental area is located in Taigu District, with an average altitude of about 780 m, belonging to a temperate continental monsoon climate, with an average annual temperature of 6-10 °, an average annual rainfall of 410-450 mm

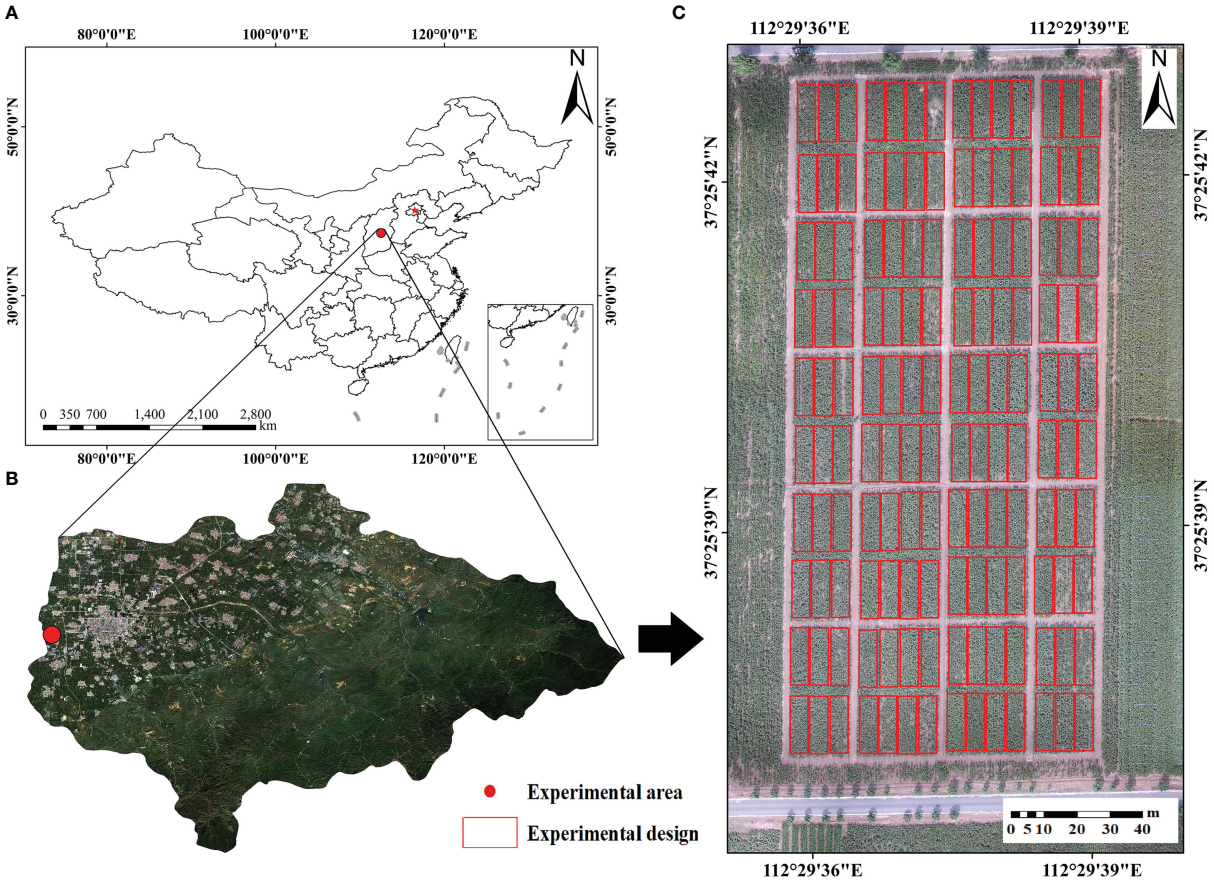


FIGURE 1
Overview of the study area. **(A)** Geographical location of Taigu District **(B)** the location of the experiential area **(C)** the design of the field experience.

and a frost-free period of 160 days. The climatic conditions such as light, heat and water are suitable for maize growth.

The experiment was conducted in a single-factor design with 140 maize varieties (Xinyu 303, Jinfeng 278, RP818, etc.), each planted on an area of 75 m^2, total 140 plots. In five key growth periods of maize, namely, tasseling period (24 July, 2021), silking period (4 August, 2021), flowering period (14 August, 2021), filling period (25 August, 2021) and milk ripening period (8 September, 2021), LAI was measured in areas with high vegetation coverage and consistent growth.

2.2 UAV multispectral image acquisition

In this study, a Meridian M210 V2 quadcopter UAV (DJI Innovations, Shenzhen, China) (Figure 2A) with a RedEdge-MX



FIGURE 2
UAV near ground remote sensing platform. **(A)** Four rotors UAV **(B)** Multispectral imaging system **(C)** The calibration panel.

imaging system (MicaSense, Seattle, WA, USA) (Figure 2B) was used as a high-throughput remote sensing platform to acquire multispectral images of maize during critical fertility periods. RedEdge-MX dual-camera imaging system has 10 spectral channels with a spectral range of 444-842 nm and can simultaneously obtain 10 discontinuous multispectral images with a resolution of 1280 × 960 pixels. The detailed band parameters are shown in Table 1. The flight time is between 10:00 and 12:00, the flight altitude is set to 60 m, and the forward overlap rate and side overlap rate are set to 85%.

2.3 UAV multispectral image processing

After the flight, the original multispectral images acquired by the UAV and the calibration plate (Figure 2C) images taken before takeoff were imported into Pix4D mapper (Pix4D S.A., Lausanne, Switzerland) together for image stitching and radiometric calibration. After the stitching was completed, the software automatically completed the radiometric calibration according to the DN value of the calibration gray plate and the reflectance calibration fitting equation. After completing the above processing process, the orthoimage of each waveband were obtained. The orthoimages were processed using ENVI Classic 5.5 (Harris Geospatial Solutions, Inc., Broomfield, CO, USA) for band fusion to obtain multispectral image data.

2.4 Field data collection

In this study, we selected the SunScan Canopy Analyser to measure maize LAI data while acquiring UAV multispectral images. During the data collection process, maize leaf area index was measured by selecting three inter-row shaded locations of 1 meter in length at random within each plot. The average of the three measurements was used as the LAI value for the plot.

2.5 Multispectral image features extraction

2.5.1 Selection of vegetation index

By combining different bands linearly or nonlinearly, the vegetation indices constructed has certain indicative significance for the dynamic changes of vegetation canopy information, which not only reduces the influence of atmospheric and soil environmental factors but also enhances the sensitivity of LAI to canopy reflectivity. According to the previous research results (Qi et al., 2020; Hang et al., 2021), we selected eight commonly used vegetation indices to estimate LAI, and the specific calculation formulas are shown in Table 2.

2.5.2 Texture features extraction

To improve computational efficiency, we select three bands of red 650 nm, red edge 705 nm and near-infrared 842 nm for texture feature extraction. Grey level co-occurrence matrix (GLCM) is one of the most widely used methods in texture feature extraction (Haralick et al., 1973). This method was proposed by Haralick in 1973, and is mainly used in machine vision, image classification, image recognition and so on (Kavdir and Guyer, 2004; Adjed et al., 2018; Vani et al., 2018). After radiation correction and image fusion, eight texture features such as Mean (mean), variance (var), homogeneity (hom), contrast (con), dissimilarity (dis), entropy (ent), second moment (sm) and correlation (cor) were extracted using the GLCM. A total of 24 texture feature values were selected and the mean value extracted from the region of interest was used as the texture feature value for the corresponding image.

In order to fully explore the application potential of texture features in UAV multispectral images in maize LAI estimation, this study used Matlab 2020a software(MathWorks, Natick, Massachusetts, USA) to traverse and combine eight texture feature values of three-band images and calculates three texture indexes: normalized difference texture index(NDTI) (Zheng et al., 2019), ratio texture index(RTI) and difference texture index(DTI) (Hang et al., 2021). The specific calculation formulas are as follows:

TABLE 1 RedEdge-MX Dual multispectral cameras bands.

Band number	Band name	Center wavelength (nm)	Bandwidth (nm)
B1	Coastal Blue*	444	28
B2	Blue	475	32
B3	Green*	531	14
B4	Green	560	27
B5	Red*	650	16
B6	Red	668	14
B7	Red edge*	705	10
B8	Red edge	717	12
B9	Red edge*	740	18
B10	Near infrared	842	57

*Indicates the bands of the RedEdge-MX Blue camera.

TABLE 2 Formula for calculating vegetation indices.

Vegetation Index	Description Formula	Reference
Normalized Difference Vegetation Index (NDVI)	$(NIR_{842}-R_{650})/(NIR_{842}+R_{650})$	(Rouse et al., 1974)
Red Edge Normalized Difference Vegetation Index (NDRE)	$(NIR_{842}-RE_{705})/(NIR_{842}+RE_{705})$	(Gitelson et al., 1996)
Modified Triangular Vegetation Index (MTCI)	$(NIR_{842}-RE_{705})/(RE_{705}+R_{650})$	(Dash and Curran, 2004)
Difference Vegetation Index (DVI)	$NIR_{842}-R_{650}$	(Naito et al., 2017)
Ratio Vegetation Index (RVI)	NIR_{842}/R_{650}	(Liang et al., 2013)
Red edge chlorophyll index ($CI_{red\ edge}$)	$NIR_{842}/RE_{705}-1$	(Gitelson et al., 1996)
Enhanced Vegetation Index (EVI)	$2.5((NIR_{842}-R_{650})/(NIR_{842}+6R_{650}-7.5B_{475}+1))$	(Prabhakara et al., 2015)
Soil-adjusted Vegetation Index (OSAVI)	$(1+0.16)(NIR_{842}-R_{650})/(NIR_{842}+R_{650}+0.16)$	(Rondeaux et al., 1996)

B_{475}, R_{650}, RE_{705} and NIR_{842} in the table represent the reflectance at blue, red, red edge and near-infrared bands, respectively.

$$NDTI = (T_1-T_2)/(T_1+T_2) \quad (1)$$

$$RTI = T_1/T_2 \quad (2)$$

$$DTI = T_1-T_2 \quad (3)$$

In the formula, T_1 and T_2 are texture eigenvalues of random bands.

2.6 Model construction and evaluation

In this study, a total of 700 datasets were collected from five key fertility periods of maize, each containing ground-truthed LAIs and UAV image features such as vegetation indices and texture indices. To ensure that each dataset could be involved in modeling and validation, the datasets were randomly divided into 10 parts using ten-fold cross-validation, with 90% (630 datasets) used for modeling and 10% (70 datasets) for model validation. Each model was trained 10 times to ensure robustness.

We developed 12 LAI estimation models for multiple fertility stages by combining ground truth LAIs with various inputs using four machine learning algorithms, namely SVM, RF, BPNN, and PLSR. The model inputs consisted of univariate and multivariate factors, where the former comprised vegetation indices and texture indices, and the latter involved their combination.

(1) The support vector machine is a popular machine learning algorithm for pattern recognition and nonlinear regression (Cortes and Vapnik, 1995). In this study, we used the SVM algorithm with a radial basis function (RBF) to construct maize LAI estimation models using various predictors. The SVM model requires tuning of the penalty factor c and the kernel function parameter g. After continuous testing, we determined the optimal values of c=1.00 and g=3.03.

(2) Random forest is a nonlinear regression modeling method based on multiple decision trees. It consists of two methods, Bootstrap sampling and Random subspace(Breiman, 2001), and is effective in handling high-dimensional data and covariance problem among variables, with strong noise resistance. The number of decision trees (ntrees) and the number of predictors randomly selected for each split (mtry) are the main parameters that need to be tuned to optimize the performance of a random forest model. After repeated testing, this study set the random forest parameters to ntrees=200 and tuned mtry according to different input variable types.

(3) Back Propagation (BP) based neural network consists of three parts: input layer (input), hidden layer (hidden), and output layer (output). To ensure the model monitoring accuracy, the BPNN model was trained several times and the model parameters were iteratively tuned. Finally, the learning rate of 0.01, 10 hidden layers and 1 output layer were used as the best parameters.

(4) PLSR originates from the nonlinear iterative partial least squares (NUPALS) algorithm proposed by Herman Wold et al. (2001). The number of latent variables (LV) is one of the important influencing factors to determine the prediction accuracy of the PLSR model, and in this study the model automatically adjusted the number of latent variables according to the input variable types.

In this study, coefficient of determination (R^2), root mean square error (RMSE) and relative percent error (RPD) were used as evaluation indexes to evaluate the model performance. When the estimated model has higher R^2, RPD and smaller RMSE in modeling and validation datasets, it indicates that the model has higher goodness of fit, accuracy and stability.

3 Results

3.1 Correlation analysis between spectral parameters and LAI

The correlation analysis of the 18 spectral parameters with LAI is shown in Figure 3. The correlation between LAI and canopy reflectance was significantly negative ($p < 0.01$) in the range of 444-717 nm and positive ($p < 0.01$) in the range of 740-842 nm. All eight selected vegetation indices were highly significantly positively

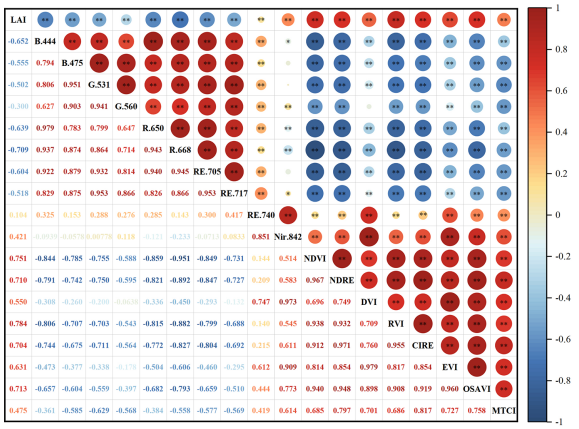


FIGURE 3
Correlation coefficient between specific parameters and maize LAI. * and * * are significant at 0.05 and 0.01 levels, respectively. The red area indicates positive correlation, the blue area indicates negative correlation. Darker colors and larger circles mean a stronger correlation between LAI and spectral parameters.

correlated with LAI ($P < 0.01$), with RVI being the most strongly correlated with LAI ($r = 0.784$) and NDVI, OSAVI, NDRE and $CI_{red\ edge}$ being more strongly correlated with LAI, with correlation coefficients above 0.700. All five vegetation indices included the three bands, red, red edge and NIR, indicating that the band combinations could be better for maize LAI monitoring.

3.2 Correlation analysis between LAI and texture features

The correlation analysis of the three bands of texture features with LAI (Table 3) showed that the mean texture values in the Red and NIR band were significantly correlated with LAI, with a strong correlation ($r = -0.687$ and -0.703).

Due to the weak correlation between most texture features and LAI, this study constructed three texture indices composed of different texture feature values in order to improve the potential application of texture features in monitoring maize LAI. The results shown in Figure 4 indicated that the correlation between the linearly transformed texture features and LAI was significantly enhanced. Among them, the RTI ($mean_{705}$, ent_{705}) had the strongest correlation with LAI, with a correlation coefficient of -0.804, which was a 14.370% increase in the absolute value of the correlation coefficient compared to the red edge mean.

3.3 Comparison of LAI estimation models based on vegetation indices and texture indices

In order to fully investigate the potential of combining UAV spectral and texture features with machine learning algorithms for LAI estimation. Based on the strength of the correlation between different image features and LAI, we selected five vegetation indices (RVI, NDVI, OSAVI, NDRE and $CI_{red\ edge}$) and three texture indices (RTI ($mean_{705}$, ent_{705}), DTI ($mean_{705}$, con_{705}) and NDTI ($mean_{650}$, ent_{705})) as independent variables, and constructed LAI estimation models by using four machine learning algorithms: SVM, RF, BPNN and PLSR respectively. Table 4 shows the training results of machine learning models with different input variables. In the single variable model, from the perspective of modelling methods, RF performed best in the dataset of vegetation indices and SVM performed best in the dataset of texture indices; From the perspective of the different input variables, the estimation model based on the texture indices performs better overall than the vegetation indices when using the same modelling approach. On the whole, the estimation of the SVM model based on TIs was optimal ($R^2 = 0.790$, RMSE = 0.312, RPD = 2.010).

Using vegetation and texture indices as multivariate input variables to construct the LAI estimation model. From the perspective of modelling method, the RF model in the calibration set performed the best from the perspective of the modelling approach ($R^2 = 0.906$, RMSE = 0.208, RPD = 3.149), with the SVM model performing second best ($R^2 = 0.806$, RMSE = 0.315, RPD = 1.856). However, in the validation set, the SVM model performed best ($R^2 = 0.813$, RMSE = 0.297, RPD = 2.084). In

TABLE 3 Correlation coefficients between texture features of the three bands and LAI.

Texture Features	Correlation Coefficients		
	Red_{650}	Red $edge_{705}$	NIR_{842}
Mean(mean)	-0.687**	-0.703**	-0.095*
Variance (var)	-0.311**	-0.121**	0.055
Homogeneity (hom)	0.285**	0.073	-0.206**
Contrast (con)	-0.285**	-0.076*	0.084*
Dissimilarity (dis)	-0.291**	-0.071	0.131**
Entropy (ent)	-0.369**	-0.144**	0.188**
Second moment (sm)	0.357**	0.165**	-0.196**
Correlation (cor)	-0.088	0.036	0.237*

*and ** are significant at the 0.05 and 0.01 levels.

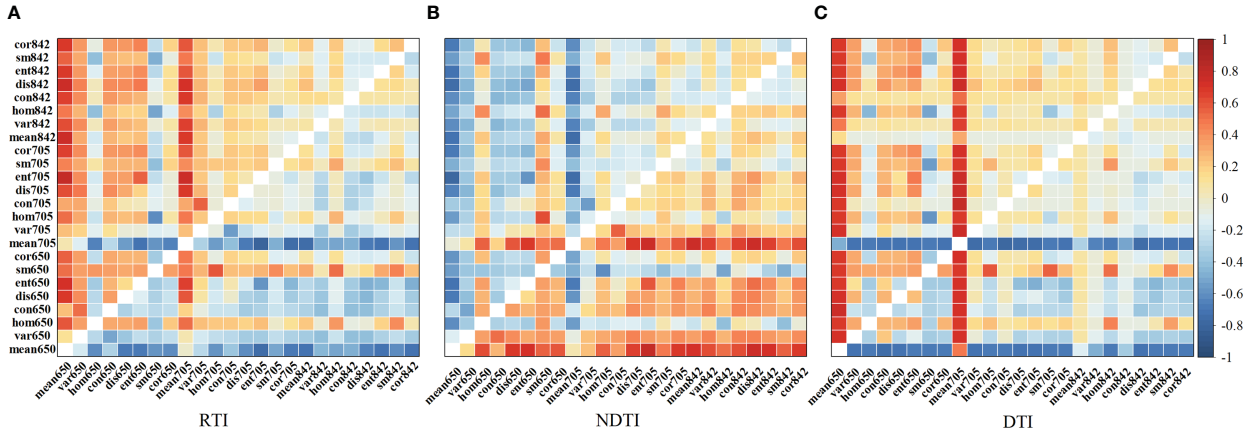


FIGURE 4
Correlation coefficient matrix between LAI and three types of texture index indices **(A)** the ratio texture index, **(B)** the normalized difference texture index, **(C)** the difference texture index. Each cell in the figure represents the correlation coefficient between the texture index, which is obtained by linearly transforming the original texture parameters corresponding to the x and y coordinates of each cell, and the LAI.

contrast, the R^2 of the RF model plummeted from 0.906 to 0.786, the RMSE increased by 66.827% and the RPD decreased by 40.044%. Although the monitoring effect of BPNN and PLSR is slightly weaker than that of SVM and RF, the estimation accuracy and model stability are also better (R^2 > 0.75, RPD > 1.75). The above results show that the SVM model has the best estimation accuracy and stability, and the other three models also have great prediction results.

When analyzed from the perspective of the input variables, the machine learning models constructed by fusing the two types of indices explained significantly more variance in the LAI compared to the single-factor input variables of the vegetation or texture indices. Combining the VIs with the TIs resulted in R^2 means of 0.817 and 0.788, RMSE means of 0.292 and 0.330, and RPD means of 2.165 and 1.921 for the calibration and validation sets, respectively. Compared with the single vegetation indices data source model, R^2 increased by 14.810% and 19.757%, RMSE decreased by 19.337% and 19.118%, and RPD increased by 33.807% and 29.622%. Compared with the single texture indices model, R^2 increases by 11.507% and 11.142%, RMSE decreases by 16.092% and 14.063%, and RPD increases by 31.212% and 15.237%. The above results show that the estimation effect of the model is obviously improved and more stable after fusing different image features.

The scatter plot in Figure 5 showed good consistency between the predicted LAI values from the machine learning estimation model and the measured LAI values in the validation dataset, with an RMSE ranging from 0.297 to 0.433 and an RPD ranging from 1.328 to 2.084. Combining vegetation indices and texture indices resulted in the best estimation results among the four types of machine learning models.

4 Discussion

UAV remote sensing has great potential in the process of crop phenotype information mining and analysis due to its high spatial and temporal resolution and simple operation (Xie and Yang,

TABLE 4 Summary of the results of estimating LAI by machine learning models based on different inputs.

Technique	Data set	VIs			TIs			VIs+TIs		
		R^2	RMSE	RPD	R^2	RMSE	RPD	R^2	RMSE	RPD
SVM	Cal	0.698	0.371	1.487	0.735	0.348	1.656	0.806	0.315	1.856
	Val	0.621	0.433	1.338	0.790	0.312	2.010	**0.813**	**0.297**	**2.084**
BPNN	Cal	0.648	0.403	1.335	0.689	0.380	1.445	0.791	0.311	1.899
	Val	0.664	0.392	1.716	0.688	0.389	1.787	0.795	0.330	1.936
RF	Cal	0.840	0.271	2.257	0.844	0.269	2.095	0.906	0.208	3.149
	Val	0.733	0.396	1.545	0.699	0.411	1.399	0.786	0.347	1.888
PLSR	Cal	0.649	0.402	1.392	0.652	0.397	1.405	0.753	0.335	1.755
	Val	0.616	0.413	1.328	0.660	0.425	1.474	0.759	0.347	1.775

The value in bold type indicates that the model is the best LAI estimation model. The Cal and Val in the table represent the calibration set data set and validation set data set respectively.

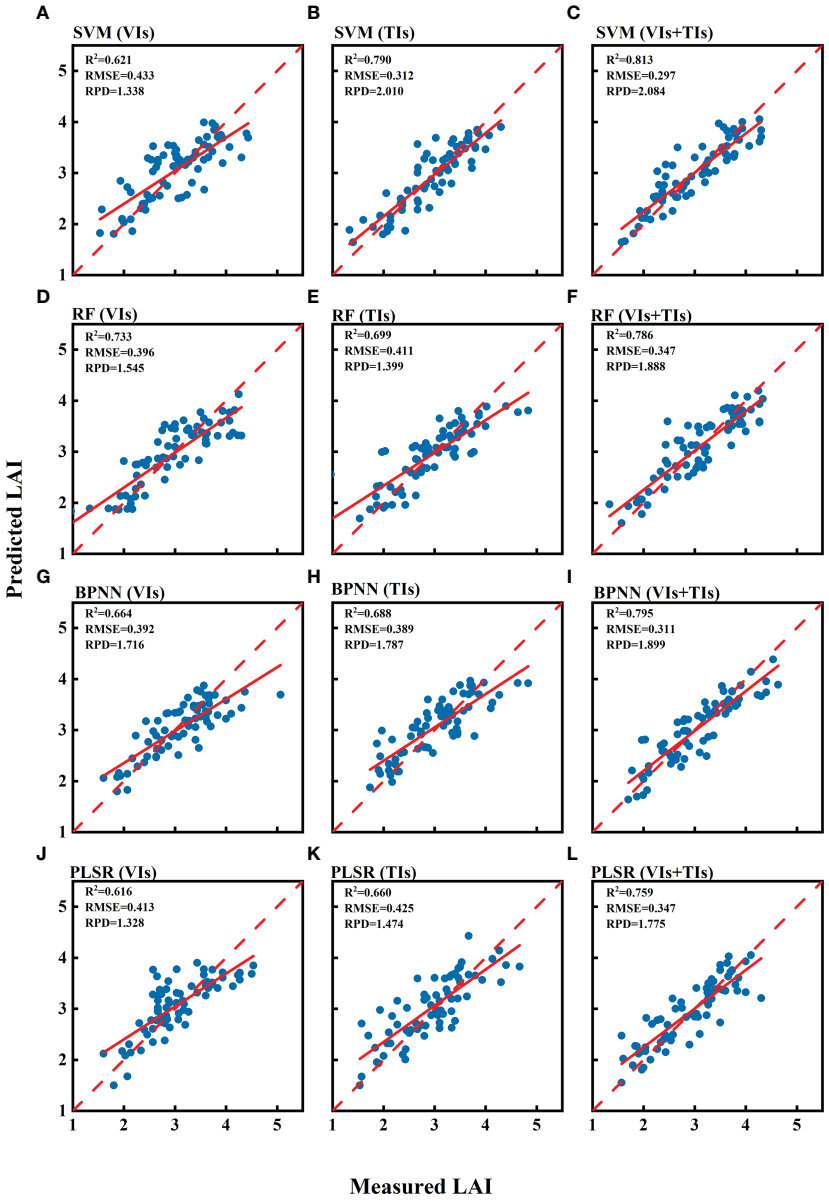


FIGURE 5
Accuracy evaluation results of LAI estimation models based on vegetation indices (VIs), texture indices (TIs) and combined vegetation indices and texture indices (VIs+TIs) in the validation set. The models evaluated are SVM, RF, BPNN, and PLSR, shown respectively in **(A–C)**, **(D–F)**, **(G–I)**, and **(J–L)**.

2020). In this study, multispectral images of the study area were acquired by using a UAV with a multispectral camera. The non-destructive and rapid estimation of maize LAI at the plot scale was achieved by extracting different types of image features and combining them with machine learning algorithms.

4.1 Analysis of monitoring LAI by vegetation indices

Vegetation indices are widely used in crop chlorophyll content (Jiang et al., 2022), LAI (Li et al., 2019), biomass (Gnyp et al., 2014) and yield prediction (Fu et al., 2020; García-Martínez et al., 2020). Crop canopy reflectance is easily influenced by leaf pigmentation in the visible bands, which can lead to "oversaturation" of the vegetation indices (Hatfield et al., 2008). In contrast, red-edge and near-infrared reflectance are mainly influenced by canopy structure and have a stronger penetration effect. Hence, researchers usually choose red-edge and near-infrared bands to construct vegetation indices. Shi et al. (2022) demonstrated that vegetation indices based on red light bands and near-infrared bands correlate well with LAI and AGB of red and green beans, allowing for growth monitoring of intercropped crops in plot tea plantations. Zheng et al. (2019) found that the vegetation indices based on red-edge bands were important parameters for rice biomass estimation before tasseling. However, the estimation was significantly reduced after tasseling, mainly because canopy leaf biomass was sensitive to red-edge bands but not stems. Qi et al. (2020) used fixed-wing UAV to monitor peanut

growth, and found that red light and near-infrared bands were sensitive bands of peanut LAI, which can effectively predict the changes of peanut LAI. In this study,we found strong correlations (r > 0.700) between maize LAI and five vegetation indices: RVI, NDVI, OSAVI, NDRE, and $CI_{red\ edge}$. These indices were identified as effective for monitoring LAI in various maize varieties. The results are in general agreement with the results of previous studies, indicating that spectral indices based on red light, red edge and near-infrared bands are of good application in crop monitoring and can achieve rapid and non-destructive monitoring of crop growth parameters.

In this study, DVI and MTCI performed poorly in estimating LAI, and the correlation coefficients were only 0.550 and 0.475. There may be two possible reasons for this result: (1) the influence of other disturbing factors such as soil background and vegetation shading on the multispectral reflectance; (2) the high LAI level in the middle and late stages of maize growth, resulting in an underestimation of some vegetation indices. In the process of monitoring crop growth using UAV multispectral imagery, the use of spectral features alone may not achieve satisfactory results (Zheng et al., 2019; Yang et al., 2021; Fei et al., 2022). The use of vegetation indices alone can only quantitatively analyze the structural characteristics, biochemical components and productivity trends of crop canopies from a spectral perspective. It cannot deeply explore the effective information of other data dimensions in UAV multispectral imagery.

4.2 Estimation potential analysis of texture features

Texture features can reveal changes in crop canopy information from the data dimension of image spatial features. Previous work has used texture parameters from satellite remote sensing data combined with spectral and topographic features to estimation above-ground biomass in forests, demonstrating the feasibility of applying texture parameters in agricultural remote sensing (Lu and Batistella, 2005; Li et al., 2008). In this study, most of the texture parameters had poor correlation with LAI. However, the correlation between the constructed texture indices and LAI were significantly improved after linear processing. The texture indices RTI ($mean_{705}$, ent_{705}), DTI ($mean_{705}$, con_{705}) and NDTI ($mean_{650}$, ent_{705}) effectively improved the estimation of LAI. Similar conclusions were obtained by Hang et al. (2021) and Fei et al. (2022) when using texture parameters for rice growth monitoring and yield estimation of wheat. This is mainly because the texture is linearly combined and transformed to reduce the influence of soil background, vegetation shading and topographic factors, which can better highlight the changing patterns of feature characteristics.

4.3 Influence of UAV image features fusion on LAI estimation potential

The multi-input model based on vegetation and texture indices in this study was superior to the model with a single input variable, with significant improvements in fit, estimation accuracy and stability. In particular, the model combining UAV texture and spectral features outperformed the model using only the vegetation indices, with a 19.757% increase in R^2, a 19.118% decrease in RMSE and a 29.622% increase in RPD in the validation set. The results of this study are similar to those of previous studies. Ma et al. (2022) used color indices and texture features from UAV RGB images to accurately estimate cotton yield, with the RF_ELM model based on color indices and texture features having the highest accuracy (R^2 = 0.911). Yang et al. (2021) used vegetation indices and texture features to achieve an estimation of LAI for rice at full fertility. The combination of spectral features and texture features had superior predictive power than vegetation indices. In summary, combining the UAV spectral features with texture features is an effective method to improve the accuracy of LAI estimation.

4.4 Comparison of different machine learning models

Machine learning algorithms combined with remote sensing data have been widely used in areas such as crop growth monitoring(Li et al., 2019; Zheng et al., 2019; Zhang et al., 2021), yield estimation (Fu et al., 2020; García-Martínez et al., 2020; Ma et al., 2022) and disease identification (Guo et al., 2021; Zhu et al., 2022). This study used four machine learning algorithms, SVM, RF, BPNN and PLSR, to construct LAI monitoring models for different maize varieties. The results show that the SVM model performs best as a whole, and the model's training results and verification results have a high degree of explanation for the variation of LAI. In the machine learning models constructed based on vegetation and texture indices, the validation sets R^2 and RPD of SVM were improved and RMSE decreased compared to RF, PLSR and BPNN models, indicating the high performance of SVM modelling. Zhu et al. (2022) realized high-precision monitoring of wheat scab by using machine learning method combined with spectral and texture features of drones, and the model built by SVM in collaboration with VIs + TFs can provide the most accurate monitoring results; Omer et al. (2016) used WorldView-2 multispectral imagery combined with SVM and ANN algorithms to achieve monitoring of LAI of forest endangered tree species, where the SVM model showed excellent prediction accuracy and model stability. The excellent performance of SVM model in LAI estimation may be related to the model structure. SVM uses the principle of structure minimization (Camps-Valls et al., 2006) to solve the nonlinear mapping problem between input variables and response variables. However, to address the issue of unstable RF model performance, a similar situation has been found in other studies, Han et al. (2019) used the cooperative machine learning algorithm of canopy structure information and spectral information to construct maize AGB, the performance of RF model in the training set and test set was quite different. There are two main reasons for the instability of the RF model: (1) the amount of data in the validation set is far less than that in the calibration set, and RF will perform better in extensive sample data; (2) the existence of outliers in the validation set due to human measurement problems reduces the stability of the model.

In this paper, when using different machine learning algorithms to estimate LAI with multivariate input variables, all four machine learning models achieved good performance, indicating that there is a non-linear relationship between the response variables and the various predictors. However, the input variables were selected without deeper mining of the input variable feature selection, and the contribution of different predictors to the LAI estimation model was not considered. To improve the monitoring accuracy of LAI, it is necessary to study the above shortcomings in future research.

5 Conclusions

Rapid and non-destructive plot-scale maize LAI estimation is important for UAV remote sensing monitoring of crop growth as well as precise agricultural management. In this study, we used image analysis techniques to extract spectral and texture features from UAV multispectral images and used machine learning methods (SVM, RF, BPNN, PLSR) to achieve fast and accurate estimation of maize LAI. Most Vegetation indices based on red, red-edge, and NIR bands exhibited strong correlation with LAI, whereas most texture features demonstrated limited association with LAI. Nevertheless, after applying linear transformation, texture indices displayed a substantially enhanced correlation with LAI. Among the different types of estimation models, the model constructed by SVM method combined with vegetation indices and texture indices was the best for LAI estimation (R^2 = 0.813, RMSE=0.297, RPD=2.084), and this result revealed that there was a non-linear relationship between LAI and spectral parameters and texture parameters. The results of this study show that the use of UAV near-ground remote sensing combined with image analysis techniques can achieve accurate monitoring of the growth of different maize varieties and provide guidance for maize variety selection.

Data availability statement

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

Author contributions

Conceptualization, XinS, WY and MF. data curation, XinS, ZY, PS and ZW. formal analysis, XinS and CY. funding acquisition, WY and MF. investigation, XinS, ZY, PS, KW and ZW. project administration, WY and MF. resources, WY and MF. writing—original draft, XinS. writing—review and editing, WY and MF. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by the Key Research and Development Program of Shanxi Province, China (201903D211002-01, 201903D211002-05).

Acknowledgments

We are grateful to the Shanxi Academy of Agricultural Sciences for providing the trial site. We are also grateful to the editor 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.

References

Adjed, F., Safdar Gardezi, S. J., Ababsa, F., Faye, I., and Chandra Dass, S. (2018). Fusion of structural and textural features for melanoma recognition. *IET Comput. Vis.* 12, 185–195. doi: 10.1049/iet-cvi.2017.0193

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/a:1010933404324

Camps-Valls, G., Bruzzone, L., Rojo-Álvarez, J. L., and Melgani, F. (2006). Robust support vector regression for biophysical variable estimation from remotely sensed images. *IEEE Geosci Remote Sens Lett.* 3, 339–343. doi: 10.1109/lgrs.2006.871748

Chen, P., Feng, H., Li, C., Yang, G., Yang, J., Yang, W., et al. (2019). Estimation of chlorophyll content in potato using fusion of texture and spectral features derived from UAV multispectral image. *Trans. CSAE.* 35, 63–74. doi: 10.11975/j.issn.1002-6819.2019.11.008

Chen, X., Meng, J., Du, X., Zhang, F., Zhang, M., and Wu, B. (2010). Remote sensing monitoring model of winter wheat leaf area index based on environmental satellite CCD data. *Remote Sens Land Resou* 2, 55–58+62. doi: 1001-070X(2010)02-0055-04

Coburn, C., and Roberts, A. C. (2004). A multiscale texture analysis procedure for improved forest stand classification. *Int. J. Remote Sens.* 25, 4287–4308. doi: 10.1080/0143116042000192367

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018

Dash, J., and Curran, P. (2004). The MERIS terrestrial chlorophyll index. *Int. J. Remote Sens.* 23, 5403–5413. doi: 10.1080/0143116042000274015

Duan, B., Fang, S., Zhu, R., Wu, X., Wang, S., Gong, Y., et al. (2019). Remote estimation of rice yield with unmanned aerial vehicle (UAV) data and spectral mixture analysis. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00204

Duncan, W. J. C. S. (1971). Leaf angles, leaf area, and canopy photosynthesis. *Crop Sci.* 11, 482–485. doi: 10.2135/cropsci1971.0011183x001100040006x

Fei, S., Hassan, M. A., Xiao, Y., Su, X., Chen, Z., Cheng, Q., et al. (2022). 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

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

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. *Agric* 10, 277. doi: 10.3390/agriculture10070277

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 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

Hang, Y., Su, H., Yu, Z., Liu, H., Guan, H., and Kong, F. (2021). Estimation of rice leaf area index combining UAV spectrum, texture features and vegetation coverage. *Trans. CSAE.* 37, 64–71. doi: 10.11975/j.issn.1002-6819.2021.09.008

Haralick, R. M., Shanmugam, K., and Dinstein, (1973). Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* 6, 610–621. doi: 10.1109/tsmc.1973.4309314

Hatfield, J. L., Gitelson, A. A., Schepers, J. S., and Walthall, C. L. (2008). Application of spectral remote sensing for agronomic decisions. *Agron. J.* 100, S–117-S-131. doi: 10.2134/agronj2006.0370c

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

Kang, S., Gu, B., Du, T., and Zhang, J. (2003). Crop coefficient and ratio of transpiration to evapotranspiration of winter wheat and maize in a semi-humid region. *Agric. Water Manage.* 59, 239–254. doi: 10.1016/s0378-3774(02)00150-6

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

Kavdir, I., and Guyer, D. (2004). Comparison of artificial neural networks and statistical classifiers in apple sorting using textural features. *Biosyst. Eng.* 89, 331–344. doi: 10.1016/j.biosystemseng.2004.08.008

Ke, L., Zhou, Q., Wu, W., Tian, X., and Tang, H.-j. (2016). Estimating the crop leaf area index using hyperspectral remote sensing. *J. Integr. Agric.* 15, 475–491. doi: 10.1016/s2095-3119(15)61073-5

Li, M., Tan, Y., Pan, J., and Peng, S. (2008). Modeling forest aboveground biomass by combining spectrum, textures and topographic features. *Front. For China.* 3, 10–15. doi: 10.1007/s11461-008-0013-z

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

Liang, D., Guan, Q., Huang, W., Huang, L., and Yang, G. (2013). Remote sensing inversion of leaf area index based on support vector machine regression in winter wheat. *Trans. CSAE.* 29, 117–123. doi: 10.3969/j.issn.1002-6819.2013.07.015

Liu, C., Yang, G., Li, Z., Tang, F., Wang, J., Zhang, C., et al. (2018). Biomass estimation of winter wheat by fusing UAV spectral information with texture information. *Sci. Agric. Sin.* 51, 3060–3073. doi: 10.3864/j.issn.0578-1752.2018.16.003

Liu, J., Zhao, C., Yang, G., Yu, H., Zhao, X., Xu, B., et al. (2016). Review of field-based phenotyping by unmanned aerial vehicle remote sensing platform. *Trans. CSAE.* 32, 98–106. doi: 10.11975/j.issn.1002-6819.2016.24.013

Lu, D., and Batistella, M. (2005). Exploring TM image texture and its relationships with biomass estimation in rondônia, Brazilian Amazon. *Acta Amazonica.* 35, 249–257. doi: 10.1590/s0044-59672005000200015

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. doi: 10.3389/fpls.2022.925986

Naito, H., Ogawa, S., Valencia, M. O., Mohri, H., Urano, Y., Hosoi, F., et al. (2017). Estimating rice yield related traits and quantitative trait loci analysis under different nitrogen treatments using a simple tower-based field phenotyping system with modified single-lens reflex cameras. *ISPRS J. Photogramm Remote Sens.* 125, 50–62. doi: 10.1016/j.isprsjprs.2017.01.010

Oguntunde, P. G., Fasinmirin, J. T., and Abiolu, O. A. (2012). Performance of the SunScan canopy analysis system in estimating leaf area index of maize. *Agric. Eng. Int: CIGR Journa.* 14, 1–7.

Omer, G., Mutanga, O., Abdel-Rahman, E. M., and Adam, E. (2016). Empirical prediction of leaf area index (LAI) of endangered tree species in intact and fragmented indigenous forests ecosystems using WorldView-2 data and two robust machine learning algorithms. *Remote Sens.* 8, 324. doi: 10.3390/rs8040324

Prabhakara, K., Hively, W. D., and McCarty, G. W. (2015). Evaluating the relationship between biomass, percent groundcover and remote sensing indices across six winter cover crop fields in Maryland, united states. *Int. J. Appl. Earth Obs.* 39, 88–102. doi: 10.1016/j.jag.2015.03.002

Qi, H., Zhu, B., Wu, Z., Liang, Y., Li, J., Wang, L., et al. (2020). Estimation of peanut leaf area index from unmanned aerial vehicle multispectral images. *Sensors* 20, 6732. doi: 10.3390/s20236732

Richardson, A. D., Dail, D. B., and Hollinger, D. (2011). Leaf area index uncertainty estimates for model–data fusion applications. *Agr For. Meteorol.* 151, 1287–1292. doi: 10.1016/j.agrformet.2011.05.009

Rondeaux, G., Steven, M., and Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sens Environ.* 55, 95–107. doi: 10.1016/0034-4257(95)00186-7

Rouse, J. W., Haas, R. W., Schell, J. A., Deering, D. W., and Harlan, J. C. (1974). *Monitoring the vernal advancement and retrogradation (Green wave effect) of natural vegetation. NASA/GSFCT type III final report no. NASA-CR-139243* (College Station, TX: Texas A&M University).

Sarkar, S., Cazenave, A.-B., Oakes, J., McCall, D., Thomason, W., Abbott, L., et al. (2021). Aerial high-throughput phenotyping of peanut leaf area index and lateral growth. *Sci. Rep.* 11, 1–17. doi: 10.1038/s41598-021-00936-w

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

Tester, M., and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822. doi: 10.2307/40509897

Vani, K., Poongodi, S., and Harikrishna, B. (2018). K-Means cluster based leaf disease identification in cotton plants. *Indian J. Public Health Resea Dev.* 9, 1117–1120. doi: 10.5958/0976-5506.2018.01288.3

Wang, J., and Jiang, Y. (2021). Inversion of leaf area index of soybean based on UAV multispectral remote sensing. *Chin. Agric. Sci. Bull.* 37, 134–142. doi: casb2020-0229

Wilhelm, W., Ruwe, K., and Schlemmer, M. R. (2000). Comparison of three leaf area index meters in a corn canopy. *Crop Sci.* 40, 1179–1183. doi: 10.2135/cropsci2000.4041179x

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometr Intell. Lab.* 58, 109–130. doi: 10.1016/s0169-7439(01)00155-1

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, K., Gong, Y., Fang, S., Duan, B., Yuan, N., Peng, Y., et al. (2021). Combining spectral and texture features of UAV images for the remote estimation of rice LAI throughout the entire growing season. *Remote Sens.* 13, 3001. doi: 10.3390/rs13153001

Zhang, J., Qiu, X., Wu, Y., Zhu, Y., Cao, Q., Liu, X., et al. (2021). 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

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

Zhu, W., Feng, Z., Dai, S., Zhang, P., and Wei, X. (2022). Using UAV multispectral remote sensing with appropriate spatial resolution and machine learning to monitor wheat scab. *Agric* 12, 1785. doi: 10.3390/agriculture12111785

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 June 2023
DOI 10.3389/fpls.2023.1202536

Check for updates

OPEN ACCESS

EDITED BY
Zhenjiang Zhou,
Zhejiang University, China

REVIEWED BY
Hengbiao Zheng,
Nanjing Agricultural University, China
Kiril Manevski,
Aarhus University, Denmark

*CORRESPONDENCE
Mitchell R. Tuinstra
mtuinstr@purdue.edu

RECEIVED 08 April 2023
ACCEPTED 06 June 2023
PUBLISHED 20 June 2023

CITATION
Tolley SA, Carpenter N, Crawford MM, Delp EJ, Habib A and Tuinstra MR (2023) Row selection in remote sensing from four-row plots of maize and sorghum based on repeatability and predictive modeling. *Front. Plant Sci.* 14:1202536. doi: 10.3389/fpls.2023.1202536

COPYRIGHT
© 2023 Tolley, Carpenter, Crawford, Delp, Habib and Tuinstra. 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.

Row selection in remote sensing from four-row plots of maize and sorghum based on repeatability and predictive modeling

Seth A. Tolley[1], Neal Carpenter[2], Melba M. Crawford[1,3], Edward J. Delp[4], Ayman Habib[3] and Mitchell R. Tuinstra[1]*

[1]Department of Agronomy, Purdue University, West Lafayette, IN, United States, [2]Analytics and Pipeline Design, Bayer Crop Science, Chesterfield, MO, United States, [3]Lyles School of Civil Engineering, Purdue University, West Lafayette, IN, United States, [4]School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, United States

Remote sensing enables the rapid assessment of many traits that provide valuable information to plant breeders throughout the growing season to improve genetic gain. These traits are often extracted from remote sensing data on a row segment (rows within a plot) basis enabling the quantitative assessment of any row-wise subset of plants in a plot, rather than a few individual representative plants, as is commonly done in field-based phenotyping. Nevertheless, which rows to include in analysis is still a matter of debate. The objective of this experiment was to evaluate row selection and plot trimming in field trials conducted using four-row plots with remote sensing traits extracted from RGB (red-green-blue), LiDAR (light detection and ranging), and VNIR (visible near infrared) hyperspectral data. Uncrewed aerial vehicle flights were conducted throughout the growing seasons of 2018 to 2021 with data collected on three years of a sorghum experiment and two years of a maize experiment. Traits were extracted from each plot based on all four row segments (RS) (RS1234), inner rows (RS23), outer rows (RS14), and individual rows (RS1, RS2, RS3, and RS4). Plot end trimming of 40 cm was an additional factor tested. Repeatability and predictive modeling of end-season yield were used to evaluate performance of these methodologies. Plot trimming was never shown to result in significantly different outcomes from non-trimmed plots. Significant differences were often observed based on differences in row selection. Plots with more row segments were often favorable for increasing repeatability, and excluding outer rows improved predictive modeling. These results support long-standing principles of experimental design in agronomy and should be considered in breeding programs that incorporate remote sensing.

KEYWORDS

remote sensing, high-throughput phenotyping, RGB, lidar, hyperspectral, UAV, plot trimming, border effect

1 Introduction

Maize (*Zea mays* L.) and sorghum (*Sorghum bicolor* (L.) Moench) are among the most important crops in the world that are utilized for food, animal feed, biofuels, and other applications. These crops are desirable because of their C4 photosynthetic pathways that enable greater photosynthetic potential in hot, dry environments. These crops are likely to be of increasing global importance in the coming years and continued crop improvement are needed to improve food security (Ray et al., 2013).

Remote sensing using sensors attached to uncrewed aerial vehicles (UAV) enables in-season field measurements with high spatial and temporal resolutions necessary for small-plot research (Maes and Steppe, 2019; Ravi et al., 2019; Masjedi et al., 2020). Remote sensing has been used for plant phenotyping for a number of crops including maize (Zaman-Allah et al., 2015; Pugh et al., 2018; Anderson et al., 2019; Anche et al., 2020), sorghum (Watanabe et al., 2017; Pugh et al., 2018; Masjedi et al., 2020; Yang K. et al., 2021), soybean (Moreira, 2020), and wheat (Tattaris et al., 2016; Crain et al., 2018; Hassani et al., 2023). Various sensors are now widely flown on a UAV including RGB (red-green-blue) cameras, multispectral/hyperspectral sensors, and LiDAR (light detection and ranging) units. With an onboard Global Navigation Satellite System/Inertial Navigation System (GNSS/INS) and system calibration, data from these sensors can be georeferenced directly with high spatial accuracy (Ravi et al., 2019). These sensors collect large amounts of data that provide opportunities for the measurement, estimation, or prediction of a wide range of attributes such as grain yield, biomass productivity, leaf area index, canopy cover, plant counts, nitrogen content, and disease detection (Yang et al., 2007; Zarco-Tejada et al., 2012; Mathews and Jensen, 2013; Ribera et al., 2017; Li et al., 2018; Toledo et al., 2022; Hassani et al., 2023).

In maize and sorghum breeding programs, experimental hybrids are evaluated in field trials, and the genotypes with the most desirable combinations of traits are selected and advanced in the respective breeding programs. Nevertheless, traditional hybrid testing programs are time consuming and expensive; creating a phenotyping bottleneck for progress in crop improvement (Cobb et al., 2013; Yang et al., 2017). Phenotyping by remote sensing has the potential to relieve the phenotyping strain on breeding programs by reducing the time-consuming and laborious nature of low-throughput phenotyping while simultaneously increasing the throughput on the number of plots that can be evaluated (Furbank and Tester, 2011; Araus and Cairns, 2014; Araus et al., 2018). However, given the observation geometry and sensor technologies, certain traits cannot be computed directly from remotely sensed data, as with traditional phenotyping methods.

High-throughput phenotyping by remote sensing enables the evaluation of more plots and can improve genetic gain in plant breeding. More genotypes can be evaluated in the same land area by either decreasing the plot size or by maintaining plot size and increasing land area and consequently cost. Often, this tradeoff is balanced in a breeding program by using small plots in preliminary trials and larger plots in advanced breeding trials where yield is a primary trait of interest (Acquaah, 2012). A common agronomic practice is to make phenotyping measurements on the inner rows of a multi-row plot to minimize border effects between competing genotypes. The border row effects have been well defined and as a result these outer plot rows are harvested separately or ignored in yield trials (Bird, 1929; Genter, 1958; Gomez, 1972, Kramer et al., 1982; Bowman, 1989; Petersen, 1994; Ceccarelli and Grando, 1996; Reynolds & Braun, 2022) (Supplemental Figure 1). Zhang et al. (2019) found that soybean yield was best predicted where hyperspectral data was analyzed for a subset of 20 to 80% of the plot area. More work is needed to understand the border effect on remote sensing traits and their use in breeding programs.

Trait repeatability is an important factor in plant breeding programs (Bernardo, 2014). Repeatability is defined as the signal to noise ratio for phenotypic measurements, indicating the relative importance of genetic effects in comparison to environmental effects in phenotypic observations. Proper experimental design including replication, blocking, randomization, and plot size all influence the trait repeatability. Barmeier and Schmidhalter (2016) demonstrated that high-throughput spectral phenotyping can be used in small-plot research comprising single or multiple rows. However, no study has evaluated the impact of row selection or plot trimming on the repeatability of remote sensing traits. Additionally, a primary goal of in-season remote sensing is to predict end-season yield (biomass or grain). Selection of rows from remote sensing data products could be an important consideration impacting the results of predictive modeling.

In this study, experimental hybrids of maize and sorghum were evaluated in four-row plots in field trials from 2018 to 2021. Remote sensing flights were conducted with RGB, LiDAR, and VNIR (visible near infrared) hyperspectral sensors throughout the growing season. Biomass and grain yield were evaluated in sorghum and maize, respectively, at the end of the growing seasons. The objectives of this experiment were to (1) determine the impact of row selection on repeatability based on remote sensing traits, (2) compare repeatability of remote sensing-based traits with end-trimming and no trimming, and (3) demonstrate the impact of row selection on end-season biomass and grain yield prediction.

2 Materials and methods

2.1 Field experiment and germplasm

Field trials were conducted from 2018 to 2021 in West Lafayette, Indiana, US at the Agronomy Center for Research and Education at Purdue University (40°28'37.18"N, 86°59'22.67"W). The experiments were planted as four-row plots with a length of 3.05 m by 3 m with 76 cm row spacing (Figure 1). Alleys were 76 cm in length between ranges. Nutrients, herbicides, and insecticides were applied according to best agronomic practices to not limit plant growth and development in these experiments. In each study, the experiments were planted in a crop rotation with soybean.

In 2018, 2019, and 2020, 619 inbred lines from the sorghum diversity panel were assessed for their testcross hybrid performance with ATx623 (sorghum reference genome). This experiment was



FIGURE 1
Hyperspectral orthophoto colored as red-green-blue showcasing the 2020 sorghum (large, blue box) and maize (small, red box) experiments used in this study.

planted on May 8, 2018; June 4, 2019; and May 12, 2020 with a consistent seeding rate of 220,000 seeds ha^{-1}. On June 7, 2018, 28% liquid UAN was applied at 16 g N m^{-2}. In October 2018, 34 g m^{-2} of potash (0-0-60) and 1,121 g m^{-2} of lime for the 2019 growing season. Additionally, 18 g N m^{-2} of anhydrous ammonia was applied on May 9, 2019. On April 6, 2020, 18 g N m^{-2} of anhydrous ammonia was applied. The sorghum experiments followed a randomized complete block design with 2 replications. End-season biomass was harvested from rows two and three of each plot on August 14, 2018; September 12, 2019; and August 19, 2020 using a Wintersteiger Cibus Biomass Harvester (Wintersteiger Inc., Salt Lake City, UT, USA). Moisture content was determined by sampling ~500 g of chopped biomass from each plot from which fresh weight and dry weight (after samples were dried for ten days at 72°C) were recorded. Moisture content was used to adjust biomass from fresh weight to dry weight.

In 2020 and 2021, a maize field experiment was grown comprised of ten temperate and ten tropical inbred lines evaluated for their testcross hybrid performance with PHP02. This experiment was planted on May 12, 2020 and May 23, 2021 at a population of 74,000 seeds ha^{-1} using a randomized complete block design with 3 replications. On April 6, 2020, 18 g N m^{-2} of anhydrous ammonia was applied. On April 14, 2021, 28% liquid UAN was applied at 16.25 g N m^{-2}. Grain yield was harvested from rows two and three of each plot on October 14, 2020 and October 21, 2021 using a Kincaid plot combine (Kincaid 8-XP, Haven KS, USA) with grain yields adjusted to 15% moisture.

2.2 Sensors

A DJI M600 pro equipped with an onboard Applanix APX-15 GNSS/INS unit, which allowed for direct-georeferencing, and integrated RGB, LiDAR, and VNIR sensors, was used in each of the studies. Boresight calibrations were performed for each sensor for co-alignment of the sensors and flight dates as described by Habib et al. (2018); Ravi et al. (2018), and LaForest et al. (2019). RGB data were collected using a Sony Alpha 7R digital mirrorless camera with a Sony Sonnar T* FE 35 mm lens. The Sony Alpha 7R camera features a full-frame 36.4 MP sensor delivering high-resolution images. Georeferenced orthophotos were generated using a structure-from-motion strategy introduced by He et al. (2018) and Hasheminasab et al. (2020). LiDAR data were collected with a Velodyne VLP-16 instrument, which has 16 beams and a 360-degree horizontal field of view with a maximum range of 100 m. It can scan up to 300,000 points $second^{-1}$ with a typical accuracy of ±3cm. VNIR data were collected with a Headwall Photonics Nano-Hyperspec, a hyperspectral push-broom scanner collecting data in 272 spectral bands at 2.2 µm $band^{-1}$ from 400 nm to 1000 nm. It has 640 spatial channels with a7.4µm pixel pitch. Spectral targets calibrated using re Spectra Vista SVC 1024i were used in post-processing to convert from radiance to reflectance via the Empirical Line Method. VNIR orthophotos were obtained using the digital surface model from the georeferenced LiDAR point cloud through the approach developed by Lin and Habib (2021).

2.3 Flight information

Flight information including date, growing degree days (GDD), and sensor data collection is given in Table 1 and Supplemental Tables 1, 2. Flights were conducted at an altitude of ~40 m with a speed of ~4 m s^{-1} for a ground sampling distance of ~1 cm in RGB orthophotos and ~4cm in VNIR orthophotos. Generally, flights

TABLE 1 Number of flights with RGB (red-green-blue), LiDAR (light detection and ranging), and VNIR (visible near infrared hyperspectral) at each growth stage (vegetative, 0 – 650 GDD; flowering, 650 – 900 GDD; grain filling (900+ GDD) in maize and sorghum experiments from 2018 – 2021.

Crop	Growth Stage	Sensor Flights		
		RGB	LiDAR	VNIR
Maize	**Vegetative**	8	6	6
	Flowering	6	4	3
	Grain Filling	8	6	6
Sorghum	**Vegetative**	9	7	5
	Flowering	7	6	4
	Grain Filling	6	5	6

were conducted around solar noon on clear days with little wind. GDD for each day was calculated using the formula, GDD = $[(T_{max} + T_{min})/2]$-10. When the maximum and minimum temperatures were greater than 30°C or less than 10°C, then T_{max} and T_{min} were set to 30°C and 10°C, respectively (Gilmore and Rogers, 1958). Flights were organized across years based on GDD into vegetative (0-650 GDD), flowering (650-900 GDD), and grain filling (900 +GDD) growth stages. These GDD windows were based on the average beginning and end of flowering across the years of the maize and sorghum experiments, although there were a few late maturing sorghum genotypes that did not begin flowering until later in the growing season each year.

2.4 Remote sensing traits

Remote sensing traits were collected from RGB, LiDAR, and VNIR sensors and were extracted for each row segment (RS) of a plot 2D canopy cover (CC) was estimated from RGB imagery as the ratio of the number of pixels attributed to vegetative material, based on thresholds in the HSV (Hue, Saturation Value) color space, to the total number of pixels in a RS (Chen, 2019) and 3D canopy cover by LiDAR as the number of points above the 10th (CC10) percentile divided by the total number of points in each RS (Masjedi et al., 2020). LiDAR was also used to quantify variation in height and plot volume. Height was determined at the 95th (Height 95%) percentile of the nonground points. Plot volume was estimated by assigning cell sizes 8 cm x 8 cm to each plot. Within each grid, height was calculated as the average of the 95th percentile height and the minimum height of nonground points. Average height was multiplied by the cell size to estimate the volume of vegetation in a plot. Vegetation indices were calculated from the plant material in hyperspectral orthophotos where the ground was masked using thresholds of vegetation indices (Table 2).

2.5 Row selection and plot trimming

Row segmentation was performed as described by Yang et al. (2021a) to enclose the vegetation in rectangles. Briefly, the intersections of vertical and horizontal lines result in a grid of N x M coordinates where N is the number of ranges in the field and M is the number of rows. These grids result in bounding boxes around each individual RS of each plot in an early season RGB orthophoto (Figure 2). Bounding boxes do not all have exactly the same size because of variations in the location of emerging plants at the ends of the row segment. The plot means of a given remote sensing trait can be computed from all four RS (RS1234), the inner RS (RS23),

TABLE 2 Summary of vegetation indices obtained from the VNIR-hyperspectral sensor used in predictive modeling.

Abbreviation	Index Name	Formula	Reference
Carte1	Carter Index 1	R_{695}/R_{420}	Carter (1994)
GNDVI	Green NDVI	$(R_{750} - R_{550})/(R_{750} + R_{550})$	Gitelson et al., 1996
mRENDVI	Modified Red Edge Simple NDVI	$(R_{750} - R_{705})/(R_{750} + R_{705} - 2R_{445})$	Sims and Gamon (2002)
NDVI	Normalized Difference Vegetation Index	$(R_{750} - R_{705})/(R_{750} + R_{705})$	Gitelson and Merzlyak (1994)
OSAVI	Optimized Soil Adjusted Vegetative Index	$(1 + 0.16)(R_{800} - R_{670})/(R_{800} + R_{670} + 0.16)$	Rondeaux et al. (1996)
PRI	Photochemical Reflectance Index	$(R_{531} - R_{570})/(R_{531} + R_{570})$	Gamon et al. (1992)
PSRI	Plant Senescence Reflectance Index	$(R_{680} - R_{500})/R_{750}$	Merzlyak et al. (1999)
SR800680	Simple Ratio Index	R_{800}/R_{680}	Sims and Gamon (2002)
SR700670	Simple Ratio Index	R_{700}/R_{670}	McMurtrey et al. (1994)
VOG1	Vogelmann Red Edge Index	R_{740}/R_{720}	Vogelmann et al. (1993)

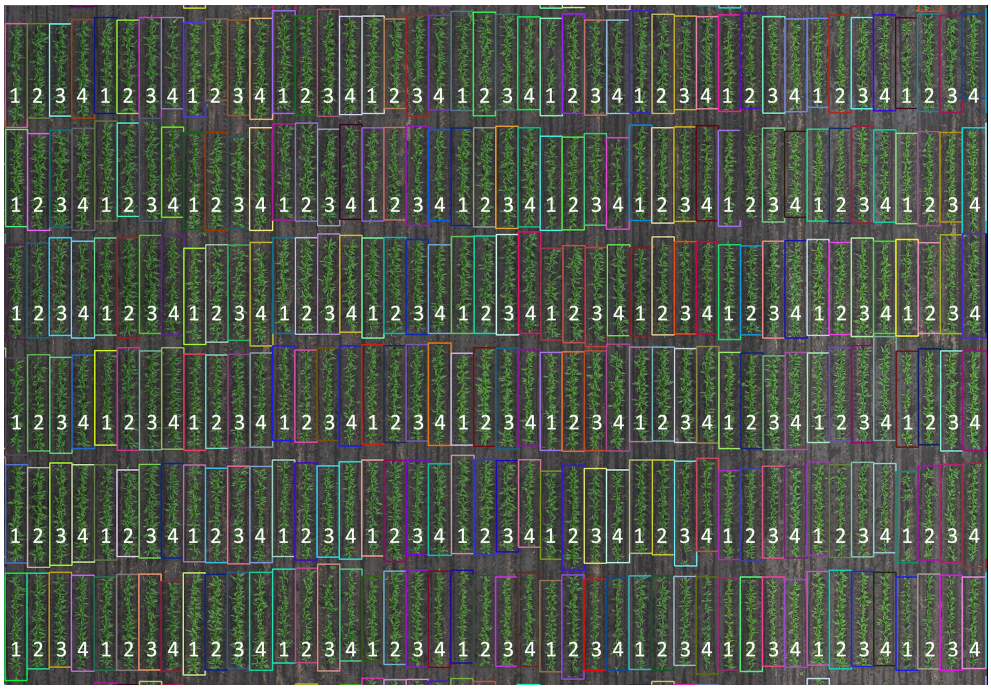


FIGURE 2
Row segmentation from a section of the sorghum experiment in 2018. Plots were planted as four-row plots. Row segments (1, 2, 3, and 4) within each plot were individually segmented. Row segment length was automatically determined based on the proximal and terminal ends of the plants in the individual row segments. Plot trimming was performed by removing 40 cm from the top and bottom of each row segment.

the outer RS (RS14), and individual RS (RS1, RS2, RS3, RS4). For this study, plots were trimmed by removing 40 cm from the two ends of the RS boundary (Supplemental Figure 2).

2.6 Repeatability

Broad-sense heritability was calculated on an entry-mean basis (H^2) to estimate repeatability of representative geometric and spectral remote sensing traits including: CC, CC10, plant height 95th percentile (Height 95th), plot volume, and the normalized difference vegetation index (NDVI) (Nyquist and Baker, 1991; Piepho and Möhring, 2007). Variance components were estimated through restricted maximum likelihood (REML) from Equation 1 using R package 'lme4' (Bates et al., 2015). As remote sensing traits were observed on different dates across years and experiments, variance components were predicted for each year individually using the following model:

$$Y_{ij} = \mu + H_i + R_j + \varepsilon_{ij} \quad (1)$$

where Y_{ij} is the phenotypic measurement of the i^{th} hybrid in the j^{th} rep. Components of the model include μ that represents the grand mean, H_i as the random effect of the i^{th} hybrid, R_j as the fixed effect of the j^{th} replicate, and ϵ_{ij} as the residual error for the i^{th} hybrid in the j^{th} rep. Variance components for the random effects, hybrid and error, were estimated from Equation 1 and used to evaluate repeatability using Equation 2.

$$H^2 = \frac{\sigma_H^2}{\sigma_H^2 + \frac{\sigma_\varepsilon^2}{rep}} \quad (2)$$

where H^2 represents repeatability of a given trait. Hybrid and error variance components are denoted by σ_H^2 and σ_ε^2, respectively. The number of replications (2, sorghum; 3, maize) were *rep* in Equation 2.

2.7 Predictive modeling

Classical machine learning models including Support Vector Regression (SVR) and Partial Least Squares Regression were evaluated, and SVR was selected for further analysis based on its predictive performance as previously shown in Masjedi et al. (2020). Nevertheless, the goal of this study was to understand the impact of row selection and plot trimming on repeatability and predictive models, not to develop an optimum predictive model. SVR is a non-parametric regression technique with no statistical assumptions. SVR transforms the original feature space to find a linear hyperplane in a higher dimension for predictive modeling (Cristianini and Shawe-Taylor, 2000). Optimal values of hyperparameters sigma and cost were determined through cross-validation in a grid search. Parameters evaluated included sigma (0.001, 0.001, 0.01, 0.1) and cost (10, 50, 100, 150, 200, n) where n is the number of features used in the model (Masjedi, 2020). The correlation (r) of predicted yield with known yield in the testing set was used to assess model performance. SVR models were developed

in R package 'caret' (Kuhn, 2008) implementing the model from R package 'kernlab' (Karatzoglou et al., 2004).

Remote sensing traits throughout the growing season were standardized (centered and scaled) and used to predict end-season biomass in sorghum and end-season grain yield in maize. In sorghum, 10-fold cross-validation repeated 100 times was used to assess model accuracy. In maize, 3-fold cross-validation repeated 500 times was used to assess model accuracy. An equal number of temperate and tropical germplasms were selected in each fold of cross validation to ensure similar population structures of training and testing datasets in maize. Plots of the same genotype from different replicates were considered as unique entries, and this was not a factor controlled in cross validation. Maize was repeated more than sorghum because of computational efficiency of the smaller experiment and to improve the precision of the prediction accuracy estimates.

2.8 Statistical analysis

Statistical analyses were performed in R (R Core Team, 2022). For the repeatability of a remote sensing trait from different dates and years, data were grouped based on GDD for degrees of freedom in analysis of variance (ANOVA). Flights were grouped as from 0 to 650 GDD representing vegetative growth stage, 0 to 900 GDD for flowering growth stage, and greater than 900 GDD for grain filling growth stage. Multiple flights in each growth stage interval in each year provide a robust dataset to perform statistical analyses (Table 1). ANOVA was used to determine the presence of significant interaction effects between year, row selection, and plot trimming in maize and sorghum. Where there was a significant difference between year, row selection, plot trimming, or the interaction effects in ANOVA results, a Least Significant Difference Test in R package 'agricolae' (de Mendiburu, 2021) was used to determine which treatments were significantly different at $\rho < 0.05$. Remote sensing data, yield data, and R code used for this study are available at the Purdue University Research Repository (10.4231/PF9S-4G38).

3 Results

3.1 Impact of row selection and plot trimming on repeatability

The repeatability of remote sensing traits was evaluated in multi-year sorghum (Table 3) and maize (Table 4) experiments. The average repeatability of all remote sensing traits was 0.84 in maize and 0.67 in sorghum. The repeatability values of remote sensing traits increased at later stages of development. The average repeatability values in sorghum were 0.54, 0.71, and 0.76 during the vegetative, flowering, and grain filling growth stages respectively. The average repeatability values in maize similarly increased at later growth stages with values of 0.81, 0.84, and 0.87 during the vegetative, flowering, and grain filling growth stages, respectively. Plot trimming and interaction effects involving plot trimming never resulted in a significant change in repeatability in either maize or sorghum.

The interaction effect between row segment and year was significant in sorghum for CC10 ($\rho<0.05$) at flowering stage and NDVI ($\rho< 0.01$) at flowering stage (Supplemental Table 3). The repeatability value of CC10 at flowering growth stage was best in RS1234 in 2018, RS1234 in 2019, and RS23 in 2020. The least repeatable value for CC10 at flowering was from RS4 in 2018, RS3 in 2019, and RS1 in 2020. While the interaction effect was significant for NDVI at flowering stage in sorghum, the most and least repeatable data came from RS1234 and RS4, respectively, in all years of the sorghum experiment. NDVI at the vegetative growth stage was the only trait with a significant ($\rho< 0.001$) row segment x year interaction in maize with RS4 in 2020 being worse than other row segment combinations in either year (Supplemental Table 4).

The row segments used to extract remote sensing data was a significant factor impacting repeatability for ten of the fifteen (five traits at three growth stages) remote sensing traits in sorghum (Table 3). Average repeatability across all remote sensing traits was 0.83 in RS1234, 0.82 in RS23, 0.79 in RS2, 0.79 in RS3, 0.77 in RS14, 0.72 in RS4, and 0.71 in RS1. The highest repeatability values were generally observed for RS1234, but repeatability for RS1234 was only significantly greater than RS23 for NDVI at flowering growth stage. Repeatability from RS14 was significantly lower than repeatability of RS1234 for nine of the fifteen remote sensing traits. Repeatability from a single outer row was significantly less repeatable than RS14 (outer rows) for eight of the fifteen traits for RS1 and six of the fifteen traits for RS4. Often, repeatability of traits measured in the inner row segments (RS2 and RS3) was significantly higher than repeatability of traits from the outer row segments (RS1 and RS4).

Row segment was a significant factor impacting repeatability for nine of the fifteen (five traits at three growth stages) remote sensing traits in maize (Table 4). Average repeatability across all remote sensing traits in all three growth stages was 0.90 in RS1234, 0.89 in RS23, 0.89 in RS3, 0.88 in RS14, 0.87 in RS2, 0.84 in RS1, and 0.83 in RS4. Repeatability of traits from RS1234 and RS23 were never significantly different, and repeatability of traits from RS14 was only significantly lower for PV at the grain filling stage. Repeatability of traits extracted from RS1 and RS4 (outside single rows) was significantly lower than repeatability of traits from RS1234 for seven and eight remote sensing traits, respectively, and significantly less repeatable than RS14 for four and eight remote sensing traits, respectively. Often, repeatability of traits from the inner row segments (RS2 and RS3) was significantly higher than traits measured in the outer row segments (RS1 and RS4).

3.2 Impact of row selection on predictive modeling

End-season sorghum biomass yield averaged 1,737 g m^{-2} (range: 745 to 3,989) in 2018; 1,720 g m^{-2} (range: 1,039 to 2,879) in 2019; and 1,746 g m^{-2} (range: 619 to 2,653) in 2020. The interaction effect

TABLE 3 Repeatability of remote sensing traits from RGB (red-green-blue), LiDAR (light detection and ranging), and VNIR (visible near infrared hyperspectral) sensors in sorghum experiments from 2018, 2019 and 2019.

	Vegetative (0 - 650 GDD)					Flowering (650 - 900 GDD)					Grain Filling (900+ GDD)				
	CC^R	CC 10%^{L}	Height 95%^{L}	Plot Volume^L	$NDVI^V$	CC^R	CC 10%^{L}	Height 95%^{L}	Plot Volume^L	$NDVI^V$	CC^R	CC 10%^{L}	Height 95%^{L}	Plot Volume^L	$NDVI^V$
Mean	0.42	0.37	0.71	0.65	0.56	0.4	0.48	0.94	0.95	0.8	0.56	0.59	0.92	0.92	0.83
Sorghum 2018	0.3 b						0.47 b	0.93 b	0.94 b	0.72 c		–		–	0.8 b
Sorghum 2019	0.5 a						0.46 b	0.96 a	0.95 a	0.81 b		0.64 a		0.92 b	0.83 ab
Sorghum 2020	0.53 a						0.58 a	0.94 ab	0.94 b	0.85 a		0.54 b		0.93 a	0.84 a
RS1234	0.47 a					0.45 a	0.6 a	0.97 a	0.96 a	0.84 a		0.69 a	0.96 a	0.95 a	0.88 a
RS23	0.45 ab					0.45 a	0.56 ab	0.98 a	0.96 a	0.83 b		0.68 ab	0.97 a	0.95 a	0.87 ab
RS14	0.43 abc					0.38 b	0.49 cd	0.93 b	0.94 b	0.82 c		0.57 bc	0.91 b	0.92 b	0.84 b
RS1	0.36 d					0.3 c	0.38 e	0.87 c	0.93 bc	0.77 e		0.48 c	0.84 c	0.88 c	0.74 c
RS2	0.42 abc					0.43 ab	0.51 bc	0.97 a	0.96 a	0.8 d		0.6 ab	0.96 a	0.94 ab	0.83 b
RS3	0.41 bc					0.4 ab	0.44 de	0.97 a	0.96 a	0.8 d		0.63 ab	0.96 a	0.94 ab	0.85 ab
RS4	0.4 cd					0.39 ab	0.39 e	0.9 b	0.91 c	0.74 f		0.47 c	0.87 c	0.88 c	0.78 c
Trim ANOVA	0.722 NS†	0.87 NS	0.993 NS	0.936 NS	0.822 NS	0.294 NS	0.169 NS	0.906 NS	0.835 NS	0.09 NS	0.939 NS	0.575 NS	0.87 NS	0.727 NS	0.322 NS
Year ANOVA	<0.001 ***	0.403 NS	0.216 NS	0.244 NS	0.541 NS	0.12 NS	<0.001 ***	<0.001 ***	0.04 *	<0.001 ***	0.984 NS	0.002 **	0.518 NS	0.011 *	0.025 *
RS ANOVA	<0.001 ***	0.844 NS	0.901 NS	0.83 NS	0.885 NS	<0.001 ***	<0.001 ***	<0.001 ***	<0.001 ***	<0.001 ***	0.987 NS	0.002 **	<0.001 ***	<0.001 ***	<0.001 ***
RS x Y ANOVA	0.292 NS	1 NS	1 NS	1 NS	1 NS	0.355 NS	0.014 *	0.198 NS	0.492 NS	0.002 **	0.907 NS	0.999 NS	0.389 NS	1 NS	0.712 NS

Significant differences were declared using ANOVA between plot trimming (Trim) techniques, years (Y), row segments (RS), and interaction (RS x Y) effects. Where ANOVA was significant, letters following the traits indicate significant differences between treatments at $\rho < 0.05$. The same letters signify no significant differences between treatments. Where there were not significant differences in ANOVA, values were removed and the mean can be used for the trait.

CC, Canopy Cover; R, Trait from RGB sensor; L, Trait from LiDAR sensor; V, Trait from VNIR-hyperspectral sensor; RS, Row Segment; Y, Year.

† = ANOVA significance based on p-value: >0.05= NS,<0.05 = *,<0.01 = **,<0.001 = ***.

10.3389/fpls.2023.1202536

TABLE 4 Repeatability of remote sensing traits from RGB (red-green-blue), LiDAR (light detection and ranging), and VNIR (visible near infrared hyperspectral) sensors in maize experiments from 2020 and 2021.

	Vegetative (0 - 650 GDD)					Flowering (650 - 900 GDD)					Grain Filling (900+ GDD)				
	CC^R	CC 10%^{L}	Height 95%^{L}	Plot Volume^L	$NDVI^V$	CC^R	CC 10%^{L}	Height 95%^{L}	Plot Volume^L	$NDVI^V$	CC^R	CC 10%^{L}	Height 95%^{L}	Plot Volume^L	$NDVI^V$
Mean	0.83	0.71	0.83	0.79	0.86	0.75	0.7	0.94	0.92	0.91	0.77	0.84	0.97	0.95	0.84
Maize 2020	0.85 a	0.81 a		0.82 a			0.86 a		0.89 b					0.94 b	0.77 b
Maize 2021	0.81 b	0.57 b		0.75 b			0.53 b		0.95 a					0.96 a	0.91 a
RS1234	0.87 a		0.86 a	0.82 a	0.89 a		0.8 a		0.94 a			0.9 a	0.98 ab	0.97 a	
RS23	0.86 ab		0.86 a	0.84 a	0.88 a		0.75 ab		0.94 a			0.9 a	0.98 a	0.97 a	
RS14	0.85 ab		0.83 a	0.78 a	0.89 a		0.75 ab		0.92 a			0.85 ab	0.97 bc	0.95 b	
RS1	0.78 c		0.83 a	0.79 a	0.9 a		0.66 bc		0.91 a			0.78 bc	0.94 d	0.94 c	
RS2	0.82 bc		0.86 a	0.84 a	0.86 a		0.63 bc		0.94 a			0.86 ab	0.98 ab	0.96 a	
RS3	0.82 abc		0.84 a	0.82 a	0.87 a		0.76 ab		0.93 a			0.88 a	0.98 a	0.96 a	
RS4	0.79 c		0.76 b	0.65 b	0.74 b		0.54 c		0.87 b			0.75 c	0.97 c	0.91 d	
Trim ANOVA	0.907 NS†	0.725 NS	0.812 NS	0.807 NS	0.879 NS	0.904 NS	0.74 NS	0.827 NS	0.923 NS	0.567 NS	0.951 NS	0.722 NS	0.792 NS	0.328 NS	0.937 NS
Year ANOVA	0.009 **	<0.001 ***	0.077 NS	0.029 *	0.058 NS	0.131 NS	<0.001 ***	0.072 NS	<0.001 ***	0.192 NS	0.081 NS	0.328 NS	0.938 NS	<0.001 ***	<0.001 ***
RS ANOVA	0.003 **	0.617 NS	0.035 *	0.003 **	<0.001 ***	0.134 NS	0.005 **	0.143 NS	0.023 *	0.569 NS	0.27 NS	0.005 **	<0.001 ***	<0.001 ***	0.851 NS
RS x Y ANOVA	0.342 NS	0.93 NS	0.59 NS	0.968 NS	<0.001 ***	0.262 NS	0.093 NS	0.967 NS	0.258 NS	0.585 NS	0.996 NS	0.756 NS	0.177 NS	0.202 NS	0.996 NS

Significant differences were declared using ANOVA between plot trimming (Trim) techniques, years (Y), row segments (RS), and interaction (RS x Y) effects. Where ANOVA was significant, letters following the traits indicate significant differences between treatments at $\rho < 0.05$. The same letters signify no significant differences between treatments. Where there were not significant differences in ANOVA, values were removed and the mean can be used for the trait.

CC, Canopy Cover; R, From RGB sensor; L, From LiDAR sensor; V, From VNIR-hyperspectral sensor; RS, Row Segment; Y, Year.

† = ANOVA significance based on p-value: >0.05= NS,<0.05 = *,<0.01 = **,<0.001 = ***.

10.3389/fpls.2023.1202536

between row segment and year was significant for both sorghum and maize. Thus, the effect of row segment was assessed for each year individually (Figure 3). Prediction accuracy of sorghum biomass ranged from 0.74 to 0.76 in 2018, 0.64 to 0.68 in 2019, and 0.69 to 0.75 in 2020. The highest prediction accuracy in 2018 used data from RS1234, though not significantly different from RS23. In 2019 and 2020, prediction accuracy using data from RS23 was significantly greater than all other row segment combinations. Remote sensing data from outer rows (RS14, RS1, or RS4) generally resulted in a significant decrease in prediction accuracy compared to remote sensing data including inner rows (RS23, RS1234, RS2, or RS3), except in 2018 where prediction accuracy from RS14 and RS23 and RS2, RS3, and RS4 were not significantly different.

Average maize grain yield was 10.9 Mg ha^{-1} (range: 3.2 to 14.8) in 2020 and 9.3 Mg ha^{-1} (range: 4.8 to 12.9) in 2021. Prediction accuracy of maize grain yield ranged from 0.53 to 0.68 in 2020 and 0.48 to 0.66 in 2021. Remote sensing data from RS3 and RS4 resulted in a prediction accuracy significantly higher than all other row segment combinations evaluated in 2020. Prediction accuracy using remote sensing data from RS23 was significantly greater than RS1234 which was significantly greater than RS14. The prediction accuracy of maize grain yield using data from RS1 was significantly lower than all other row segment combinations in 2020. However, prediction accuracy of maize grain yield in 2021 from RS1234 was significantly greater than all other RS combinations. Prediction accuracy from remote sensing data from RS14 was significantly greater than RS23. Prediction accuracy from RS3 was not significantly different from RS23, but it was significantly greater than RS1, RS2, and RS4. Remote sensing data from RS14 (outer rows) resulted in a significantly higher prediction accuracy when compared to RS1 and RS4 (single outer rows).

4 Discussion

4.1 Implications of row selection and plot trimming in remote sensing

Phenotyping capabilities have been the limiting factor to greater genetic gain in breeding programs based on the time consuming, laborious, costly, and often destructive nature of the endeavor (Cobb et al., 2013; Yang et al., 2017). Additionally, some traits, such as canopy cover, are difficult to quantitatively measure and are reduced to subjective estimates. High-throughput phenotyping through remote sensing can be used to produce large quantities of data on breeding plots that can be used *per se* or in end-season yield predictions to increase genetic gain in plant breeding. Key research questions related to processing of remote sensing data in breeding trials include the impact of row selection, plot trimming, and number of RS in a plot have not been quantitatively explored.

The border effect has been well documented in previous agronomic studies and has led to an emphasis on measurements of traits from the inner rows of multi-row plots (Bird, 1929; Genter, 1958; Gomez, 1972; Kramer et al., 1982; Bowman, 1989; Ceccarelli and Grando, 1996; Petersen, 1994; Reynolds – Braun, 2022) (Supplemental Figure 1). Nevertheless, border effect of remote sensing traits has been less documented. In this experiment, the border effect was evaluated based on repeatability and yield prediction accuracy using representative remote sensing traits related to geometric and chemistry-related responses (Tables 3, 4; Figure 3). The impact of border rows can be observed by comparing all rows RS1234, inner rows (RS23), and outer rows (RS14) or comparing single inner rows (RS2 and RS3) with single outer rows (RS1 and RS4). Repeatability of remote sensing traits was often improved or not significantly different when considering whole

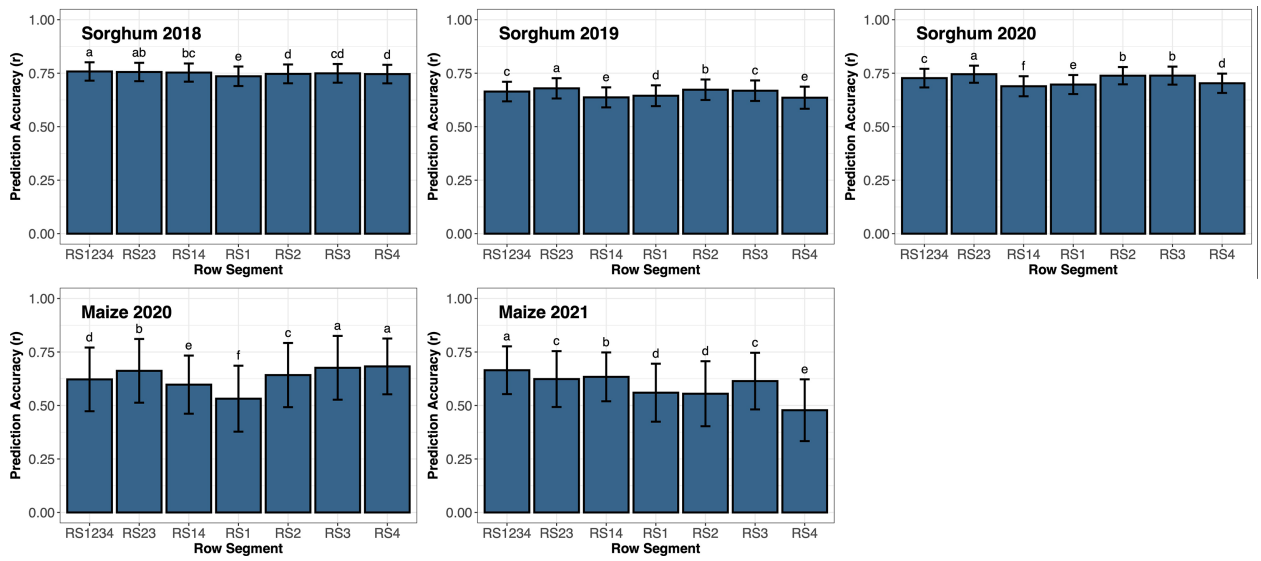


FIGURE 3
Prediction accuracy based on correlation (r) of support vector regression using all remote sensing traits from various row segments throughout the growing season to predict end-season biomass in sorghum or grain yield in maize. Error bars represent the standard deviation of the prediction accuracies. Letters above the bars represent significant differences between row segments using LSD ($\rho < 0.05$). Different letters represent significant differences between different row segments.

plots rather than inner-plot rows. Prediction accuracy of end-season yield was often improved by considering just the inner-rows of a four-row plot. These results agree with Zhang et al. (2019) where soybean yield was best predicted when 20 to 80% of the plot area was used as it minimized border effect. Since yield was harvested from the inner rows in our study, it is possible that prediction accuracy from the inner-row segments was favorably biased and different results could have been obtained if yield had been determined using the harvest of all rows. Nevertheless, as yield is often harvested from the inner-plot rows in breeding experiments, this comparison has the most relevance. A follow-up study of interest could be to harvest each row of a plot individually to better associate yield and remote sensing data collection.

The significance of row selection indicated the presence of a border effect, which suggested that trimming the two ends of the plots could similarly impact results. Anderson et al. (2019) did not use any plot trimming technique. Malambo et al. (2017) trimmed 20 cm from the proximal and terminal ends of plots with a length 7.6 m in maize and 5.6 m in sorghum. Masjedi et al. (2018) trimmed 40 cm from the proximal and terminal ends of each plot of length 3 m in sorghum. Krause et al. (2019) trimmed 50 cm from the proximal and terminal ends of each plot in wheat experiments with plot lengths of 2.8 to 4 m. Tirado et al. (2020) created 20 bins within each row segment and evaluated the middle 12 bins for their analysis in maize with plot lengths of 3.65 m. While various methods have been used, none of these studies evaluated differences between plot trimming techniques. In our study, there was not a significant difference in repeatability values in all remote sensing data with and without 40 cm removed from the proximal and terminal ends of the plot boundary (Tables 3, 4). The row selection grid in our experiment was created early in the growing season (V2-V4) when row segments were clearly visually separated. The plot segmentation process created segments that were defined by the end plants in each row. While plants continue to grow throughout the growing season, the grid remained unchanged. One hypothesis that explained the non-significant plot trimming effect was that plants grew outside of the bounding boxes developed early in the growing season, and this material was not included in the bounding box for trait measurements. Nevertheless, there is limited downside to plot trimming and it remains an effective measure to eliminate potential alley effect.

The number of row segments in a plot is an important consideration in experimental design constrained by the number of genotypes, replications, seed, and land area available. Generally, fewer row segments are used in the early stages of a breeding program to evaluate many genotypes and remove unfavorable genotypes based on highly heritable traits (i.e. disease resistance, plant height, etc.) (Acquaah, 2012). More row segments are used in advanced stages of a breeding program where yield performance is of primary concern. Nevertheless, remote sensing could enable researchers to evaluate plots with fewer row segments with greater accuracy and remove the necessity of plots with more row segments. In our study, RS1 and RS4 (single outer rows) and RS14 (outer rows) best represent one-row and two-row plots, respectively, as the border effect was present in these row segments. Remote sensing trait repeatability and prediction accuracy of either biomass in sorghum or grain yield in maize were generally improved as the number of RSs increased in both maize and sorghum. While this study only evaluated a maximum of four-row plots, it is likely that repeatability of remote sensing traits and prediction accuracy could be further increased using plots with more RS.

4.2 Implications and future work

The results of this study indicate that basic agronomic principles should be implemented to maximize the value of remote sensing data for plant breeding purposes. Plot trimming and excluding exterior rows should be used to limit the border effect from the alleys and neighboring plots. The number of RS in a plot is an important consideration when designing an experiment and should be increased when possible. Nevertheless, this experiment did not evaluate the tradeoff value of increasing the number of RS in a plot in comparison to increasing the number of replications in the experiment. Additionally, future studies should evaluate larger plot sizes extending beyond the four-row plot size used in this study. Finally, while plot trimming was not a significant factor in this study, it should be evaluated in experiments with different plot lengths or using other row segmentation techniques.

5 Conclusions

Remote sensing is a rapidly advancing area of phenomics enabling an increase in the amount of in-season data that can be evaluated. Within breeding programs, this data can be used *per se* or to predict end-season yield in the growing season. This study was performed to evaluate the importance of a border effect, plot trimming, and number of row segments used in remote sensing data in maize and sorghum. Generally, repeatability was improved when remote sensing data from more row segments was used in the analysis, and prediction accuracy was improved when excluding outer rows. While results in this study obtained using trimmed plots were not significantly different from when they were not trimmed, it should be considered in future studies to minimize any potential alley effect. Implementing these basic practices could help to maximize the value of remote sensing data and increase selection efficiency in a breeding program.

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 conceived, initiated, and coordinated all of the experiments. MC, ED, and AH collected and processed the remote sensing data

from the field trials. ST and NC analyzed the data and wrote the paper with contributions from the authors. All authors contributed to the article and approved the submitted version.

Funding

This project was partially funded by grants from the Heat Tolerant Maize for Asia project supported by the Feed the Future Initiative of USAID organized by the International Maize and Wheat Improvement Center (CIMMYT) (Grant no. 15100072), the U.S. Department of Energy ARPA-E Program (Award Number DE-AR0001135), and the USDA National Institute of Food and Agriculture Hatch Program (Number 1023186).

Acknowledgments

We thank Andrew Linvill, Eugene Glover, and the Tuinstra Lab for help in collecting agronomic data in field trials. We thank the MC, AH, and ED labs for help with collection and processing of remote sensing data in the maize and sorghum field trials.

Conflict of interest

Authors MT, MC, ED, and AH are founders of the agriculture startup company GRYFN.

The remaining 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.2023.1202536/full#supplementary-material

References

Acquaah, G. (2012). *Principles of plant genetics and breeding (2nd Ed)* (Maryland: John Wiley & Sons).

Anche, M., Kaczmar, N., Morales, N., Clohessy, J., Ilut, D., Gore, M., et al. (2020). Temporal covariance structure of multi-spectral phenotypes and their predictive ability for end-of-season traits in maize. *Theor. Appl. Genet.* 133 (10), 2853–2868. doi: 10.1007/s00122-020-03637-6

Anderson, S., Murray, S., Malambo, L., Ratcliff, C., Popescu, S., Cope, D., et al. (2019). Prediction of maize grain yield before maturity using improved temporal height estimates of unmanned aerial systems. *Plant Phenome J.* 2 (1), 1–15. doi: 10.2135/tppj2019.02.0004

Araus, J., and Cairns, J. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19 (1), 52–61. doi: 10.1016/j.tplants.2013.09.008

Araus, J., Kefauver, S., Zaman-Allah, M., Olsen, M., and Cairns, J. (2018). Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23 (5), 451–466. doi: 10.1016/j.tplants.2018.02.001

Barmeier, G., and Schmidhalter, U. (2016). High-throughput phenotyping of wheat and barley plants grown in single or few rows in small plots using active and passive spectral proximal sensing. *Sensors* 16 (11), 1–14. doi: 10.3390/s16111860

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67 (1), 1–48. doi: 10.18637/jss.v067.i01

Bernardo, R. (2014). *Essentials of plant breeding* (Minnesota: Stemma Press).

Bird, J. N. (1929). *Competition between adjacent rows of corn* (Canada: McGill University).

Bowman, D. T. (1989). Plot configuration in corn yield trials. *Crop Sci.* 29 (5), 1202–1206. doi: 10.2135/cropsci1989.0011183X002900050023x

Carter, G. A. (1994). Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. *Int. J. Remote Sens.* 15 (3), 697–703. doi: 10.1080/01431169408954109

Ceccarelli, S., and Grando, S. (1996). Drought as a challenge for the plant breeder. *Plant Growth Regul.* 20 (2), 149–155. doi: 10.1007/BF00024011

Chen, Y. (2019). Estimating plant phenotypic traits from RGB imagery. purdue university graduate school. *Thesis*, 37–41. doi: 10.25394/PGS.10312115.v1

Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., and McCouch, S. (2013). Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theor. Appl. Genet.* 126, 867–887. doi: 10.1007/s00122-013-2066-0

Crain, J., Mondal, S., Rutkoski, J., Singh, R. P., and Poland, J. (2018). Combining high-throughput phenotyping and genomic information to increase prediction and selection accuracy in wheat breeding. *Plant Genome* 11 (1), 1–14. doi: 10.3835/plantgenome2017.05.0043

Cristianini, N., and Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods* (Cambridge: Cambridge university press). doi: 10.1097/00003246-198507000-00006

de Mendiburu, F. (2021) *Agricolae: statistical procedures for agricultural research. r package version 1.3-5.* Available at: https://CRAN.R-project.org/package=agricolae.

Furbank, R., and Tester, M. (2011). Phenomics - technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16 (12), 635–644. doi: 10.1016/j.tplants.2011.09.005

Gamon, J. A., Peñuelas, J., and Field, C. B. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens. Environ.* 44 (1), 35–44. doi: 10.1016/0034-4257(92)90059-S

Genter, C. (1958). Plot competition between corn hybrids. *Agron. J.* 50 (4), 205–206. doi: 10.2134/agronj1958.00021962005000040010x

Gilmore, E., and Rogers, J. (1958). Heat units as a method of measuing maturity in corn. *Agron. J.* 50 (10), 62–81. doi: 10.2134/agronj1958.00021962005000100014x

Gitelson, A. A., Kaufman, Y. J., and Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* 58 (3), 289–298. doi: 10.1016/S0034-4257(96)00072-7

Gitelson, A., and Merzlyak, M. (1994). Quantitative estimation of chlorophyll-a using reflectance spectra: experiments with autumn chestnut and maple leaves. *J. Photochem. Photobiol. B: Biol.* 22 (3), 247–252. doi: 10.1016/1011-1344(93)06963-4

Gomez, K. A. (1972). Border effects in rice experimental plots. II. varietal competition. *Exp. Agric.* 8 (4), 295–298. doi: 10.1017/S001447970000541X

Habib, A., Zhou, T., Masjedi, A., Zhang, Z., Flatt, J. E., and Crawford, M. (2018). "Boresight calibration of gnss/ins-assisted push-broom hyperspectral scanners on UAV platforms," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 11. 1734–1749. doi: 10.1109/JSTARS.2018.2813263

Hasheminasab, S., Zhou, T., and Habib, A. (2020). GNSS/INS-assisted structure from motion strategies for UAV-based imagery over mechanized agricultural fields. *Remote Sens.* 12 (3), 1–37. doi: 10.3390/rs10121952

Hassani, K., Gholizadeh, H., Taghvaeian, S., Natalie, V., Carpenter, J., and Jacob, J. (2023). Application of UAS-based remote sensing in estimating winter wheat phenotypic traits and yield during the growing season. *PFG* 91, 77–90. doi: 10.1007/s41064-022-00229-5

He, F., Zhou, T., Xiong, W., Hasheminnasab, S., and Habib, A. (2018). Automated aerial triangulation for UAV-based mapping. *Remote Sens.* 10 (12), 1–34. doi: 10.3390/rs10121952

Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004). Kernlab – an S4 package for kernel methods in r. *J. Stat. Softw.* 11 (9), 1–20. doi: 10.18637/jss.v011.i09

Kramer, T., Van Ooijen, J. W., and Spitters, C. J. T. (1982). Selection for yield in small plots of spring wheat. *Euphytica* 31 (3), 549–564. doi: 10.1007/BF00039193

Krause, M., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R., et al. (2019). Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3* 9 (4), 1231–1247. doi: 10.1534/g3.118.200856

Kuhn, M. (2008). Building predictive models in r using the caret package. *J. Stat. Softw.* 28 (5), 1–26. doi: 10.18637/jss.v028.i05

LaForest, L., Hasheminasab, S., Zhou, T., Flatt, E., and Habib, A. (2019). New strategies for time delay estimation during system calibration for UAV-based GNSS/INS-assisted imaging systems. *Remote Sens.* 11 (15), 1–36. doi: 10.3390/rs11151811

Li, J., Shi, Y., Veeranampalayam-Sivakumar, A.-N., and Schachtman, D. P. (2018). Elucidating sorghum biomass, nitrogen and chlorophyll contents with spectral and morphological traits derived from unmanned aircraft system. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01406

Lin, Y. C., and Habib, A. (2021). Quality control and crop characterization framework for multi-temporal UAV LiDAR data over mechanized agricultural fields. *Remote Sens. Environ.* 256, 112299. doi: 10.1016/j.rse.2021.112299

Maes, W. H., and Steppe, K. (2019). Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends Plant Sci.* 24 (2), 152–164. doi: 10.1016/j.tplants.2018.11.007

Malambo, L., Popescu, S., Murray, S., Putman, E., Pugh, N., Horne, D., et al. (2017). Multitemporal field-based plant height estimation using 3D point clouds generated from small unmanned aerial systems high-resolution imagery. *Int. J. Appl. Earth Observ. Geoinform.* 64 (2018), 31–42. doi: 10.1016/j.jag.2017.08.01

Masjedi, A., Crawford, M., Carpenter, N., and Tuinstra, M. (2020). Multi-temporal predictive modelling of sorghum biomass using UAV-based hyperspectral and LiDAR data. *Remote Sens.* 12, 1–35. doi: 10.3390/rs12213587

Masjedi, A., Zhao, J., Thompson, A. M., Yang, K., Flatt, J. E., Crawford, M. M., et al. (2018). "Sorghum biomass prediction using UAV-based remote sensing data and crop model simulation," in *38th IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (Valencia Spain. Piscataway, NJ United States). 7719–7722. doi: 10.1109/IGARSS.2018.8519034

Mathews, A. J., and Jensen, J. L. (2013). Visualizing and quantifying vineyard canopy LAI using an unmanned aerial vehicle (UAV) collected high density structure from motion point cloud. *Remote Sens.* 5, 2164–2183. doi: 10.3390/rs5052164

McMurtrey, J. E., Chappelle, E. W., Kim, M. S., Meisinger, J. J., and Corp, L. A. (1994). Distinguishing nitrogen fertilization levels in field corn (*Zea mays l.)* with actively induced fluorescence and passive reflectance measurements. *Remote Sens. Environ.* 47 (1), 36–44. doi: 10.1016/0034-4257(94)90125-2

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. *Physiol. Plantarum* 106 (1), 135–141. doi: 10.1034/j.1399-3054.1999.106119.x

Moreira, F., Oliveira, H., Lopez, M., Abughali, B., Gomes, G., Cherkauer, K., et al. (2021). High-throughput phenotyping and random regression models reveal temporal genetic control of soybean biomass production. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021/715983

Nyquist, W., and Baker, R. (1991). Estimation of heritability and prediction of selection response in plant populations. *Crit. Rev. Plant Sci.* 10 (3), 235–322. doi: 10.1080/07352689109382313

Petersen, R. G. (1994). *Agricultural field experiments: design and analysis* (Boca Raton: CRC Press).

Piepho, H. P., and Möhring, J. (2007). Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* 177, 1881–1888. doi: 10.1534/genetics.107.074229

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

Ravi, R., Hasheminasab, S. M., Zhou, T., Masjedi, A., Quijano, K., Flatt, J. E., et al. (2019). "UAV-based multi-sensor multi-platform integration for high throughput phenotyping," in *Proc. SPIE 11008, Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV*, 110080E. p. 13. doi: 10.1117/12.2519190

Ravi, R., Lin, Y.-J., Elbahnasawy, M., Shamseldin, T., and Habib, A. (2018). "Simultaneous system calibration of a multi-lidar multicamera mobile mapping platform," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 11. 1694–1714. doi: 10.1109/JSTARS.2018.2812796

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PloS One* 8 (6), 1–8. doi: 10.1371/journal.pone.0066428

R Core Team (2022) *R: a language and environment for statistical computing*. Available at: http://www.r-project.org/ (Accessed January 20, 2022).

Reynolds, M. P., and Braun, H. J. (2022). "Wheat improvement," in *Wheat improvement*. Eds. M. P. Reynolds and H. J. Braun (Cham: Springer). doi: 10.1007/978-3-030-90673-3_1

Rondeaux, G., Steven, M., and Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* 55 (2), 95–107. doi: 10.1016/0034-4257(95)00186-7

Sims, D. A., and Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures, and developmental stages. *Remote Sens. Environ.* 81 (2-3), 337–354. doi: 10.1016/S0034-4257(02)00010-X

Tattaris, M., Reynolds, M. P., and Chapman, S. C. (2016). A direct comparison of remote sensing approaches for high-throughput phenotyping in plant breeding. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01131

Tirado, S., Hirsch, C., and Springer, N. (2020). UAV-based imaging platform for monitoring maize growth throughout development. *Plant Direct* 2020 (00), 1–11. doi: 10.1002/pld3.230

Toledo, C. A., Crawford, M., and Vyn, T. (2022). "Maize yield prediction based on multi-modality remote sensing and LSTM models in nitrogen management practice trials," in *IEEE 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)* Rome, Italy. pp. 1–7. doi: 10.1109/WHISPERS56178.2022.9955086

Vogelmann, J. E., Rock, B. N., and Moss, D. M. (1993). "Red edge spectral measurements from sugar maple leaves," in *International Journal of Remote Sensing*, Vol. 14. 1563–1575. doi: 10.1080/01431169308953986

Watanabe, K., Guo, W., Arai, K., Takanashi, H., Kajiya-Kanegae, H., Kobayashi, M., et al. (2017). High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00421

Yang, C., Baireddy, S., Cai, E., Crawford, M., and Delp, E. J. (2021a). "Field-based plot extraction using UAV RGB images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1390–1398. doi: 10.48550/arXiv.2109.00632

Yang, K., Chapman, S., Carpenter, N., Hammer, G., McLean, G., Zheng, B., et al. (2021b). Integrating crop growth models with remote sensing for predicting biomass yield of sorghum. *In Silico Plants* 3, 1–19. doi: 10.1093/insilicoplants/diab001

Yang, C.-M., Cheng, C.-H., and Chen, R.-K. (2007). Changes in spectral characteristics of rice canopy infested with brown planthopper and leaffolder. *Crop Sci.* 47, 329–335. doi: 10.2135/cropsci2006.05.0335

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. doi: 10.3389/fpls.2017.01111

Zaman-Allah, M., Vergara, O., Araus, J. L., Tarekegne, A., Magorokosho, C., Zarco-Tejada, P. J., et al. (2015). Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. *Plant Methods* 11 (35), 1–10. doi: 10.1186/s13007-015-0078-2

Zarco-Tejada, P. J., González-Dugo, V., and Berni, J. A. (2012). Fluorescence, temperature and narrow- band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera. *Remote Sens. Environ.* 117, 322–337. doi: 10.1016/j.rse.2011.10.007

Zhang, X., Zhao, J., Yang, G., Liu, J., Cao, J., Li, C., et al. (2019). Establishment of plot-yield prediction models in soybean breeding programs using UAV-based hyperspectral remote sensing. *Remote Sens.* 11 (23), 2752. doi: 10.3390/rs11232752

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 26 June 2023
DOI 10.3389/fpls.2023.1146490

Check for updates

OPEN ACCESS

EDITED BY
Jiangang Liu,
Chinese Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Rui Xu,
University of Georgia, United States
Jieli Duan,
South China Agricultural University, China

*CORRESPONDENCE
Yang Zhou
zybuaa@163.com
Wei Long
ylslongwei@caf.ac.cn

RECEIVED 17 January 2023
ACCEPTED 05 June 2023
PUBLISHED 26 June 2023

CITATION
Zhou H, Zhou Y, Long W, Wang B, Zhou Z and Chen Y (2023) A fast phenotype approach of 3D point clouds of Pinus massoniana seedlings.
Front. Plant Sci. 14:1146490.
doi: 10.3389/fpls.2023.1146490

COPYRIGHT
© 2023 Zhou, Zhou, Long, Wang, Zhou 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 fast phenotype approach of 3D point clouds of Pinus massoniana seedlings

Honghao Zhou[1,2], Yang Zhou[1,2]*, Wei Long[3]*, Bin Wang[3], Zhichun Zhou[3] and Yue Chen[4]

[1]College of Electronic and Information Engineering, Zhejiang University of Science and Technology, HangZhou, ZheJiang, China, [2]College of Biological and Chemical Engineering, Zhejiang University of Science and Technology, HangZhou, ZheJiang, China, [3]Research Institute of Subtropical Forestry, Chinese Academy of Forestry, HangZhou, ZheJiang, China, [4]Horticulture Institute, Zhejiang Academy of Agricultural Sciences, HangZhou, ZheJiang, China

The phenotyping of Pinus massoniana seedlings is essential for breeding, vegetation protection, resource investigation, and so on. Few reports regarding estimating phenotypic parameters accurately in the seeding stage of Pinus massoniana plants using 3D point clouds exist. In this study, seedlings with heights of approximately 15-30 cm were taken as the research object, and an improved approach was proposed to automatically calculate five key parameters. The key procedure of our proposed method includes point cloud preprocessing, stem and leaf segmentation, and morphological trait extraction steps. In the skeletonization step, the cloud points were sliced in vertical and horizontal directions, gray value clustering was performed, the centroid of the slice was regarded as the skeleton point, and the alternative skeleton point of the main stem was determined by the DAG single source shortest path algorithm. Then, the skeleton points of the canopy in the alternative skeleton point were removed, and the skeleton point of the main stem was obtained. Last, the main stem skeleton point after linear interpolation was restored, while stem and leaf segmentation was achieved. Because of the leaf morphological characteristics of Pinus massoniana, its leaves are large and dense. Even using a high-precision industrial digital readout, it is impossible to obtain a 3D model of Pinus massoniana leaves. In this study, an improved algorithm based on density and projection is proposed to estimate the relevant parameters of Pinus massoniana leaves. Finally, five important phenotypic parameters, namely plant height, stem diameter, main stem length, regional leaf length, and total leaf number, are obtained from the skeleton and the point cloud after separation and reconstruction. The experimental results showed that there was a high correlation between the actual value from manual measurement and the predicted value from the algorithm output. The accuracies of the main stem diameter, main stem length, and leaf length were 93.5%, 95.7%, and 83.8%, respectively, which meet the requirements of real applications.

KEYWORDS

3D point cloud, phenotyping, skeletonization, Pinus massoniana seedlings, slicing

Introduction

Pinus massoniana is the main timber tree species in southern China; its roots are rich in resin and can be used for flavor synthesis, which is widely used in the construction, papermaking, and artificial fiber industries. After the standardized cultivation of Pinus massoniana seedlings, the obtained adult plants provide high economic benefits. Pinus massoniana is one of the main tree species for afforestation in barren hills in China, and its wood is also used to make furniture and landscape decorations. At present, one of the most important problems in ecology is the local-scale coexistence mechanism driving plants of the same genus (Silvertown, 2004). By researching the phenotypic parameters of Pinus massoniana, the coexistence mechanism can be investigated, which is significant in ecological research. Simultaneously, the phenotypic morphology of Pinus massoniana seedlings can be used to predict the future growth of plants by selecting healthy seeds for cultivation. However, manually measuring phenotypic parameters is time-consuming and laborious and errors can easily occur. Three-dimensional (3D) plant models and automatic phenotypic algorithms provide an efficient and convenient method for plant structural digitization. Therefore, developing an automated method to obtain plant phenotypic parameters from 3D plant models greatly improves measurement efficiency.

In phenotypic applications, two-dimensional (2D) images or 3D point clouds are widely used to characterize the individual morphology of plants (Vázquez-Arellano et al., 2016). 2D images have been widely used in phenotypic analysis; however, due to the limitation of dimensions, 2D images cannot accurately include plant information as well as 3D data (Kaminuma et al., 2004; Gibbs et al., 2018). In particular, phenotype measurements of the leaf area, stem diameter, stem volume, and other parameters require 3D information on plants to obtain more accurate results. Usually, there are two methods of obtaining the 3D phenotypic data of plants. One method is to reconstruct 3D models by synthesizing multi-view 2D images (Park, 2005), and this method has been successfully applied to the 3D reconstruction of wheat and paddies (Pound et al., 2014). However, the limitation of this method is that the 3D reconstruction of the plant surface lacks information from texture and light changes (de Moraes Frasson and Krajewski, 2010). Another approach is to obtain the 3D modeling of plant phenotypes directly from 3D-based sensors, such as lidar, 3D laser scanners, time-of-flight (ToF) cameras, and structured light cameras. These devices have the advantages of high precision, high signal update, and strong robustness. Applications of lidar include extracting plant skeletons from images captured by lidar scanners to estimate the leaf length, leaf inclination, leaf tip length, leaf azimuth (Jimenez-Berni et al., 2018; Wu et al., 2019), leaf area index (LAI) (Zhao and Popescu, 2009), and stand volume (Dziubich et al., 2016). Laser scanners are used to estimate the leaf area, leaf angle, chlorophyll content (Eitel et al., 2010), plant height, leaf width, and main stem volume (Paulus et al., 2014), while the ToF camera is used to evaluate plant traits (Xiang et al., 2019) and classify crops (Li and Tang, 2018). Some research groups have reported that structured light cameras are also a useful tool to extract plant phenotypic data (Nguyen et al., 2015) and predict farm crop growth conditions (Rosell-Polo et al., 2015).

According to the actual environment and the phenotypic features of plants and compared with the 3D images of plants obtained by cameras based on different principles, it was found that the 3D images obtained by ToF cameras were more suitable for our study. The Microsoft Azure Kinect camera, which has high color and depth resolution and performs well in displaying 3D images of Pinus massoniana in plant phenotype applications, is a common camera that is based on the ToF principle. Teng et al. (2021) developed a 3D image acquisition system for oilseed rape plants using an Azure Kinect camera. This system collected point cloud images from six angles and obtained complete point cloud images of plants by rotating the registration. Combined with this system and the hardware facilities of the experimental environment, a set of similar image acquisition devices that obtain three-dimensional images of different plants through non-destructive methods was built.

In agricultural and forestry applications, such as plant biomass analysis, some phenotypic parameters, such as plant height, stem diameter, and leaf length, are important indicators for evaluating plant health, growth status, and effective photosynthesis ability. Because of the morphological characteristics of Pinus massoniana, its leaves grow along the main stem. To obtain the phenotypic parameters of leaves and stems, separation is very important in the whole extraction process. Some algorithms, such as the locally convex connected patches (LCCP) algorithm (Hu et al., 2020), random sample consensus (RANSAC) cylinder fitting algorithm (Fischler and Bolles, 1981), color-based region growth segmentation algorithm (Tang, 2010), curvature-based region growth segmentation algorithm (Besl and Jain, 1988), and skeletonization-based methods, are widely used in leaf and stem segmentation. Wang et al (2020). used the LCCP algorithm to segment the registered vegetable seedlings and then calculated the length, width, and surface area of segmented leaves. However, the limitation of this method is that some leaves cannot be separated from the main stem and require manual segmentation. Another segmentation method is to use RANSAC to fit the cylinder as the main stem of plants (Ghahremani et al., 2021). This method has a good effect on direct-stem plants; however, the main stem of Pinus massoniana is bent, and part of the main stem can be extracted by this method. This method also needs to manually set the RANSAC searching radius without knowing any main stem information, which is still difficult to perform. The application of curvature-based region growing segmentation is to calculate the curvature of the leaf and stem and set a threshold to separate the leaf and stem (Lin et al., 2016). This method will set different thresholds for different individuals of the same plant; thus, this method cannot be universally used. Regarding skeletonization, recent developments adopt slice clustering as the skeleton points of a plant. Then, a Hough plane (Dalitz et al., 2017) is searched according to the skeleton points, and the distance threshold to the Hough plane is used to find the main stem (Xiang et al., 2019). Another application of skeletons is to use the Laplace transform (Cao

et al., 2010) to shrink the point cloud of plants to obtain the skeleton graph (Wu et al., 2019). There are also some novel algorithms, such as the raindrop algorithm, that also achieve leaf and stem separation (Zermas et al., 2020). All the above algorithms provide a reference for the phenotypic analysis of Pinus massoniana seedlings.

Since they have a special morphological structure, Pinus massoniana has significant structural differences from traditional woody plants. At present, most of the phenotypic measurements of Pinus massoniana and plants with similar structures can be made through manual estimation. In a small number of papers, the plant height, ground diameter, and crown width parameters are simply estimated using depth cameras. However, the XOY plane is used in these papers as the projection plane to calculate the stem diameter. Because the near soil part of the stem of Pinus massoniana is not parallel to the soil plane, selecting the XOY plane as the projection plane will cause a position shift of the stem on the plane and ultimately cause the low accuracy of the ground diameter, with a value of approximately 75%. Moreover, no research on the other phenotypic parameters of the leaves has been conducted in these studies. Therefore, in our research, we aim to find a more accurate method to predict these parameters and improve the prediction performance. Moreover, effectively estimating the leaf length and leaf number is proposed in our study.

In this study, based on the idea of the skeletonization of slice clustering, slices in the horizontal direction were added, and slices with skeleton points of the main stem were selected with grayscale values. Combining the idea of the projection method, the regional leaf length and leaf number of Pinus massoniana seedlings were estimated. Finally, an automated method for obtaining plant phenotypic parameters from the 3D models of plants is proposed. The overall objective of this study is to automatically extract phenotypic parameters of Pinus massoniana seedlings through 3D point cloud analysis. The specific objectives are (1) to reconstruct a 3D point cloud model of plants; (2) to develop a set of processing flows to analyze the structural characteristics of plants; and (3) to automatically extract phenotypic parameters, including plant height, stem diameter, main stem length, leaf number, and regional leaf length.

Materials

A set of non-destructive 3D image collection devices, including an Azure Kinect camera, precision rotary table, bracket, black curtain, and computer, was obtained and assembled by our research group. The configuration of the computer was Intel CPU E5-2670/16G/1TB/Quadro K2000 4G DDR4, the operating system was Windows 10 Professional, and the required software was Microsoft Visual Studio 2017, OpenCV4.5.3, and PCL1.8.1. The Aruze Kinect camera integrated a color camera of 4096×3072 pixels, a depth camera of 1024×1024 pixels, and an infrared camera of 1024×1024 pixels, and it was based on the principle of ToF (L. J. T. w. p. Li, 2014). The point cloud collection device is shown in Figure 1. The camera position was fixed, and the Pinus massoniana seedlings were placed on the turntable, while the camera height was set to 0.5 meters, and the camera angle at the bracket was adjusted to 20 degrees down to the horizontal plane. The distance between the camera and the plant was approximately 0.4-0.5 meters. The whole plant was in the center of the camera vision view, and appropriate adjustments were made artificially according to different plants. The distance between the turntable and the curtain was 1.2-1.5 meters, and the back shadow of the plant was within the scope of the black curtain to reduce other background interference. The point cloud image was acquired every 180 degrees, and two point clouds were obtained for one plant by registration. The whole acquisition was completed by the self-developed program based on the software development kit (SDK) of the camera.

Methods

The pipeline for phenotypic parameter acquisition involved three main steps, namely, (1) point cloud preprocessing, (2) stem and leaf segmentation, and (3) morphological trait extraction. Among them, the point cloud preprocessing step used pass-through filtering, RANSAC fitting plane and removal, radius filtering method to remove the background, experimental platform, and flowerpots. Then, the improved iterative closest

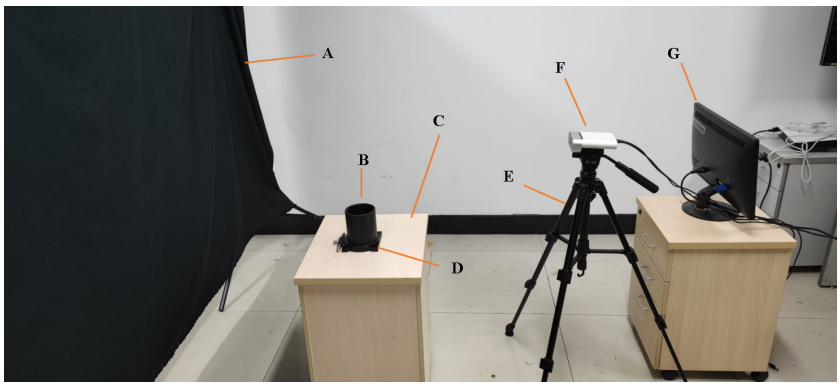


FIGURE 1
Nondestructive 3D image acquisition setup. **(A)** Curtain. **(B)** Plant placement. **(C)** Experimental platform. **(D)** Rotary table. **(E)** Yunteng691 bracket. **(F)** Aruze Kinect camera. **(G)** Computer.

point (ICP) algorithm based on feature point consistency was proposed to register the point cloud images of two angles, and complete plant point cloud images were obtained. In the stem and leaf segmentation step, the slices in the vertical direction and horizontal direction were sliced, and the skeleton points were extracted by clustering each slice. Then, the alternative skeleton points of the main stem were extracted by the directed acyclic graph (DAG) longest path algorithm, and the canopy length was estimated by local plane projection and convex hull fitting. The skeleton points of the canopy contained in the alternative skeleton points of the main stem were removed according to the canopy length, and the remaining skeleton points were regarded as points at the main stem. Finally, the skeleton points of the main stem were interpolated, and the point cloud of the main stem was obtained by k-nearest search to perform stem and leaf separation. In the morphological evaluation step, five phenotypic parameters (plant height, stem diameter, main stem length, regional leaf length, and leaf number) were estimated from the point cloud after stem and leaf separation. Our research used the Point Cloud Library (Rusu and Cousins, 2011) and OpenCV Library (Bradski and Kaehler, 2008) to implement these steps in our program.

Point cloud preprocessing

The pipeline presented involved six steps, namely, (1) 3D point cloud generation, (2) background removal, (3) experimental platform removal, (4) discrete point removal, (5) soil and flowerpot removal, and (6) registration. In the first process, the depth image obtained by the image acquisition device was converted into 3D point cloud data. In the second step, the background curtain was removed by pass-through filtering. The third step extracted and removed the experimental platform by the RANSAC fitting plane. The fourth stage used the radius filter method to remove the discrete points in the image. The fifth stage removed the soil and the parts below the soil. Finally, the point clouds processed in the first five stages were registered by the improved ICP algorithm.

3D point cloud generation

The image acquisition device obtained the depth image of the plant. For further analysis and study, the depth image was transformed into 3D point cloud data according to Equations (1) to (3) through coordinate axis transformation.

$$Y = d \tag{1}$$

$$X = Z(x - c_x)/fx \tag{2}$$

$$Z = Z(y - c_y)/f_y \tag{3}$$

where x and y represent the 2D coordinates provided by the depth image according to the camera coordinate system, d represents the depth value information directly provided by the depth image, c_x and c_y represent the coordinates on the X-axis and Y-axis of the lens aperture center, respectively, and f_x and f_y are the focal lengths on the X-axis and Y-axis of the camera, respectively. The three-dimensional coordinates (X, Y, Z) of the corresponding cloud points could be obtained by combining the above formula, the X-axis represents the horizontal information of the point cloud, the Z-axis represents the height information of the point cloud, the Y-axis represents the depth information of the point cloud. The real image of the plant is shown in Figure 2A, and the depth image of the plant is shown in Figure 2B. The point cloud image of the plant is shown in Figure 2C.

Background removal

The main object was the plant, but the point cloud data generated contained some background objects, such as background curtains, loading platforms, and flowerpots. These background objects were not the targets to be measured and interfered with the point cloud analysis of subsequent plants. The background was removed by pass-through filtering. The upper and lower boundaries of the pass-through filtering were obtained using Equations (4) and (5):

$$PT_{lb} = maxPt.z \tag{4}$$

$$PT_{ub} = (d_{cc} - d_{cp})/d_{cc} * maxPt.z \tag{5}$$

where PT_{lb} and PT_{ub} represent the near-far boundary in the depth direction of the direct filtering, as well as the distance between the lens to the curtain and the distance between the loading platform to the curtain, respectively. d_{cc} represents the distance between the camera and the experimental curtain, d_{cp} represents the distance between the camera and the plant, and $maxPt.z$ represents the maximum value of the depth direction of all points in the point cloud. Hence, PT_{lb} and PT_{ub} were two boundaries, and the point clouds outside of the boundaries were removed. Data obtained from the above two formulas were used as the input parameters to perform direct filtering to remove the background. The background removed point cloud image is shown in Figure 2D.

Experimental platform removal

The loading platform, which is considered an irrelevant variable, had no effect on the measurement process and interfered with the point cloud analysis of subsequent plants. The point cloud of the experimental platform was detected by RANSAC plane fitting (Fischler and Bolles, 1981). The experimental platform removed point cloud was obtained by subtracting the background removed point cloud from the experimental platform point cloud. The experimental platform removed point cloud image is shown in Figure 2E.

Discrete point removal

Due to the limitations of the camera accuracy and the RANSAC plane fitting algorithm, the point cloud without the background and loading platform objects contained some sparse noise points, which had low point density. A radius-based outlier filter was used to remove these discrete noise points. The setting parameters of the radius filter were a custom search radius r and the number of adjacent points k in the search radius r. If the number of points in



FIGURE 2
Image of point cloud preprocessing. **(A)** Real image. **(B)** Depth image. **(C)** 3D point cloud image. **(D)** Background removed point cloud image. **(E)** Experimental platform removed point cloud image; parts of discrete points are marked by red circles. **(F)** Discrete points removed point cloud image. **(G)** Soil detect image; point clouds in the soil are marked in red. **(H)** Registered point cloud image of 0 degrees. **(I)** Registered point cloud image of 180 degrees.

the search radius was less than the specified value k, the point was removed as a noise point (Dziubich et al., 2016). In this study, $r = 0.01m$ and $K = 10$ were tested to effectively reduce the number of noise points without damaging the original point cloud image structure. The point cloud after removing discrete points by radius filtering is shown in Figure 2F.

Soil and flowerpot removal

In the next registration step, the method of feature point consistency was used; however, the flowerpots interfered with the registration step, and it was necessary to remove the soil of the plant and the flowerpots below the soil. The height of the flowerpots was fixed at 0.12 m in our experiment, and 0.07 m to 0.12 m above the experimental platform was selected as the soil area for pass-through filtering. The points above the soil, which are regarded as the point cloud area after the soil and flowerpot were removed, were the inputs of the registration step, as shown in Figure 2G.

Registration

This step registered point clouds collected from two angles after the preprocessing step. The registration step processed coarse registration first and then fine registration. Two preprocessed point clouds, the source point cloud, and the target point cloud, were the input of the registration step. In the coarse registration process, the KD-tree search method was used to find k adjacent points at a point in the point cloud, and the k points were fitted to the minimum quadratic plane, where the k value was set to 10. Principal component analysis (PCA) (Feng et al., 2014) was applied to the cloud points of the quadratic plane, and the normal vector corresponding to the nonzero minimum eigenvalue was considered the local normal vector. The above steps were looped until all points of the source and target point clouds were traversed. After the normal vectors of all points were obtained, the angle of normal vectors between the self-point and its adjacent points was calculated, and the mean value of these angles was obtained at the same time. The appropriate angle mean threshold was set, and if the angle mean was greater than the threshold, the point was used as the feature point in the source point cloud and the target point cloud. The fast point feature histogram (FPFH) descriptors of the feature point cloud were calculated, and the coarse registration of the target point cloud was carried out using the sampling consistency method (SAC-IA) (Rusu et al., 2009).

Improved ICP registration based on the KD-tree adjacent search algorithm was used for fine registration of the target point

cloud after coarse registration (Zhang, 1994). Finally, the target point cloud and source point cloud after registration and rotation were combined into a complete point cloud image. The registered point clouds of 0 degrees and 180 degrees are shown in Figures 2H, I, respectively.

Stem and leaf segmentation

The pipeline presented involved four steps, namely, (1) skeletonization, (2) main stem alternative skeleton point extraction, (3) main stem skeleton point extraction, and (4) main stem cloud point restoration. In the first stage, the preprocessed point cloud was sliced, clustering was performed on each slice, and the slice of the stem was selected by grayscale values. In the second stage, the minimum spanning tree (MST) was established according to the skeleton points, and the alternative skeleton points of the main stem were extracted by the DAG longest path algorithm. In the third stage, the length of the plant canopy was estimated by the local projection method. According to the length of the canopy, the canopy skeleton points contained in the alternative skeleton points of the main stem were removed, and the skeleton points of the main stem were obtained. Finally, the skeleton points of the main stem were interpolated, and the point cloud of the main stem was obtained by k-nearest search searching. The remaining point cloud was regarded as the area in which the leaf part was located to perform stem and leaf segmentation.

Skeletonization

An improved skeletonization method based on the phenotypic characteristics of Pinus massoniana seedlings was proposed. First, the slice along the Z-axis of the point cloud was performed, and a total of 30 slice layers were obtained. Then, Euclidean clustering was applied to the point clouds of each slice layer, and the corresponding clusters in each slice layer were obtained (Rusu and Cousins, 2011). However, due to the structural characteristics of Pinus massoniana, the clusters of Z-axis slices involved leaves and main stem parts. It was difficult to distinguish whether the cluster was from the leaf part or the main stem part, while the centroid was extracted from the different clusters of the slice layer. Therefore, further processing was needed at the slice layer along the X-axis.

The maximum differential value of the X-axis coordinates of the point cloud along the Z-axis of each slice layer in the cluster was calculated, and the corresponding cluster along the Z-axis slice layer, whose differential value was greater than the threshold, was sliced twice along the X-axis direction (the number of slices along the X-axis was 5). The slicing point cloud image along the Z-axis and its horizontal direction subdivision for each Z-axis slice layer is shown in Figure 3A. The average grayscale values of all slice layers in the same cluster along the X-axis were calculated. The slice layer with the largest average grayscale value of five horizontal slice layers was taken as the main stem location. The centroid of this X-slice layer was calculated and used as the skeleton alternative point of the corresponding cluster for the main stem of Pinus massoniana on this Z-axis slice. If the maximum differential value of the X-axis coordinates was less than the threshold, the centroid of the cluster was directly calculated and used as the skeleton alternative point of the cluster corresponding to the main stem of Pinus massoniana on this Z-axis slice. According to the above steps, traversing all slice layers along the Z-axis, the skeleton alternative points of the main stem were collected. The algorithm for skeletonization is shown in Figure 4.

Nomenclature in Algorithm1	
Z_{max} Z_{min}	Maximum and minimum of the Z-axis values in the preprocessed point cloud
X_{max} X_{min}	Maximum and minimum of the X-axis values in the point cloud
P_c	Preprocessed point cloud
S_w	Slice layer width

Main stem alternative skeleton point extraction

The alternative skeleton point clouds obtained after the skeletonization step were discrete, and it was necessary to establish adjacent relationships between the skeleton points. The MST was used to establish this relationship for alternative skeleton point clouds, and the Kruskal algorithm (Kruskal, 1956) was used to build the MST of all alternative skeleton point sets of the main stem. By searching up from the root node of the MST, the directed acyclic graph between the points was generated according to the adjacent relation formed by the MST. The diagram of the relationship between the skeleton points is shown in Figure 3B. Then, the DAG longest path algorithm was used to search the longest path from the root node in a directed acyclic graph (Healy and Nikolov, 2001). The skeleton points in the longest path were set as the skeleton points of the main stem with canopy, which is shown in Figure 3C.

Main stem skeleton point extraction

As shown in Figure 3C, the top of Pinus massoniana is a canopy, and it did not belong to the main stem part. As a result, it was necessary to remove the canopy part involved in the skeleton points of the main stem with the canopy. Therefore, according to the distance information between the skeleton points provided by the MST, the canopy length was estimated, and the skeleton points in the canopy range were removed from the top based on the estimated length. Thus, the skeleton points of the main stem and the canopy were separated.

In this study, the number of point clouds of the main stem skeleton with a canopy was investigated. It was found that the boundary point between the canopy and the main stem was near the fourth or fifth point of the skeleton points with the canopy from top to bottom in our samples. Therefore, the fifth point was selected as the center of the following local projection method. The highest point of the point cloud in the main stem skeleton without the canopy was the starting point, and the MST was traversed down to the position of the fifth skeleton

10.3389/fpls.2023.1146490

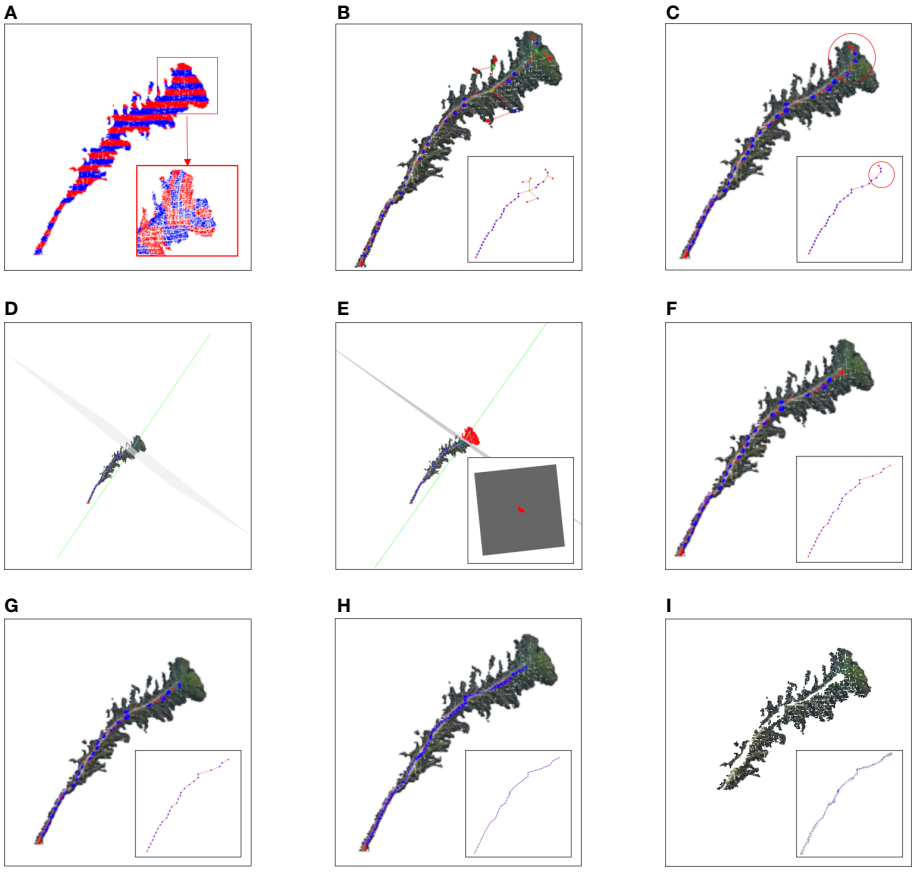


FIGURE 3
Image of the stem and leaf segmentation. **(A)** Point cloud slices along the Z-axis and X-axis with adjacent slice layers distinguished by different colors. **(B)** Diagram of the relationship between the skeleton points. The green point, red point, and blue point represent the junction, vertex, and internal node, respectively. **(C)** Skeleton points of the main stem with the canopy; the red circle represents the canopy. **(D)** The green line represents a normal composed of a fifth skeleton point and the adjacent skeleton points above it. The gray normal represents a projection plane perpendicular to the normal. **(E)** The red part represents the projection of the top canopy on the tangent plane. **(F)** Skeleton point of the main stem after removing the canopy. **(G)** Main stem skeleton points before interpolation. **(H)** Main stem skeleton points after interpolation. **(I)** Point cloud after removing the main stem.

Algorithm 1 Skeletonization.

Input: *The point cloud after preprocessing method.*
Output: *The main stem alternative skeleton point set.*

$S_w = (Z_{max} - Z_{min}) / 30$
for each $i \in [0, 30]$ **do**
  Generate a temporary cloud T_c to store points in each slice layer of Z-axis
  for each *point* **in** P_c **do**
    if point.z >= (Z_{min} + S_w * i) and point.z <= (Z_{min} + S_w * (i+1)) **then**
      add this point to T_c
    end if
  end for
  cluster set = **euclideanClustering**(T_c)
  for each *cluster* **in** *cluster set* **do**
    calculate X_{max} and X_{min} of cluster
    if (X_{max} - X_{min}) > threshold **then**
      make 5 slices in x-axis of cluster
      calculate average grayscale value A_g of each slice
      calculate centroid point $\mathbf{C}_p$ of slices with the maximum A_g
      add $\mathbf{C}_p$ to main stem alternative skeleton point set
    end if
    else
      calculate centroid point $\mathbf{C}_p$ of cluster
      add $\mathbf{C}_p$ to main stem alternative skeleton point set
    end else
  end for
end for

FIGURE 4
Algorithm for skeletonization.

point. Then, the plane was formed under the centroid of the fifth skeleton point and the skeleton point above it, which was perpendicular to the normal. The normal and projection planes are shown in Figure 3D. The point cloud above the plane was projected onto the plane. Points projected onto the plane are shown in Figure 3E. The convex boundary of the 2D projection point cloud was extracted based on the convex hull algorithm (Cupec et al., 2020). The average distance from all points on the boundary to the search center (fifth point) was determined to be the threshold of the specified canopy length. From the top of the skeleton point set of the main stem with the canopy, the sum of the paths between the skeleton points according to the distance relationship provided by the MST was calculated. During searching, if the sum of the paths was greater than the threshold, the search was stopped. The skeleton point from this point to all the collections from the top was discarded, and the remaining skeleton points were the skeleton points of the main stem. The skeleton point of the main stem is shown in Figure 3F. The algorithm for main stem skeleton point extraction is shown in Figure 5.

Nomenclature in Algorithm2	
p_i	The i-th point of the alternative main stem skeleton point from top to bottom
Line()	The line between two points
P_c	Preprocessed point cloud
M_{as}	Main stem alternative skeleton point set
$M_{as}.len$	The length of the main stem alternative skeleton point set
p_f	Dividing point of the canopy and main stem part

Main stem point cloud restoration

The purpose of this step was to restore the main stem according to the skeleton point cloud of the main stem. The k-nearest search method was used to search the point cloud along the skeleton points of the main stem within the set radius range, and if the search radius was set too large, the point cloud of the leaf part might be wrongly restored to the main stem. The distance between the main stem skeleton points was uniform, and the large distance caused leaf point cloud loss in the search. Therefore, in this study, the main stem skeleton points were interpolated to solve the above problem.

Because the near-soil parts of the main stem of Pinus massoniana had no leaves, the point cloud of the main stem extracted by the k-nearest search method did not involve the leaf point cloud, and the diameter of the main stem in this region was calculated using the algorithm mentioned in the morphological evaluation step and used as an interpolation step length. The number of new skeleton points inserted between the two skeleton points was determined from the distance between the two skeleton points divided by the interpolation step length, then linear interpolation was carried out. All the adjacent skeleton points were traversed to complete the interpolation expansion of the whole skeleton point set. The main stem skeleton points before and after interpolation are shown in Figures 3G, H. For the expanded main stem skeleton points after interpolation, a k-nearest search was used to search the points within the radius (the search radius was the main stem diameter), and all the searched points could be regarded as the main stem part. When the preprocessed point cloud was subtracted from the point cloud of the main stem, the point cloud of leaves was obtained, and the stem and leaf segmentation step was realized. Point clouds that did not contain the main stem are shown in Figure 3I. The algorithm for main stem point cloud restoration is shown in Figure 6.

Algorithm 2 Main stem skeleton point extraction.

Input: *The main stem alternative skeleton point set.*
Output: *The main stem skeleton point set.*

Generate a plane to take p_5 as the center and **Line**(p_4, p_5) as the normal
for each *point* **in** P_c **do**
  substitute the coordinates of point into the plane equation
  if substitute results > 0 **then**
    project this point onto the plane
  end if
end for
construct convex boundary with peojected points
cacluate average distance A_d from all points on the boundary to p_5
for each $i \in [0, M_{as}.len]$ **do**
  dis = **distance**(p_i, $p_{(i+1)}$)
  $A_d = A_d$ - dis
  if $A_d < 0$ **then**
    $p_f = i$
    break
  end if
end for
for each $i \in [p_f, M_{as}.len]$ **do**
  add p_i to main stem skeleton point set
end for

FIGURE 5
Algorithm for main stem skeleton point extraction.

Algorithm 3 Main stem point cloud restoration.

Input: *The main stem skeleton point set.*
Output: *Restored main stem point cloud.*

```
for each i ∈ [0, M_s.len] do
    add p_i to M_e
    for each j ∈ [0, 10] do
        Create a new point
        point.x = p_i. x + (p_(i+1). x - p_i. x) / 10 * j
        point.y = p_i. y + (p_(i+1). y - p_i. y) / 10 * j
        point.z = p_i. z + (p_(i+1). z - p_i. z) / 10 * j
        add point to M_e
    end for
end for
for each point in M_e do
    use k-nearest search method to search all neighborhood points of point in P_c
    add all the searched points to restored main stem point cloud
end for
```

FIGURE 6
Algorithm for main stem point cloud restoration.

Nomenclature in Algorithm3	
p_i	The i-th point of the main stem skeleton point from top to bottom
P_c	Preprocessed point cloud
M_s	Main stem skeleton point set
$M_s.len$	The length of the main stem skeleton point set
M_e	Expanded main stem skeleton point set

Morphological trait extraction

Five phenotypic parameters, including the plant height, stem diameter, main stem length, regional leaf length, and leaf number, were calculated automatically using the point cloud in the output of the stem and leaf segmentation step. The stem diameter was calculated using the local projection method. Part of the main stem was projected onto the tangent plane formed along the skeleton point above the soil. The stem diameter was represented by the short axis length of the ellipse by 2D ellipse fitting. The main stem length was obtained by calculating the distance between adjacent skeleton points. For the regional leaf length, we took a region along the skeleton point of the main stem every 5 cm, and the point cloud in the region was projected to the local tangent plane. The convex hull method was used to estimate the length of the projection boundary to the center, and the estimated length was used as the local leaf length. The number of leaves can also be determined by searching along the main stem skeleton point, and every 5 cm was regarded as a region. The distance density of this region was calculated, and the number of leaves in this region was estimated by the linear relationship between the distance density and leaves. Then, the total number of leaves was estimated from the density.

Plant height

The plant height was defined as the distance between the soil plane and the top of the canopy, which is a useful and frequently measured trait in agronomic research (Moles et al., 2009). Using the highest point of the whole plant as the top of the plant, the plant height was calculated using **Eq. 6**:

$$H = Z_{max} - Z_{soil} \tag{6}$$

where Z_{max} and Z_{soil} represent the Z-axis of the highest point of the point cloud after preprocessing and the Z-axis of the soil plane, respectively.

Stem diameter

The measurement position of the stem diameter was close to the soil, and the measurement value of this position was for the whole stem. The routine method selected the part 2-5 cm above the soil and sliced along the Z-axis with a thickness of 1 cm. The point cloud of each selected slice layer was projected onto the XOY plane, a 2D ellipse was fitted using these 2D projected points, and the minor axis of the ellipse was taken as the estimated value of the stem diameter. However, the slice layer of the plant along the Z-axis was not perpendicular to the XOY plane, leading to a certain deviation of the points on the XOY plane. Therefore, we proposed an improved algorithm for skeletonization and projection and for calculating the stem diameter.

The expanded skeleton points of the main stem were used, and all skeleton points in the 2-4 cm region above the lowest skeleton point were selected. At the initial stage, the normal was formed by the skeleton point at the lowest position. Then, the plane that was perpendicular to the normal was generated while the lowest skeleton point was the center, and the generated plane was denoted as the lower plane. The skeleton point above the lowest skeleton point was taken as another center, and the normal was generated with the skeleton point above another center. Another plane perpendicular to the normal was generated and was denoted as the upper plane. The upper and lower planes are shown in Figures 7A, B, respectively. The points between the upper plane and the lower plane were projected to the lower plane after the preprocessing step. Every 1 cm was sliced as a layer. Then, the 2D ellipse fitting operation was performed on these projection points, the short axis length of the fitted ellipse was used as the diameter,

10.3389/fpls.2023.1146490

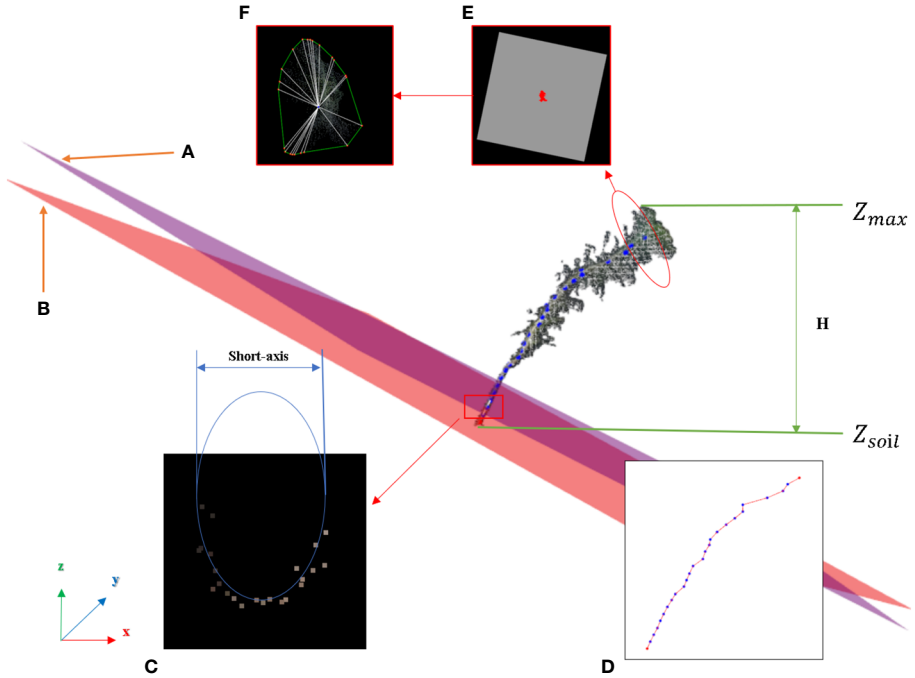


FIGURE 7
Image of the morphological traits extraction. **(A)** Upper plane. **(B)** Lower plane. **(C)** 2D projection of the main stem point cloud, short-axis of the ellipse as the stem diameter. **(D)** Main stem skeleton points for calculating the main stem length. **(E)** 2D points projected onto the lower plane. **(F)** Convex boundaries and links between the boundaries and center point.

and the average value of the diameters was the stem diameter of the main stem. The projection of points is shown in Figure 7C.

Main stem length

The stem is an important organ that connects the transport path; Pinus massoniana seedlings have only one main stem, and leaves grow around the main stem (Zhuo et al., 2010). The main stem skeleton point cloud is shown in Figure 7D. By traversing from the lowest point of the skeleton to the highest point, the distance between the adjacent points in the MST was determined, and the sum of all the distances was the main stem length.

Regional leaf length

Since the growth conditions of different parts of Pinus massoniana were different, the average leaf length of different parts was estimated in this study. The region of the skeleton point cloud of the main stem was taken every 5 cm along the main stem direction. The lowest position of the skeleton points in the region and its adjacent skeleton points were selected and used to form one normal vector. Then, the plane that was perpendicular to the normal vector was generated where the lowest position skeleton point was the center and was denoted as the lower plane. The highest skeleton point of the region was the center, with its adjacent skeleton points below the form of a normal vector, generating a plane perpendicular to the normal that is denoted as the upper plane. The points between the upper plane and the lower plane were vertically projected to the lower plane, as shown in Figure 7E. The convex boundary of the 2D projection point cloud was extracted by the convex hull algorithm (Cupec et al., 2020), which is shown in Figure 7F.

The average distance from all points on the boundary to the center of the lower plane was determined, and the calculated stem diameter was the threshold. If the average distance was greater than the threshold, the average distance was used for the average leaf length of the region. If it was less than or equal to the threshold, it was considered that there were no leaves in this region. The above algorithm could be used to estimate the average leaf length of the parts below the canopy of Pinus massoniana along every 5 cm interval of the main stem.

Leaf number

The Pinus massoniana leaves were dense and large in number. It was difficult to obtain separate leaf images even using a high-precision scanner. The point cloud distance density is an important analysis index and can be used to analyze plant phenotypic characteristics. For example, by using the different densities of the point clouds of stems and leaves in the horizontal direction, the support vector machine can be applied to classify the stems and leaves by density (Liu et al., 2020). In our study, the same regional segmentation method as the regional leaf length step was used to calculate the average distance density of the point cloud in different segmentation regions, and the number of leaves in corresponding regions was counted. It was found that the distance density of the point cloud had a certain linear relationship with the number of regional leaves. Based on this relationship, we proposed a plant leaf estimation algorithm based on the distance density method (C.-H. Lin et al., 2018). Moreover, the average distance density of the point

cloud can be expressed by **Eq. 7** and **Eq. 8**:

$$d_p = min(dis(p,q)), q = 1 \cdots N, p \neq q \quad (7)$$

$$\bar{d} = \frac{1}{N}\sum_{p=1}^{N} d_p \quad (8)$$

$dis(p,q)$ represents the Euclidean distance between Point p and any other point q in a point cloud with N point numbers, and d_p is the minimal value of all the distances. The smaller $\bar{d}$ is, the sparser the point cloud distribution; the larger $\bar{d}$ is, the denser the point cloud distribution. Based on the above formula, the algorithm for calculating the number of each leaf was as follows. The skeleton point cloud region of the main stem was used for estimating every 5 cm along the main stem, and the lowest position of the skeleton points in the region was selected as the center. With one upper adjacent skeleton point, a normal vector can be formed, and the plane that was perpendicular to the normal vector was generated and denoted as the lower plane. If there were leaves in this region (the determination method was the same as the determination method in the local leaf length step), the average distance density between these points can be calculated. Through the linear relationship between the average distance density and the number of regional leaves, the number of leaves in this region can be obtained. The leaf number of each region was calculated, and the total leaf number of Pinus massoniana was obtained by the summation of all the regions.

Results

Figure 8 shows the stem and leaf segmentation results of three representative Pinus massoniana plants with height ranges of 25-30 cm, 20-25 cm, and 15-20 cm. The three-dimensional surface visualization found that the 3D model was close to the real sample, and the skeleton of the main stem from the point cloud model is mostly consistent with the skeleton of the actual plant. We can note that the skeleton points of the third plant in Figure 8 have a small deviation from the real stem because this deviation has many leaves and shelters the main stem. Thus, the skeleton points of this part that were found by the grayscale value clustering deviate; however, no more than 5% of the skeleton points had this deviation, and the actual impact was not significant. Overall, the results show that the stem and leaf segmentation algorithm could effectively separate the main stem and leaf from the 3D point cloud of Pinus massoniana plants.

Accuracy assessment

In this study, 100 Pinus massoniana seedlings at 15–30 cm were collected as part of the experiment. The accuracy of the algorithm was evaluated by the correlation coefficient (R), root mean square error (RMSE), and mean absolute error (MAE). RMSE and MAE were defined as **Eq. 9** and **Eq. 10**.

$$RMSE = \sqrt{mean(est - act)^2} \quad (9)$$

$$MAE = mean(|est - act|) \quad (10)$$

where *est* and *act* denote the estimated values and actual values from manual measurements, respectively.

A comparison of the estimates and manual measurements of the phenotypic parameters is shown in Figure 9. The results showed that this set of algorithms was suitable for the phenotypic parameter extraction of Pinus massoniana seedlings, especially for plant heights and main stem lengths. Among them, the R between the estimated plant height and the artificially measured value was 0.96, and the RMSE and MAE were 1.35 cm and 1.21 cm, respectively; the R between the estimated plant height and the artificially measured value was 0.93, and the RMSE and MAE were 2.27 mm and 2.07 mm, respectively; the R between the estimated main stem length and the artificially measured value was 0.95, and the RMSE and MAE were 2.21 cm and 2 cm, respectively; the R between the estimated leaf numbers and the artificially measured value was 0.75, and the RMSE and MAE were 52 cm and 48 cm, respectively; the R between the average leaf length 0-5 cm below the canopy and the measured values was 0.74, and the RMSE and MAE were 1.13 cm and 0.96 cm, respectively; the R between the average leaf length 5-10 cm below the canopy and the measured values was 0.78, and the RMSE and MAE were 1.16 cm and 1.04 cm, respectively; the R between the average leaf length 10-15 cm below the canopy and the measured values was 0.73, and the RMSE and MAE were 0.89 cm and 0.8 cm, respectively; the R between the average leaf length 15-20 cm below the canopy and the measured values was 0.79, and the RMSE and MAE were 0.63 cm and 0.58 cm, respectively.

Regarding the plant height, the regression line of the data points was in good agreement with the reference line of the diagonal (R = 1), as shown in Figure 9A, which verified that the algorithm was suitable for measuring the plant height of Pinus massoniana. The average height of all artificially measured plants was 21.75 cm, and the average absolute error was 5.5% of the average. The reason for this error was that the soil plane was slightly uneven, resulting in the overestimation or underestimation of the lowest soil plane. However, this error can be ignored.

Regarding the stem diameter, Figure 9B shows that the estimated value of the system is generally greater than the measured value, which was caused by the ToF principle of the Aruze Kinect camera. Moreover, the reflection of the edge leads to the expansion of the point cloud around the stem.

Regarding the stem diameter calculation step, we have shown that the average absolute error of the stem diameter calculated by selecting the tangent plane of the local skeleton point of the main stem as the projection plane was 26.6% of the manual measurement average, while the average absolute error of stem diameter calculated using the XOY plane as the projection plane was 42.2% of the manual measurement average. Therefore, using the projection plane used in this study to calculate stem diameter greatly improves the accuracy.

Regarding the length of the main stem, Figure 9C shows that the correlation between the estimates and manual measurements is close to 1. The average length of the main stem of all plants

10.3389/fpls.2023.1146490

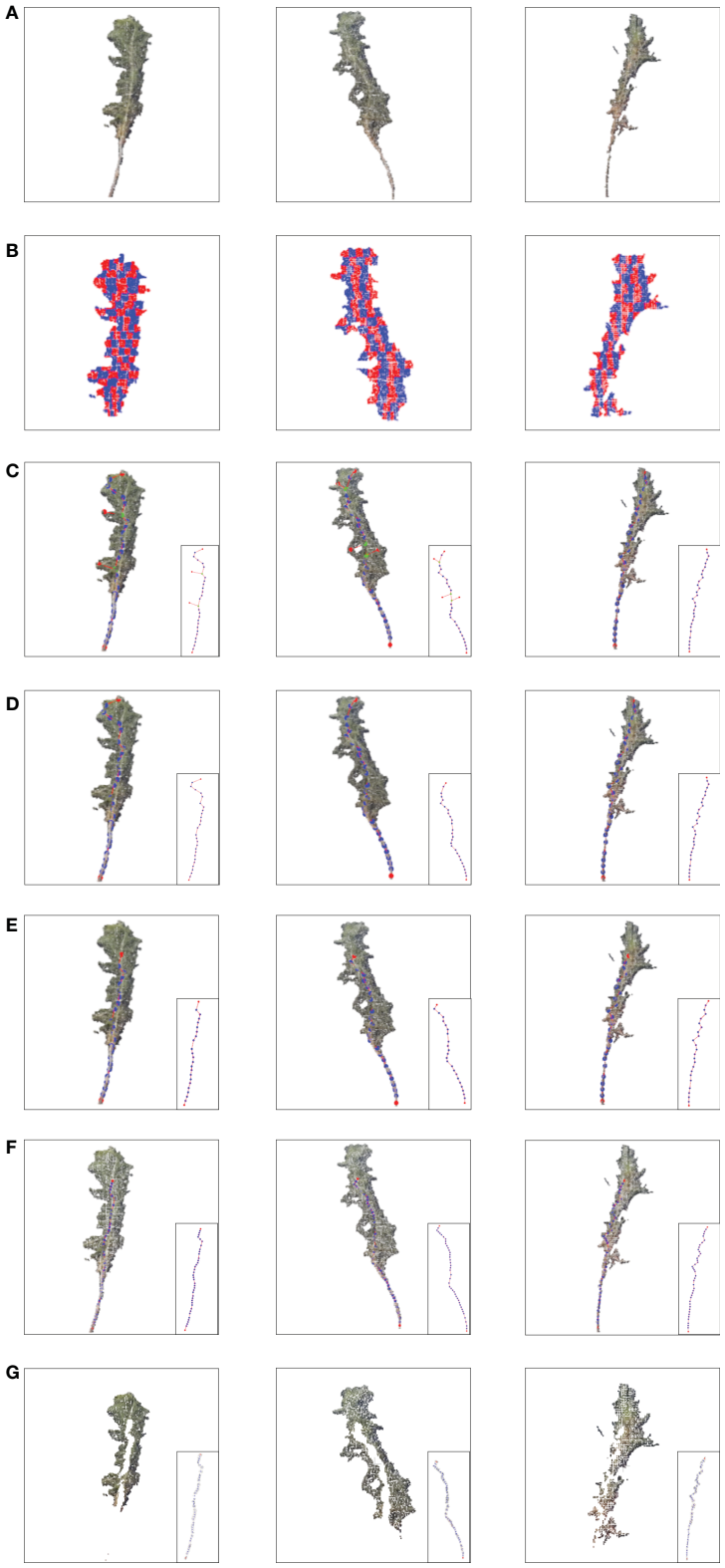


FIGURE 8
Stem and leaf segmentation procedure visualization of different heights. **(A)** Input cloud point. **(B)** Slice in the X-axis and Z-axis. **(C)** Extracting alternative skeleton points of the main stem according to the centroid and generating MST according to the skeleton points. **(D)** DAG longest path algorithm searches for the skeleton points of the main stem with a canopy. **(E)** Main stem skeleton points after removing the canopy. **(F)** Main stem skeleton points after interpolation expansion. **(G)** After removing the main stem point cloud, the plant point cloud only contains leaves.

10.3389/fpls.2023.1146490

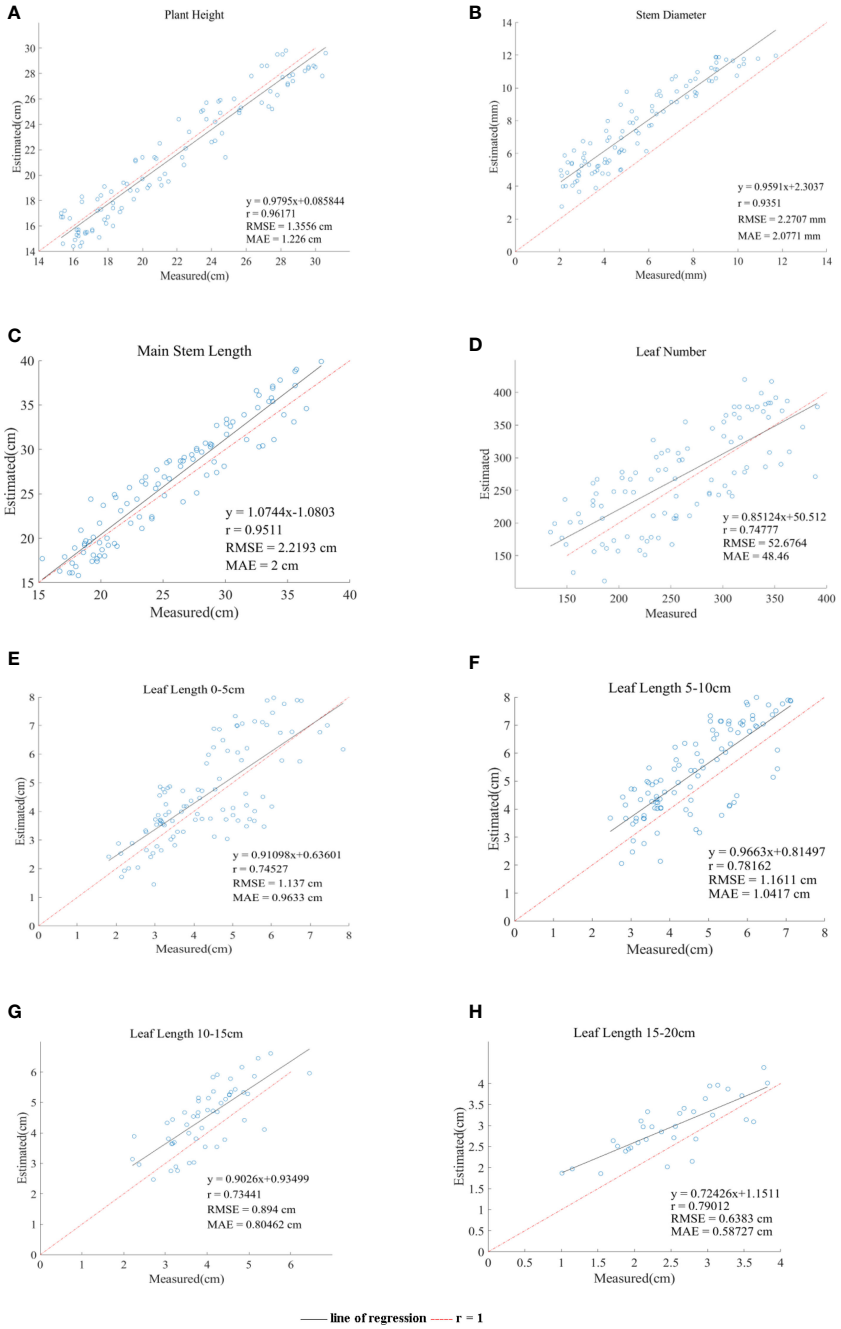


FIGURE 9
Comparison of the six phenotypic parameters system estimates and manual measurements. (**A–H**) line of regression represents the straight line fitted by the estimated value. r = 1 represents estimated value is equal to actual value and regards as a reference line.

measured manually was 25.6 cm, and the average absolute error was 7.8% of the average. It was proven that this algorithm is applicable to the length of the main stem. The random error may be caused by the following two reasons. First, the main stem might be partially blocked by the leaves, resulting in a deviation between the slice layer of the main stem selected by the grayscale value and the slice layer of the actual main stem. Second, the grayscale of withered Pinus massoniana leaves was similar to that of the main stem, thus interfering with the system. However, the object of this study was to analyze the seedling stage, with few or no withered leaves, and this impact factor can be ignored.

Regarding the leaf number, we proposed the density method to estimate the leaf number of Pinus massoniana. The regional leaf number and density of Pinus massoniana seedlings were calculated, and the linear relationship between the leaf number and density was obtained using the least square method. The relationship between the estimates and manual measurements is shown in Figure 9D. Because of the least square method (linear relationship), the number of leaves deviated from the actual value for the nonlinear relationship. The average number of leaves measured by hand was 259, and the average absolute error was 18.5% of the average measured by hand. It was proven that this estimation method had

certain feasibility. This method could avoid the high cost caused by using a high-precision scanner and reduce the time consumed by manual measurement.

Regarding the regional leaf length, Figure 9E–H shows the average leaf length of 0-5 cm, 5-10 cm, 10-15 cm, and 15-20 cm below the canopy. The average absolute errors were 22.2%, 22.3%, 20.1%, and 23% of the average manually measured values, respectively. The main reason for these errors was that there are few long leaves, and these leaves would be included by the convex hull method, causing corresponding errors. In general, the estimated leaf length of the system was close to the actual leaf length, which also proved that the algorithm in this study was feasible.

We also used a box plot for analysis. However, in the measurement of each phenotypic parameter, there were mostly no discrete points with values that were too large or too small. The line chart could show all the information of the box plot and showed some information that the box plot did not have. Therefore, in this paper, we used a line chart for analysis.

Discussion

In this study, a set of automated methods that measure the phenotypic parameters of Pinus Massoniana seedlings was developed to process images collected by a 3D image acquisition device through non-destructive means. The five parameters of interest were plant height, stem diameter, main stem length, regional leaf length, and leaf number, and they were automatically obtained from the original image. Because of the complex phenotypic characteristics of Pinus massoniana, there have been few studies on such plants. Our study focuses on the stem-leaf separation steps and phenotypic parameter extraction.

As an important step, stem and leaf segmentation could provide great convenience for the extraction of phenotypic parameters of subsequent plants. We also used three kinds of skeletonization methods to analyze Pinus massoniana seedlings based on skeleton contraction, local feature, and slicing. One of the typical skeleton contractions was Laplace skeletonization, and the core idea of Laplace skeletonization was to search the plant point cloud in the stem direction through matrix transformation, which had a good effect on gramineous plants, but it is not applicable for curved main stem. The second kind of skeletonization method was based on feature; however, the leaf part and stem part of Pinus massoniana were overlapping, and it was difficult to find the proper feature. The third type was based on slicing. However, the main stem of Pinus massoniana was curved, and some of the skeleton points of the main stem were out of the range of the extracted Hough plane, resulting in some main stem skeleton points being missed (comparative tests are included in the Supplementary Materials).

However, the idea of slicing provided much inspiration for our skeletonization, for Pinus massoniana, the Z-axis slice layer contained both the stem part point cloud and the leaf part point cloud. Then, slice subdivision was performed along the X-axis. We compared the features of each slice layer and found the characteristics of the slice layer where the main stem was located.

Regarding the phenotypic parameters, the accuracy of the plant height, stem diameter, and main stem length reached 96.3%, 84.9%, and 95.7%, respectively. Regarding the leaf length, the leaves in different regions of this study were analyzed. The growth direction of Pinus massoniana leaves was mostly perpendicular to the main stem, so the projection method was used to project the 3D point cloud to 2D for calculation to reduce the complexity of the calculation.

Although we could not obtain the accurate regional leaf length, the estimation accuracy was 83.8%, which proved that our estimation method still had feasibility. Regarding the leaf number, by comparing the local density and local leaf number of multiple groups of samples, it was found that there was a certain linear relationship between the density and leaf number. Thus, an algorithm based on distance density was designed to estimate the leaf number. Overall, there would be approximately a 15% error, but the Pinus massoniana seedling itself had hundreds of leaves; even if artificially counted, there was still approximately a 10% error. Thus, this range of error was reasonable.

Conclusion

In this study, a low-cost 3D phenotypic system based on the Azure Kinect camera was built, and an automatic measurement method for five phenotypic parameters of Pinus massoniana with a height range of 15–30 cm was proposed. The experimental results in this study provide an efficient and economical solution for plant phenotypic feature extraction, which could promote genome research and plant breeding programs.

Our future work will focus on applying this algorithm to a three-dimensional imaging platform developed by our group. In the growth process of Pinus massoniana, our algorithms were unable to handle certain lateral stems. Moreover, if there are multiple plants in the platform at the same time and the leaves of the plants overlap, the accuracy of the algorithm will also be affected. Due to the imaging technique itself, the problem of overlapping plants and the distance between the plants and the camera is far, the information of plants collected by the imaging technique will be lacked. In this regard, we may select cameras with higher imaging accuracy, optimize the algorithms to solve the problem of overlapping stems, or use segmentation methods. This set of proposed algorithms can be applied to plant seedlings with similar structures to Pinus massoniana, such as black pine, cedar, and Pinus quinquefolius. It was proven that these algorithms can be used for these plant seedlings with some feasibility. However, every structure of plant seedlings still has slight differences, and it is also necessary to optimize the internal parameters of the algorithm and carry out more tests to improve the approaches for each plant seedling.

Data availability statement

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

Author contributions

HZ proposed and developed the approach, and HZ and YZ wrote the methods section of this article. YZ wrote other parts of the article and revised the MS. WL evaluated the accuracy of the approach. HZ designed the study. BW and ZZ improved some details of the approach. HZ acquired the 3D digitizing data. All authors contributed to the article and approved the submitted version.

Funding

Three Institute and Nine Party" Agricultural Science and Technology Cooperation Plan of Zhejiang Province(2023SNJF027), Zhejiang Key Research and Development Program (2021C02038 &2021C02061), "Pioneer" and "Leading Goose" R&D Program of Zhejiang (2023C02028), Zhejiang Province Commonweal Projects (LGG21F010003), General Projects of Agricultural and Social Development in Hangzhou (202203B06).

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.2023.1146490/full#supplementary-material

References

Besl, P. J., and Jain, R. C. (1988). Segmentation through variable-order surface fitting. *IEEE Trans. Pattern Anal. Mach. Intell.* 10 (2), 167–192. doi: 10.1109/34.3881

Bradski, G., and Kaehler, A. (2008). *Learning OpenCV: computer vision with the OpenCV library* (O'Reilly Media, Inc). https://ieeexplore.ieee.org/document/5233425/

Cao, J., Tagliasacchi, A., Olson, M., Zhang, H., and Su, Z. (2010). "Point cloud skeletons via laplacian based contraction," in *2010 Shape Modeling International Conference*. 187–197. doi: 10.1109/SMI.2010.25

Cupec, R., Vidović, I., Filko, D., and Đurović, P. (2020). Object recognition based on convex hull alignment. *Pattern Recognit.* 102, 107199. doi: 10.1016/j.patcog.2020.107199

Dalitz, C., Schramke, T., and Jeltsch, M. (2017). Iterative hough transform for line detection in 3D point clouds. *Image Process. On Line* 7, 184–196. doi: 10.5201/ipol.2017.208

de Moraes Frasson, R. P., and Krajewski, W. F. (2010). Three-dimensional digital model of a maize plant. *Agric. For. Meteorol.* 150 (3), 478–488. doi: 10.1016/j.agrformet.2010.01.003

Dziubich, T., Szymański, J., Brzeski, A., Cychnerski, J., and Korłub, W. (2016). Depth images filtering in distributed streaming. *Remote Sens. Environ.* (2), 91–98. doi: 10.1016/0034-4257(95)00224-3

Eitel, J. U., Vierling, L. A., and Long, D. S. (2010). Simultaneous measurements of plant structure and chlorophyll content in broadleaf saplings with a terrestrial laser scanner. *Remote Sens. Environ.* 114 (10), 2229–2237. doi: 10.1016/j.rse.2010.04.025

Feng, C., Taguchi, Y., and Kamat, V. R. (2014). "Fast plane extraction in organized point clouds using agglomerative hierarchical clustering," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. 6218–6225. doi: 10.1109/ICRA.2014.6907776

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 (6), 381–395. doi: 10.1145/358669.358692

Ghahremani, M., Williams, K., Corke, F., Tiddeman, B., Liu, Y., Wang, X., et al. (2021). Direct and accurate feature extraction from 3D point clouds of plants using RANSAC. *Comput. Electron. Agric.* 187, 106240. doi: 10.1016/j.compag.2021.106240

Gibbs, J. A., Pound, M., French, A. P., Wells, D. M., Murchie, E., and Pridmore, T. (2018). Plant phenotyping: an active vision cell for three-dimensional plant shoot reconstruction. *Plantphysiology* 178 (2), 524–534. doi: 10.1104/pp.18.00664

Healy, P., and Nikolov, N. S. (2001). "How to layer a directed acyclic graph," in *International Symposium on Graph Drawing*. 16–30. doi: 10.1007/3-540-45848-4_2

Hu, C., Pan, Z., and Zhong, T. (2020). Leaf and wood separation of poplar seedlings combining locally convex connected patches and K-means++ clustering from terrestrial laser scanning data. *J. Appl. Remote Sens.* 14 (1), 18502. doi: 10.1117/1.JRS.14.018502

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. doi: 10.3389/fpls.2018.00237

Kaminuma, E., Heida, N., Tsumoto, Y., Yamamoto, N., Goto, N., Okamoto, N., et al. (2004). Automatic quantification of morphological traits *via* three-dimensional measurement of arabidopsis. *Plant J.* 38, 358–365. doi: 10.1111/j.1365-313X.2004.02042.x

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Am. Math. Soc.* 7 (1), 48–50. doi: 10.2307/2033241

Li, L. (2014). Time-of-flight camera–an introduction. *Tech. white paper(SLOA190B).*

Li, J., and Tang, L. (2018). Crop recognition under weedy conditions based on 3D imaging for robotic weed control. *J. Field Robotics* 35 (4), 596–611. doi: 10.1002/rob.21763

Lin, C.-H., Kong, C., and Lucey, S. (2018). "Learning efficient point cloud generation for dense 3d object reconstruction," in *Paper presented at the proceedings of the AAAI Conference on Artificial Intelligence*.

Lin, Y., Ruifang, Z., Pujuan, S., and Pengfei, W. (2016). "Segmentation of crop organs through region growing in 3D space," in *2016 Fifth international conference on agro-geoinformatics (Agro-Geoinformatics)*. 1–6. doi: 10.1109/Agro-Geoinformatics.2016.7577669

Liu, Z., Zhang, Q., Wang, P., Li, Z., and Wang, H. J. (2020). Automated classification of stems and leaves of potted plants based on point cloud data. *Biosyst. Eng.* 200, 215–230. doi: 10.1016/j.biosystemseng.2020.10.006

Moles, A. T., Warton, D. I., Warman, L., Swenson, N. G., Laffan, S. W., Zanne, A. E., et al. (2009). Global patterns in plant height. *J. Ecol.* 97 (5), 923–932. doi: 10.1111/j.1365-2745.2009.01526.x

Nguyen, T. T., Slaughter, D. C., Max, N., Maloof, J. N., and Sinha, N. J. S. (2015). Structured light-based 3D reconstruction system for plants. *Sensors* 15 (8), 18587–18612. doi: 10.3390/s150818587

Park, J.-S. J. P. (2005). Interactive 3D reconstruction from multiple images: a primitive-based approach. *Pattern Recognition Letters* 26, 16, 2558–2571. doi: 10.1016/j.tree.2004.09.003

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

Pound, M. P., French, A. P., Murchie, E. H., and Pridmore, T. P. (2014). Automated recovery of three-dimensional models of plant shoots from multiple color images. *Plant Physiol.* 166 (4), 1688–1698. doi: 10.1104/pp.114.248971

Rosell-Polo, J. R., Cheein, F. A., Gregorio, E., Andújar, D., Puigdomènech, L., Masip, J., et al. (2015). Advances in structured light sensors applications in precision

agriculture and livestock farming. *Adv. Agron.* 133, 71–112. doi: 10.1016/bs.agron.2015.05.002

Rusu, R. B., Blodow, N., and Beetz, M. (2009). "Fast point feature histograms (FPFH) for 3D registration," in *2009 IEEE international conference on robotics and automation*. 3212–3217. doi: 10.1109/ROBOT.2009.5152473

Rusu, R. B., and Cousins, S. (2011). "3d is here: point cloud library (pcl)," in *2011 IEEE international conference on robotics and automation*. 1–4. doi: 10.1109/ICRA.2011.5980567

Silvertown, J. (2004). Plant coexistence and the niche. *Trends Ecol. Evol.* 19 (11), 605–611. doi: 10.1016/j.tree.2004.09.003

Tang, J. (2010). "A color image segmentation algorithm based on region growing," in *2010 2nd international conference on computer engineering and technology*, Vol. 6. V6–634-V636-637. doi: 10.1109/ICCET.2010.5486012

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 (14), 4628. doi: 10.3390/s21144628

Vázquez-Arellano, M., Griepentrog, H. W., Reiser, D., and Paraforos, D. S. (2016). 3-d imaging systems for agricultural applications–a review. *Sensors* 16 (5), 618. doi: 10.3390/s16050618

Wang, Y., and Chen, Y. (2020). Non-destructive measurement of three-dimensional plants based on point cloud. *Plants* 9 (5), 571. doi: 10.3390/plants9050571

Wu, S., Wen, W., Xiao, B., Guo, X., Du, J., Wang, C., et al. (2019). An accurate skeleton extraction approach from 3D point clouds of maize plants. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00248

Xiang, L., Bao, Y., Tang, L., Ortiz, D., and Salas-Fernandez, M. G. (2019). Automated morphological traits extraction for sorghum plants *via* 3D point cloud data analysis. *Comput. Electron. Agric.* 162, 951–961. doi: 10.1016/j.compag.2019.05.043

Zermas, D., Morellas, V., Mulla, D., and Papanikolopoulos, N. (2020). 3D model processing for high throughput phenotype extraction–the case of corn. *Comput. Electron. Agric.* 172, 105047. doi: 10.1016/j.compag.2019.105047

Zhang, Z. (1994). Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision* 13, 2, 119–152. doi: 10.1007/BF01427149

Zhao, K., and Popescu, S. (2009). Lidar-based mapping of leaf area index and its use for validating GLOBCARBON satellite LAI product in a temperate forest of the southern USA. *Remote Sens. Environ.* 113 (8), 1628–1645. doi: 10.1016/j.rse.2009.03.006

Zhuo, K., Cui, R., Ye, W., Luo, M., Wang, H., Hu, X., et al. (2010). Morphological and molecular characterization of aphelenchoides fujianensis n. sp.(Nematoda: aphelenchoididae) from pinus massoniana in China. *Zootaxa* 2509 (1), 39–52-39–52. doi: 10.11646/zootaxa.2509.1.3

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 21 July 2023
DOI 10.3389/fpls.2023.1168732

Check for updates

OPEN ACCESS

EDITED BY
Jiangang Liu,
Chinese Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Yanbo Huang,
United States Department of Agriculture (USDA), United States
M. R. Umesh,
University of Agricultural Sciences Raichur, India

*CORRESPONDENCE
Raju Bheemanahalli
rajubr@pss.msstate.edu

RECEIVED 18 February 2023
ACCEPTED 26 June 2023
PUBLISHED 21 July 2023

CITATION
Shrestha A, Bheemanahalli R, Adeli A, Samiappan S, Czarnecki JMP, McCraine CD, Reddy KR and Moorhead R (2023) Phenological stage and vegetation index for predicting corn yield under rainfed environments.
Front. Plant Sci. 14:1168732.
doi: 10.3389/fpls.2023.1168732

COPYRIGHT
© 2023 Shrestha, Bheemanahalli, Adeli, Samiappan, Czarnecki, McCraine, Reddy and Moorhead. 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.

Phenological stage and vegetation index for predicting corn yield under rainfed environments

Amrit Shrestha[1], Raju Bheemanahalli[2*], Ardeshir Adeli[3], Sathishkumar Samiappan[4], Joby M. Prince Czarnecki[4], Cary Daniel McCraine[4], K. Raja Reddy[2] and Robert Moorhead[4]

[1]Department of Agricultural & Biological Engineering, Mississippi State University, Mississippi State, MS, United States, [2]Department of Plant and Soil Sciences, Mississippi State University, Mississippi State, MS, United States, [3]United States Department of Agriculture-Agricultural Research Service (USDA-ARS), Genetics and Sustainable Agriculture Research Unit, Mississippi State, MS, United States, [4]Geosystems Research Institute, Mississippi State University, Mississippi State, MS, United States

Uncrewed aerial systems (UASs) provide high temporal and spatial resolution information for crop health monitoring and informed management decisions to improve yields. However, traditional in-season yield prediction methodologies are often inconsistent and inaccurate due to variations in soil types and environmental factors. This study aimed to identify the best phenological stage and vegetation index (VI) for estimating corn yield under rainfed conditions. Multispectral images were collected over three years (2020-2022) during the corn growing season and over fifty VIs were analyzed. In the three-year period, thirty-one VIs exhibited significant correlations ($r \geq 0.7$) with yield. Sixteen VIs were significantly correlated with the yield at least for two years, and five VIs had a significant correlation with the yield for all three years. A strong correlation with yield was achieved by combining red, red edge, and near infrared-based indices. Further, combined correlation and random forest an alyses between yield and VIs led to the identification of consistent and highest predictive power VIs for corn yield prediction. Among them, leaf chlorophyll index, Medium Resolution Imaging Spectrometer (MERIS) terrestrial chlorophyll index and modified normalized difference at 705 were the most consistent predictors of corn yield when recorded around the reproductive stage (R1). This study demonstrated the dynamic nature of canopy reflectance and the importance of considering growth stages, and environmental conditions for accurate corn yield prediction.

KEYWORDS

corn, cover crop, phenology, remote sensing, vegetation indices, yield

1 Introduction

Feeding the growing population is an urgent challenge for the agriculture sector. However, over-fertilization to improve yield has resulted in negative consequences like water contamination and deterioration of soil health (Halliday and Wolfe, 1991; Singh, 2018). Fertilizer efficiency can be improved by applying the right amount at the right time based on

the crop's seasonal needs, and variable rate technologies (VRT) have that potential. For VRT, input decisions are usually based on a prescription map based on either intensive soil sampling (Kitchen et al., 2005; Servadio et al., 2017) or historic yield information (Lark, 1998; Jaynes et al., 2005). Spatially dense soil sampling (i.e., ≤ 2.5-acre grid which is the industry standard) is costly and not guaranteed to accurately represent soil conditions across space. At the same time, historic yield information is not reliable information for in-seasonal management decisions as yield varies from season to season (Maestrini and Basso, 2018). Therefore, in-season estimation of yield can improve management decisions for agricultural inputs.

It is estimated that 1.2 billion tons of corn was produced globally in 2021 (FAO, 2022) and the United States alone produced 382.9 million tons, worth of 82.6 billion USD (USDA, 2022). Corn is one of the important cereal crops, which is cultivated across the globe with the highest fertilization rate (IFA, 2022). The requirement for agricultural inputs (fertilizer and water) can be estimated by monitoring canopy optical properties. Green canopy has high absorption in the visible portion of the light spectrum and high reflectance in the near-infrared portion (Ustin and Jacquemoud, 2020). Plants absorb the maximum amount of light in the visible portion for photosynthesis where blue (B) and red (R) are strongly absorbed by the chlorophyll and carotenoid pigments (Lichtenthaler, 1987; Ustin and Jacquemoud, 2020), while, red edge (RE) and near-infrared (NIR) values are found to be associated with plant health (Horler et al., 1983; Fahrentrapp et al., 2019; Zahir et al., 2022). Additionally, leaf optical signatures are highly responsive to changes in soil and canopy nitrogen (Wood et al., 1992; Ziadi et al., 2008; Yang et al., 2012). Any detrimental events that change chlorophyll pigment reduce the potential assimilation capacity which adversely affects the growth, development and yield (Bheemanahalli et al., 2023). Consequently, changes in crop biochemical and physiological properties can be measured by monitoring canopy reflectance properties. The advancement of science and technologies has revolutionized high-throughput phenotyping, especially with UASs. The greatest boon has been its ability to collect high-resolution spatial and temporal information quickly, as well as flexibility on revisiting time and choice of sensor. Such high-resolution data can divulge the spatial and temporal variability (Matese et al., 2015) present in the crop during the growing season which can be integrated with VRT for management decisions. In addition, the adoption of VRT in combination with in-season spectral data for nitrogen application has promoted an increase in profitability and nitrogen saving as compared to traditional practices (Kitchen et al., 2010; Scharf et al., 2011).

A UAS comprises an uncrewed aerial vehicle (UAV), potentially one or more sensors or payloads, and a communicating and controlling device. The choice of sensor is an important consideration. A multispectral sensor that consists of blue, green (G), red, red edge, and near-infrared bands has been the first choice for studying crop bio-physiological parameters for many researchers (Yao et al., 2019; Bheemanahalli et al., 2022; Bheemanahalli et al., 2023). Various other spectral, thermal, and light detection and ranging (LiDAR) sensors have been deployed on UAV-based plant phenotyping (Xie and Yang, 2020). Both spectral and thermal sensors have been used for studying plant health (Geipel et al., 2014; Simic Milas et al., 2018) and LiDAR data has been used for plant structure (Yuan et al., 2018; Wu et al., 2019; ten Harkel et al., 2020; Luo et al., 2021). Santana et al. (2021) deployed a four-band multispectral sensor mounted on a UAV for the estimation of corn yield. Vong et al. (2021) used a red-blue-green (RGB) sensor to estimate corn stand count. Barzin et al. (2020) used a five-band multispectral sensor (blue, green, red, RE, NIR) for the estimation of corn yield. Chivasa et al. (2021) used a four-band multispectral sensor (green, red, RE, NIR) for the prediction of corn streak virus severity and yield. Zhang et al. (2019) used hyperspectral imagery collected from a UAV for detecting yellow rust disease in wheat. Crusiol et al. (2020) used a thermal sensor mounted on a UAV for determining water status of soybean plants.

The sensor is mounted on a UAV and deployed to collect the imagery data. These data are generally transformed into a ratio or combination of normalized differences of spectral bands known as a vegetation index (VI) to study crop health and performance. Vegetation indices (VIs) have been found highly related to physiological parameters (Bannari et al., 1995; Xue and Su, 2017; Ma et al., 2019) and less sensitive to atmospheric effects compared to spectral bands (Myneni and Asrar, 1994; Gitelson et al., 2002). The normalized difference vegetation index (NDVI) is commonly used to measure vegetation health. It exhibits a strong association with leaf dry biomass (Kross et al., 2015), yield (Hassan et al., 2019; Maresma et al., 2020), and leaf area index (LAI) (Shafian et al., 2018). However, its usefulness has been limited due to saturating tendency at the higher LAI (Sellers, 1985; Carlson and Ripley, 1997). Various other VIs are found to have a strong association with crop bio-physiological parameters: normalized difference spectral index (NDSI) and soil adjusted vegetation indices (SAVI) are found to have a strong correlation with leaf nutrition index (Zhao et al., 2018), modified normalized difference water index (MNDWI) with LAI (Zarate-Valdez et al., 2012), MERIS terrestrial chlorophyll index (MTCI) with chlorophyll (Dash and Curran, 2007), wide dynamic range vegetation index (WDRVI) with yield (Maresma et al., 2016), and crop water stress index (CWSI) with water status (Gonzalez-Dugo et al., 2014; Santesteban et al., 2017). In addition, different techniques were used to select vegetation indices based on variance inflation factor, recursive feature elimination, random forest, correlation, Bayesian variable selection, and genetic algorithm for yield estimation (Maya Gopal and Bhargavi, 2019; Barzin et al., 2020; Aditya Shastry and Sanjay, 2021; Saravi et al., 2021; Barzin et al., 2022). A random forest was widely used to select VIs to estimate different crop traits (Maya Gopal and Bhargavi, 2019; Barzin et al., 2020; Li et al., 2022; Luo et al., 2022) due to its robustness to outliers and noise (Breiman, 2001).

To our best knowledge, none of the studies have reported suitable VIs that can be used throughout the growing season to predict yield in corn. Thus, the objectives of our research were to i) identify the potential VIs that significantly correlate with yield across the growing season, ii) identify the suitable phenological stage for yield prediction and iii) identify the best predictive VI for yield estimation under rainfed environments.

2 Materials and methods

2.1 Site description

The field experiment was conducted for three consecutive years 2019-2022 in the R. R. Foil Plant Science Research Center, Mississippi State University (33°28'21.0"N 88°46'25.5"W), Figure 1.

2.2 Experimental design and crop husbandry

Corn - and - cotton was grown as a rotation cash crop during the growing season under rainfed conditions. Best management practices were adopted to establish the crops: cover crops were grown during the fallow period with minimum or no tillage and cash crop was planted in mid-April. The experiment consisted of three cover crop treatments, Austrian winter pea (*Pisum sativum* L.), Daikon radish (*Raphanus sativus* L. subsp. *longipinnatus*), cereal rye (*Secale cereale* L. var. Elbon), and a no cover crop (NCC) treatment. In 2022, radish was replaced with a mixture of winter pea, radish, and rye. The experiment was designed as a split-plot randomized complete block, with the cash crop as the main factor and the cover crop species as the sub-factor. Each subplot had four rows of 90 m length × 3.8 m width, with four replicates (Figure 1). Cover crops were planted in October and terminated in March across three years. Prior to the cover crop planting, two tons

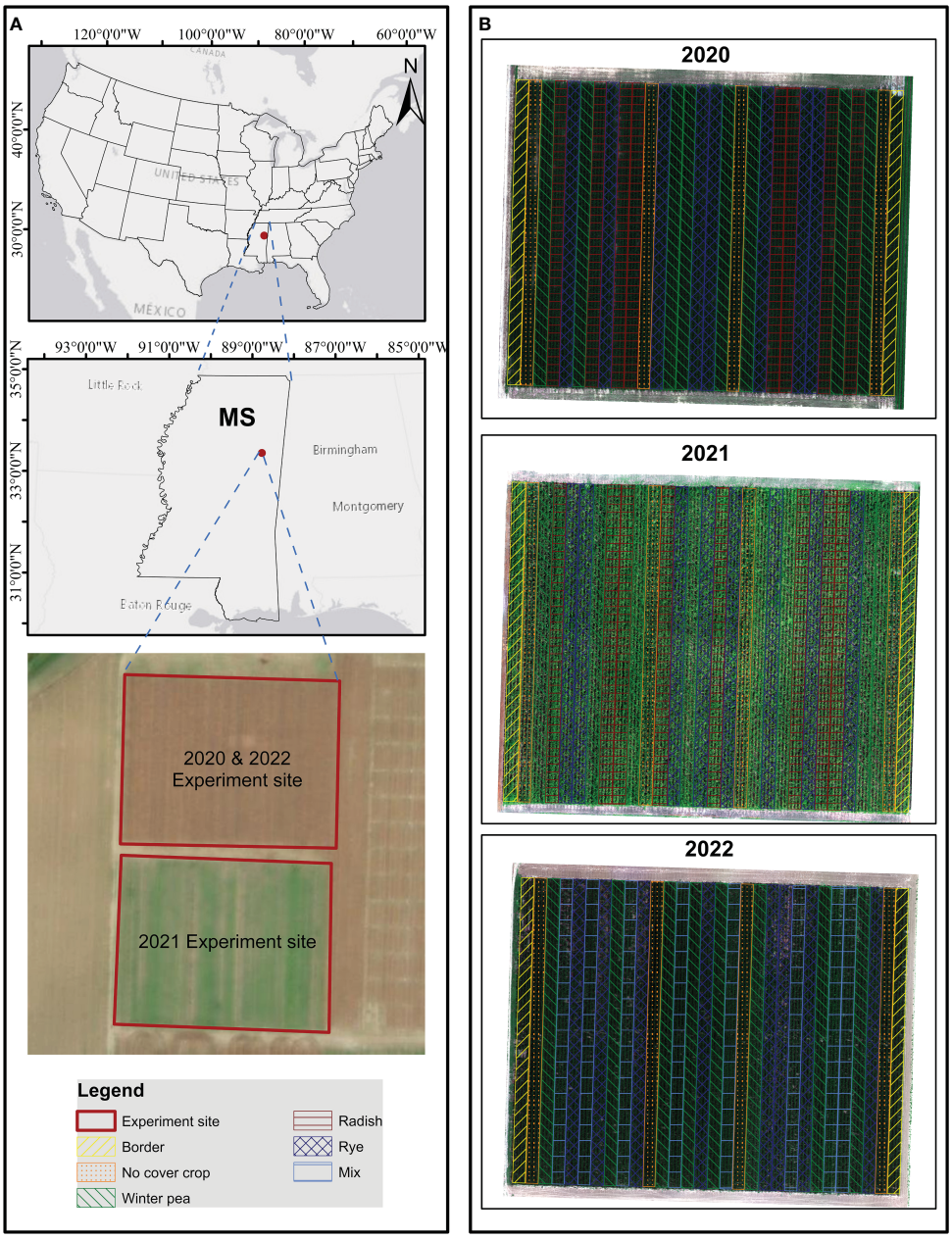


FIGURE 1
Research site location **(A)** and experiment design **(B)**.

per acre of poultry litter was surface broadcasted. After the termination of the cover crop, corn was planted on April 14, 2020, April 07, 2021, and March 28, 2022. A split fertilization procedure was adopted and divided into two parts. No cover crop (NCC) treatment received 56 kg N ha^{-1} as a starter fertilizer at the V3 and 168 kg N ha^{-1} at the V12 growth stage. The cover crop treatments did not receive starter fertilizer, but 168 kg N ha^{-1} was applied to cover crop treatments at the V12 growth stage. The source of N fertilizer was urea ammonium nitrate (UAN solution, 32% N). Corn plots were harvested in September 2020-2022 after completing physiological maturity. It is important to note that, in this study, the cover crop treatment's influence on corn yield was not the major focus. Instead, the study aims to determine the ideal phenological stages and likely highest predictive power VIs for corn yield prediction under rainfed conditions.

2.3 Data collection

2.3.1 Yield

The middle two rows of each subplot were harvested using a mechanical corn harvester and the yields were adjusted to a 15.5% moisture level. The obtained yield was then converted into the mega-gram per hectare (Mg ha^{-1}).

2.3.2 Data acquisition

A UAV mounted with a five-band multispectral camera (Rededge MX, Micasense Inc., Seattle, USA) was flown at an altitude of 61 m (200 ft) above ground level producing a spatial resolution of approximately 4 cm. The sensor has the following central wavelengths(bandwidths): 475(20) nm, 560(20) nm, 668(10) nm, 717(10) nm, and 842(40) nm for blue, green, red, RE, and NIR bands, respectively. A mission planning software (DJI GSPro, DJI LLC., Shenzen, China) was used to create the UAV flight plan in a single grid pattern. Images were acquired by setting the sensor in automatic exposure mode with 80% frontal overlapping and 70% side overlapping. The data was stored in 16-bit raw format. Before and after each flight, an image of the reference reflectance calibration tile provided by the sensor manufacturer was taken, which was later used for reflectance calibration following guidance from the sensor manufacturer. The data were collected weekly throughout the year as the weather permitted within ±2 hours of local solar noon. The UAV data were collected between the vegetative stage (V3) and the reproductive stage (R5) as shown in Supplementary Table S1. Specifically, data were collected at various time points, including V5, V6, V7, V10, V11, and V13, which represent the number of visible leaf collars on the main stem, where Vn refers to leaf collars greater than 13. In addition, data were collected at the reproductive stage, including R1, R2, R3, R4, and R5, which correspond to the silk, blister, milk, dough, and dent stages, respectively (Nleya et al., 2016).

2.4 Data preprocessing

A commercial image stitching software Pix4D Mapper (Pix4D SA, Lausanne, Switzerland) was used to generate a single orthomosaic image of the whole field. The mosaic image was geo-rectified using the ground control points that were placed around the edge of the experimental field. The accuracy of the GPS device (Trimble Geo7x, Trimble Inc., CA, USA) was ±2cm. The image digital number was converted into reflectance value using the reference reflectance calibration tile provided with the sensor. To extract the corn pixels, a support vector machine algorithm was employed using geospatial software (ENVI, version 5.6, Exelis Visual Information Solution, Boulder, CO, USA). The orthomosaic image can be classified into three major classes: corn, soil and weed pixels. For each class, at least five thousand pixels were selected for classification. The accuracy for the classification of corn pixels for all the images was greater than 0.95.

2.5 VI extraction and selection

Orthomosaic images that were affected by clouds were removed from the analysis. After removing the background pixels (i.e., soil, stubble and weeds) from orthomosaic, the middle two rows of each subplot were first digitized and split into individual subplots using ArcGIS (Environmental Systems Research Institute, Inc. (ESRI), Redland, CA, USA). Each subplot was loaded in Python using the GDAL library to compute the average spectral information. This spectral information was then used to calculate the VIs that are related to crop health, photosynthesis, biochemical, and physiology. The list of VIs used in this study is given in Supplementary Table S2. Correlation analysis between VIs and yield was performed for each flight or during the corn growing season. Further, to select the VI with the best predictive power we followed three criteria: i) any VI that had a significant correlation (absolute value of correlation coefficient $r \geq 0.7$) with yield for each date was assigned a score of 1, ii) only those VIs that scored 1 were summarized in each year, and iii) the top three VIs with the maximum cumulative score were identified as stable across stages of growth as well as the VIs with the best predictive power. We repeated the same analysis for three growing seasons and selected the VIs that were unique and common across years as the best predictors of yield. In previous studies (Asuero et al., 2006; Akoglu, 2018; Kogan et al., 2018), a correlation coefficient with an absolute value greater than 0.7 has been widely regarded as a strong correlation. Therefore, we chose a threshold limit of 0.7 to assign a score of 1.

In addition, to determine the best VI at a given phenology stage, a machine learning technique based on random forest (RF) algorithm was implemented. A RF is collection of decision trees that are trained independently. The prediction from each decision tree is averaged to a single output (Breiman, 2001). Initially, the RF model parameters were tuned by implementing a randomized search cross-validation method. In this method, a set of predefined values were randomly chosen as a possible candidate, and a hundred sets of hyperparameters were considered. Next, cross-validation was performed for each set of hyperparameters, and five-fold cross-validation was chosen. Then finally, the model returns the best set of hyperparameters. The tuned hyperparameters were the number of trees, number of samples required to split the internal node, minimum samples required to be a leaf node, and

maximum feature to be considered while splitting the node (full feature and square root of total number of feature). A bootstrap technique was selected for generating tree samples. After tuning the RF model, the model was fitted with the best set of hyperparameters. To rank the variables, a permutation-based feature importance method was implemented utilizing all the variables, and a simple linear model was developed with the top-ranked variable. Geospatial software, ArcGIS and ENVI were used for data manipulation. The modeling was done with Python (version 3.9) using the stats models (version 0.13.5) library for linear regression analysis and the sci-kit-learn library (version 1.0.2) for random forest analysis.

2.6 Yield modeling

A simple linear regression model was employed to develop a linear relationship between corn yield and selected variables as discussed above. The model performance was compared with the coefficient of determination (R^2) and mean absolute percentage error (MAPE). The variable with high R^2 with significantly lower MAPE ($p<0.01$) was considered the highest predictive power VI across the growing season. The modeling was done with Python (version 3.9) using the sci-kit-learn library (version 1.0.2).

3 Results

During the corn growing season in 2020, 2021 and 2022, the observed average air temperature was 24.2, 23.7, and 24.7°C, respectively, and precipitation was 523 mm (optimum), 859 mm (high) and 481 mm (low), respectively (Figure 2). The year 2022 was observed to be relatively drier than 2020 and 2021. There were significant differences in corn yield between treatments and years (Figure 3). Plots with winter pea and cereal rye treatments recorded higher yields than radish treatment in 2020. Cover cropping treatments (radish or mix and rye) yielded significantly lower than control plots in 2021 and 2022 except winter peas. The growing season with optimum rainfall (2020) was associated with a higher yield, while maximum rainfall (2021) during the growing season was associated with a lower yield. The annual mean yield for 2020, 2021 and 2022 were 10.72, 6.34 and 9.22 Mg ha^{-1}, respectively (Figure 3).

3.1 Correlation analysis

Results indicated a set of VIs (12 in 2020 and 2021, and 28 in 2022) with correlation coefficients $|r| \geq 0.7$ at least twice in the same growing season (Figure 4, Supplementary Table S3). In 2020, MTCI and leaf chlorophyll index (LCI) had the strongest positive correlation with the yield at four timings (R1, R2, R3, and R5 growth stages) and were ranked first. The modified normalized difference at 705 (mND705), modified simple ratio at 705 (mSR705), and red edge chlorophyll reflectance index (RIrededge) had significant correlations with corn yield at three timings. While

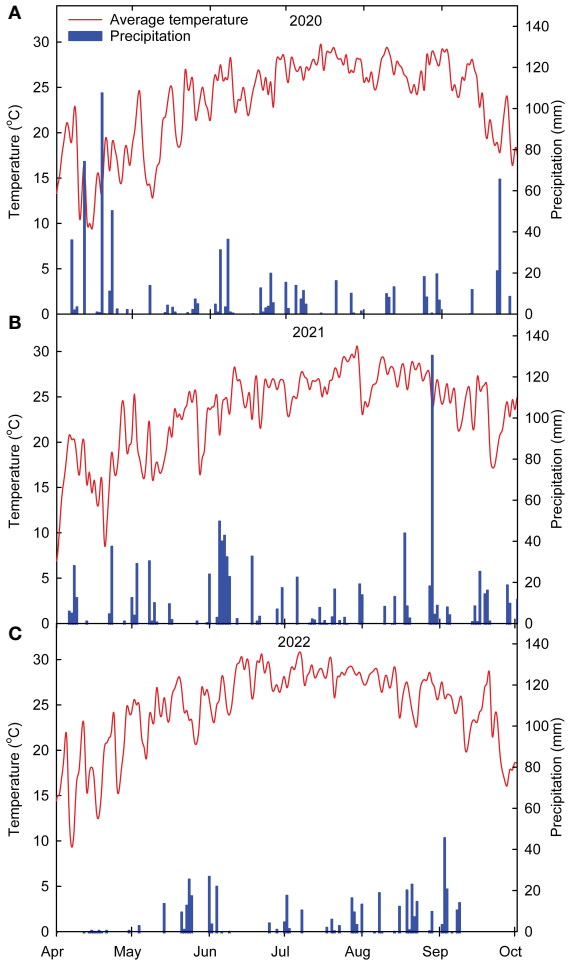


FIGURE 2
Observed daily air temperature and precipitation variability during the experimental period (2020 – **(A)**, 2021 – **(B)** and 2022 – **(C)**. Weather data were obtained from the Delta Agricultural Weather Center (http://deltaweather.extension.msstate.edu/) for the experiment site.

seven VIs such as enhanced vegetation index (EVI), modified chlorophyll absorption ratio index (MCARI2), soil adjusted vegetation index (SAVI), modified SAVI (MSAVI), optimized SAVI (OSAVI), renormalized difference vegetation index (RDVI), and green chlorophyll reflectance index (RIgreen) had significant correlations with yield at only two timings. The MCARI2 showed the highest correlation with yield at R1 growth stage (r=0.81) (Figure 4). In 2021, CVI, LCI and MTCI had strongest correlations across all (V5, V7, V11, Vn, R1, R2, and R5) growth stages (first tier). The second tier was dominated by RE-based VIs, while third tier includes green chlorophyll index (CIgreen) and triangular greenness index (TGI) (Figure 4). A strong negative correlation (r=-0.91) was noted between TGI and the yield at R5 (119 DAP). At R1 (83 DAP), the highest correlation (r=0.96) between yield and MTCI was observed. In 2022, there were 20 VIs that fall under top tier. This group was dominated with RE-based indices and showed a strong correlation with yield across the six (V6, V10, V19, V1, V3 and V5) growth stages. The second tier VIs such as EVI, LCI and MTCI showed significant correlations with yield for five times. The third-tier VIs

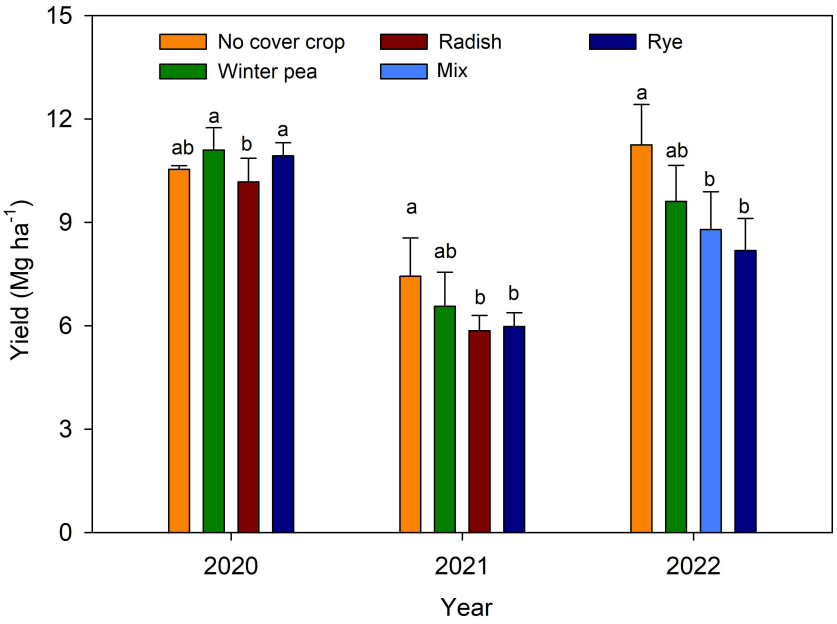


FIGURE 3
Corn yield variability in response to cover crop treatments. Vertical bars denote mean ± standard deviation (SD). Bars with common letters are not statistically significant at $p<0.05$ (LSD test) within a year.

were TCARI/OSAVI, TVI, ARI CRIrededge and mARI. Further, green normalized difference vegetation index and leaf chlorophyll index showed strong positive correlations (r=0.91) with yield at Vn (67 DAP) and R1 (78 DAP), respectively. At R5 (108 DAP), a significant negative correlation (r=-0.95) was found between SR445 and yield (Figure 4).

3.2 Unique and common VIs across the year for yield prediction

Across all three years, the five VIs (mND705, mSR705, RIrededge, LCI, and MTCI) with the highest predictive power were identified (Figure 5). In 2021, two unique VIs were identified, while

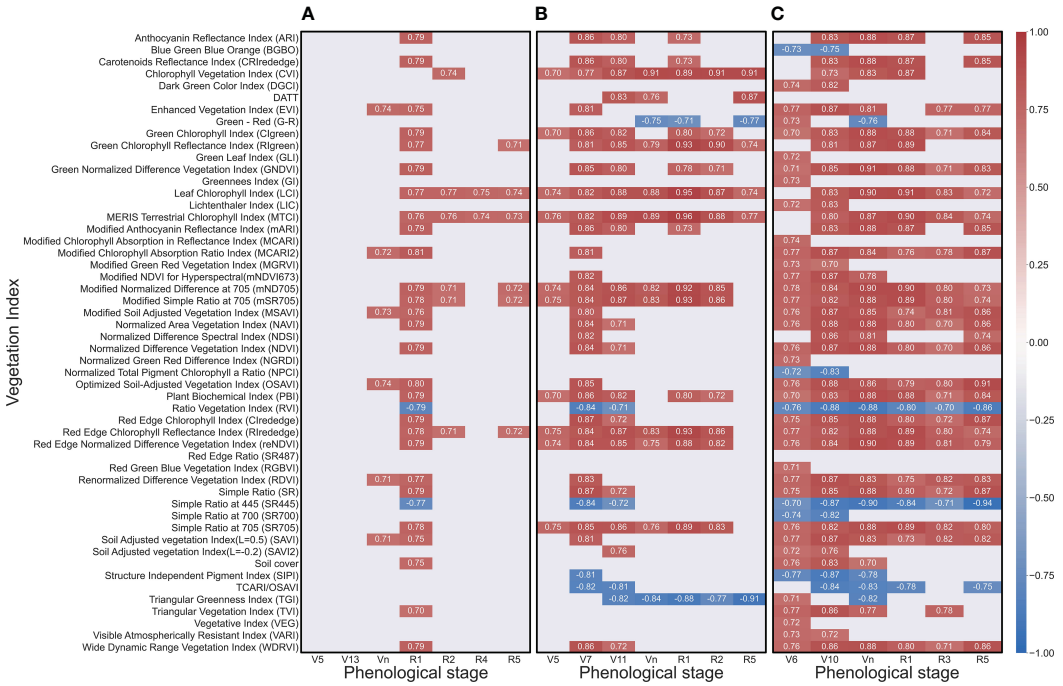


FIGURE 4
Heat map showing the significant correlation coefficient (r≥0.7) between corn yield and vegetation index at different growth stages in 2020 **(A)**, 2021 **(B)** and 2022 **(C)**. Darker color signifies a high correlation, while lighter color signifies a weaker correlation between yield and corresponding vegetation index.

thirteen were noted in 2022. Six VIs (EVI, MCARI2, MSAVI, OSAVI, RDVI and SAVI) were unique to 2020 and 2022; CIgreen, PBI, reNDVI and SR705 were unique to 2021 and 2022. RIgreen was common in both 2020 and 2021 (Figure 5). Five VIs common to all growing seasons were used to build linear regression models for each growing season (Table 1). The results of the models revealed improvement in yield estimation as the season progressed with the best prediction at the R1 stage: with mND705 (R^2 = 0.62) in 2020, with MTCI (R^2 = 0.92) in 2021, and with LCI (R^2 = 0.82) in 2022. Thereafter, the performance became relatively weaker (Table 1). The R^2 and mean absolute percentage error (MAPE) of various models ranged from 0.04 to 0.92 and 3.01% to 11.85%, respectively (Table 1). A higher error was generally noted at the early or late growth stages. Although the association was weaker, RIrededge, mSR705 and mND705 performed better than LCI and MTCI at the early growth stage. However, mND705 had higher predictive power throughout the growing season compared to RIredege and mSR705 (Table 1). In addition, MTCI had 8 times the highest performance (high R^2 with low MAPE), while LCI and mND705 had the highest performance for six times. All common VIs that are sensitive to changes in chlorophyll content had the ability to separate cover crop treatments at the R1 growth stage across years (Figure 6). In general, higher values of VIs at the R1 stage were associated with higher yields and vice versa. In 2021, all VIs followed a similar trend of yield response to cover crop treatment. At the reproductive stage, RIrededge, mSR705 and MTCI had a higher ability to differentiate cover crop treatments (winter pea and cereal rye) from the control (no cover crop) under rainfed conditions (Figure 6). Lower MTCI values in radish or mix and rye treatments at R1 were associated with lower yields (Figure 6E).

3.3 Variable assessment

A RF machine learning algorithm was implemented to identify the VIs with the highest predictive power (best) at different growth stages (Table 2). The R^2 and MAPE of the various model considered with the RF method ranged from 0.25-0.89 and 2.82-9.20%, respectively (Table 2). In 2020, the highest predictive power of VIs at V5 and V13 were the visible atmospherically resistant index (VARI), while at Vn, R1, R2, R4, and R5 were visible MSAVI, soil cover, triangular vegetation index (TGI), LCI, and MTCI, respectively. In 2021, the most predictive power VIs were MTCI (V5), RVI (V7), DATT (V11), CVI (Vn and R1), LCI (R2), and TGI (R5). However, in 2022, the VIs with the highest predictive power were RIgreen, TCARI/OSAVI, SR445, CVI, SAVI, and SR445, corresponding to growth stages V6, V10, Vn, R1, R3, and R5. Particularly, the reproductive stage exhibited the strongest prediction power for these VIs. At the R1 stage, the best variables ranked by the RF method had a strong association with yield (Table 2): soil cover (R^2 = 0.56) in 2020, CVI (R^2 = 0.82) in 2021, and CVI (R^2 = 0.76) in 2022. However, at the same growth stage, the VIs selected from correlation analysis had higher predictive power in estimating yield: (mND705, R^2 = 0.62) in 2020, (MTCI, R^2 = 0.80) in 2021, and (LCI and mND705, R^2 = 0.81) in 2022 (Table 1).

4 Discussion

Crop yield is a complex result of genetics, environmental factors, and management practices. Researchers aim to improve

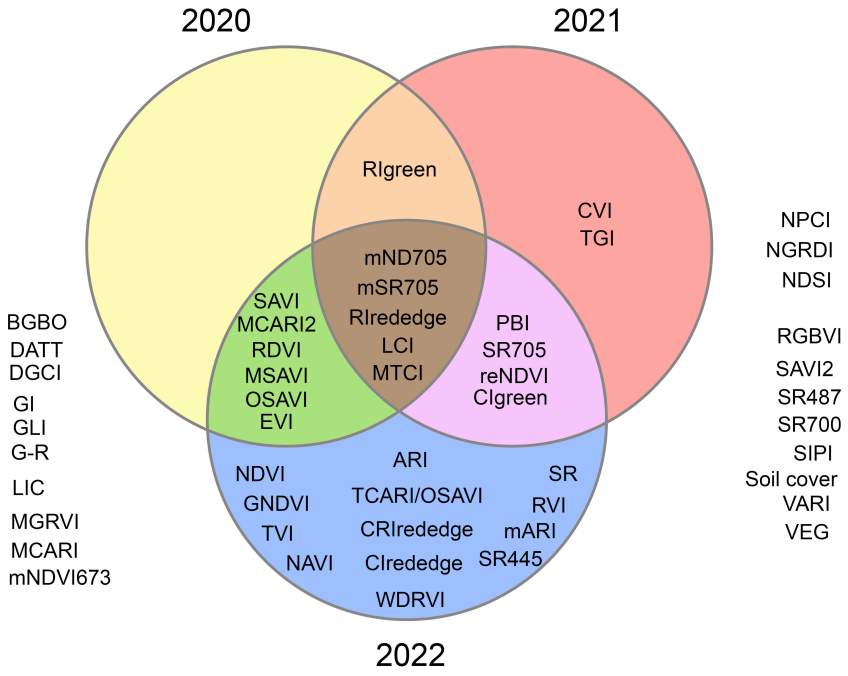


FIGURE 5
A Venn diagram showing the vegetation indices (VIs) with the highest predictive power within (unique VIs) or between years (common VIs). Full form of acronyms is given in Figure 4 and the formulation are given in Supplementary Table 2.

TABLE 1 Association between stable VIs collected at different phenological stages and corn yield under rainfed environments.

Growth stage	LCI		MTCI		RIrededge		mSR705		mND705	
	R^2	MAPE	R^2	MAPE	R^2	MAPE	R^2	MAPE	R^2	MAPE
2020										
V5	0.04ns	4.65	0.04ns	4.65	**0.18***	**4.07**	**0.18***	**4.07**	**0.18***	**4.07**
V13	**0.24****	**4.26**	0.21*	4.36	0.17**	4.46	0.17*	4.46	0.19*	4.39
Vn	0.47***	3.44	0.43***	3.58	0.44***	3.57	0.44***	3.57	**0.48*****	**3.42**
R1	0.59***	3.07	0.58***	3.12	0.60***	3.03	0.60***	3.03	**0.62*****	**2.99**
R2	**0.59*****	**3.12**	0.58***	3.18	0.51***	3.42	0.51***	3.42	0.51***	3.4
R4	**0.56*****	**3.18**	0.55***	3.23	0.46***	3.54	0.46***	3.54	0.47***	3.51
R5	**0.55*****	**2.97**	0.54***	3.01	0.51***	3.17	0.51***	3.17	0.52***	3.17
2021										
V5	0.55***	7.05	**0.58*****	**6.85**	0.57***	7.02	0.57***	7.02	0.55***	7.13
V7	0.68***	5.64	0.68***	5.73	0.70***	5.5	0.70***	5.5	**0.70*****	**5.45**
V11	0.77***	5.16	**0.79*****	**4.89**	0.76***	5.31	0.76***	5.31	0.74***	5.49
Vn	0.78***	5.14	**0.80*****	**4.81**	0.69***	5.92	0.69***	5.92	0.67***	6.15
R1	0.90***	3.51	**0.92*****	**3.12**	0.86***	3.87	0.86***	3.87	0.84***	4.16
R2	0.76***	4.81	**0.77*****	**4.72**	0.74***	5.12	0.74***	5.12	0.72***	5.35
R5	0.55***	7.42	**0.59*****	**7.0**	0.49***	7.49	0.49***	7.49	0.46***	7.72
2022										
V6	0.09ns	11.85	0.09ns	11.85	0.60***	7.8	0.60***	7.8	**0.60*****	**7.71**
V10	0.69***	7.39	0.65***	7.72	0.67***	7.44	0.67***	7.44	**0.71*****	**7.11**
Vn	**0.81*****	**5.51**	0.76***	6.29	0.77***	6.18	0.77***	6.18	0.81***	5.53
R1	**0.82*****	**5.79**	0.81***	5.89	0.79***	6.16	0.79***	6.16	0.80***	6.09
R3	0.70***	7.13	**0.70*****	**7.06**	0.64***	7.88	0.64***	7.88	0.63***	8.05
R5	0.53***	8.68	**0.54*****	**8.55**	0.54***	8.65	0.54***	8.65	0.53***	8.76

Leaf Chlorophyll Index (LCI), MERIS Terrestrial Chlorophyll Index (MTCI), Red Edge Chlorophyll Reflectance Index (RIrededge), Modified Simple Ratio at 705 (mSR705) and Modified Normalized Difference at 705 (mND705). *, ** and *** indicate the regression is significant at $p<0.05$, $p<0.01$ and $p<0.001$ respectively. 'ns' indicates non-significant. MAPE: mean absolute percentage error, DAP: day after planting. Bold values represent the best VI for a given growth stage.

yield through strategic selection of genotypes/hybrids, while farmers focus on optimizing inputs. In both scenarios, the common goal is to maximize crop yield, which is viewed as important because of its association with economic value. Studies, including the one presented herein, show significant variability in yield from year to year, with greater variability within a year rather than between years (Dhillon et al., 2022). This variability is primarily due to environmental factors such as soil (Bresler et al., 1981) and rainfall (Or and Hanks, 1992). To complement the ongoing season-based studies to make more informed management decisions, this study identified suitable growth stages and VI/s for predicting corn yield. The intention is that such a method can serve as an alternative and improvement to relying solely on historical data.

Mapping yield early help diagnose plant health or yield variability and create management zones for in-season decision making. For example, with this knowledge, a farmer can allocate sampling resources to regions of interest, reducing overall cost while improving knowledge gained from the activity. With corn being heavily reliant on fertilization, the yield map can serve as a prescription map for VRT for agricultural input management. A UAS can collect information at high resolution in contrast to satellite remote sensing, which can be beneficial for finer-scale management. Nevertheless, the extraction of information from imagery and the selection of VI have a significant effect on yield prediction or mapping. This study used a correlation-based feature selection method and compared it to a random forest approach. Results indicated that different variables are important at different growth stages and can vary by year (Table 1, 2), as seen in previous studies (Barzin et al., 2020). In addition, the results of feature selection using correlation analysis demonstrated an ability to make early predictions, compared to the RF approach.

This three-year study indicated that the VIs' ability to predict yields is weaker at the early vegetative stage compared to the

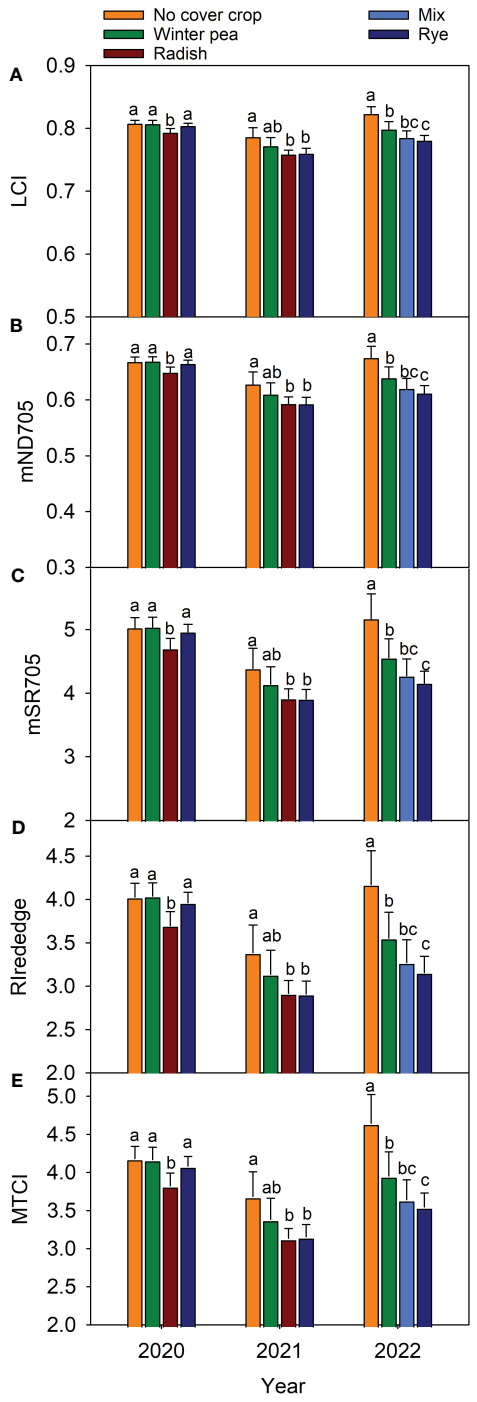


FIGURE 6
Influence of cover crop on the spectral properties of corn at R1 growth stage. Leaf Chlorophyll Index (LCI, **A**), Modified Normalized Difference at 705 (mND705, **B**), Modified Simple Ratio at 705 (mSR705, **C**), Red Edge Chlorophyll Reflectance Index (RIrededge, **D**), and MERIS Terrestrial Chlorophyll Index (MTCI, **E**). Vertical bars denote mean ± SD. Treatments with common letters are not statistically significant at $p<0.05$ (LSD test) within a year.

reproductive stage due to slower growth and canopy. This study identified five VIs (LCI, MTCI, mND705, mSR705, and RIrededge) that had a significant correlation with yield across years (Figure 5). The commonality in these VIs are NIR and RE bands. Although reNDVI is based on NIR and RE, this VI was not selected due to a weaker correlation in 2020 (Figure 4). However, this VI has been correlated with canopy chlorophyll and LAI around the reproductive stage in corn (Simic Milas et al., 2018). In addition, Li et al. (2014) and Zhao et al. (2007) reported that three-band indices were better estimators of plant nitrogen concentration and uptake compared, and LAI and above-ground biomass, respectively, compared to two-band indices. In this study, LCI, MTCI and mND705 were found to be superior compared to mSR705 and RIrededge in yield estimation at the R1 stage (Table 1).

Chlorophyll content in leaves reaches its peak during the R1-R3 stages and is a vital pigment for photosynthesis (Schepers et al., 1992; Brewer et al., 2022). Thus, LCI (Datt, 1999), MTCI (Dash and Curran, 2004) and mND705 (Sims and Gamon, 2002) have been used as proxies for canopy greenness or chlorophyll content. High correlations were reported for LCI with chlorophyll *a* (r = 0.86), and chlorophyll *a+b* (r=0.84) (Datt, 1999). Similarly, MTCI had strong relationships with chlorophyll content (Dash et al., 2010), photosynthesis (Maleki et al., 2020), nitrogen uptake (Li et al., 2021) and yield (Zhang and Liu, 2014). mND705 has also been utilized to track the senescence dynamics of wheat accessions (Anderegg et al., 2020). The blue reflectance signal from crops is a combination of chlorophyll and carotenoid pigments, while the red reflectance is mainly dominated by chlorophyll. The prevalence of chlorophyll and photosynthesis information in MTCI likely led to better prediction results when combined with the RE and NIR bands, compared to the combination of blue, RE, and NIR in mND705, mSR705, and RIrededge. Further, the prediction ability of MTCI were comparable or better than the prediction ability of the preferred VIs by the RF algorithm.

We investigated the use of various VIs as indicators of canopy greenness or photosynthetic pigments to examine crop health (Figure 6) and yield (Figure 3) responses to management. Identified five high predictive VIs (LCI, mND705, mSR705 RIrededge and MTCI) found to be linked to the plant's ability to capture and use light energy for growth and development (Boyd et al., 2011; Dong et al., 2015; Barnes et al., 2017; Tan et al., 2018). Moreover, these VIs are also sensitive to changes in chlorophyll content and canopy structure (Sims and Gamon, 2002; Wang et al., 2017; Croft et al., 2020). For instance, a higher LCI value indicates greater photosynthetic efficiency and nitrogen content in corn, which is often associated with higher yields. On the other hand, lower MTCI values under radish or mix and rye treatments at R1 indicate lower chlorophyll content and poor canopy structure, thus lower yields compared to other treatments (Figure 6). Similarly, higher values of other VIs were also associated with increased chlorophyll content and greater yields across years. It is evident that VIs that are sensitive to changes in pigments can assist in mapping differences in plant health and yield potential in corn in response to cover cropping systems under rainfed conditions (Supplementary Table 4). Identified five promising VIs can help monitor plant health and yield potential, which can be used to guide effective crop management practices, such as fertilization, irrigation, and pest management.

In summary, a combination of either blue or red, RE, and NIR-based vegetation indices had a strong correlation with corn yield. Our results showed that certain vegetation indices demonstrated high

TABLE 2 A simple linear regression model statistic developed with the top-ranked vegetation index selected by a random forest based variable selection method for three years of study.

2020				2021				2022			
Growth stage	VI	R^2	MAPE	Growth stage	VI	R^2	MAPE	Growth stage	VI	R^2	MAPE
V5	VARI	0.25**	4.25	V5	MTCI	0.58***	6.85	V6	RIgreen	0.47***	9.2
V13	VARI	0.39***	3.63	V7	RVI	0.70***	6.19	V10	TCARI/ OSAVI	0.71***	7.19
Vn	MSAVI	0.53***	3.21	V11	DATT	0.69***	5.83	Vn	SR445	0.80***	5.74
R1	Soil cover	0.56***	2.82	Vn	CVI	0.82***	4.04	R1	CVI	0.76***	6
R2	TGI	0.43***	3.61	R1	CVI	0.79***	4.54	R3	SAVI	0.67***	7.69
R4	LCI	0.56***	3.18	R2	LCI	0.76***	4.81	R5	SR445	0.89***	4.21
R5	MTCI	0.54***	3.01	R5	TGI	0.83***	4.59				

V5, V6, V7, V10, V11, and V13, represent the vegetative growth stage with 5, 6, 7, 10, 11 and 13 visible leaf collars on the main stem. Vn represent number of leaf collars greater than 13. R1, R2, R3, R4, and R5, which correspond to the silk, blister, milk, dough, and dent stages, respectively. *, ** and *** indicate the regression is significant at $p<0.05$, $p<0.01$ and $p<0.001$ respectively. 'ns' indicates non-significant. VI- vegetation index; MAPE: mean absolute percentage error. Full form of acronyms are given in Figure 4 or Supplementary Table 2.

consistency in predicting corn yield. Specifically, the indices LCI, MTCI, mND705, mSR705, and RIrededge showed the strongest predictive capabilities. Among them, MTCI emerged as the most promising VI and can be effectively used to predict corn yield during the reproductive stage. Further, weekly cloud-free imagery can be used for real-time monitoring of yield estimation under different cropping systems, which can support both research and farm decision-making as UASs become increasingly ubiquitous.

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

RB, SS, and JC conceptualized the research; RM provided the funding. AS, CM, and AA collected the data. AS and RB performed the image analysis, data analysis, data visualization, and composed the first draft. All authors contributed to editing and revision. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the USDA-Agricultural Research Service (USDA-ARS) (58-6064-9-007) and the National Institute of Food and Agriculture (MIS 043050).

Acknowledgments

We thank the Plant Stress Physiology Laboratory members for their support during data collection. We would also like to thank the Geosystems Research Institute remote sensing crew for assisting with data collection and processing.

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.2023.1168732/full#supplementary-material

References

Aditya Shastry, K., and Sanjay, H. A. (2021). A modified genetic algorithm and weighted principal component analysis based feature selection and extraction strategy in agriculture. *Knowl. Based Syst.* 232, 107460. doi: 10.1016/j.knosys.2021.107460

Akoglu, H. (2018). User's guide to correlation coefficients. *Turk. J. Emerg. Med.* 18, 91–93. doi: 10.1016/j.tjem.2018.08.001

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

Asuero, A. G., Sayago, A., and González, A. G. (2006). The correlation coefficient: an overview. *Crit. Rev. Anal. Chem.* 36, 41–59. doi: 10.1080/10408340500526766

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

Barnes, M. L., Breshears, D. D., Law, D. J., van Leeuwen, W. J. D., Monson, R. K., Fojtik, A. C., et al. (2017). Beyond greenness: detecting temporal changes in photosynthetic capacity with hyperspectral reflectance data. *PloS One* 12, e0189539. doi: 10.1371/journal.pone.0189539

Barzin, R., Lotfi, H., Varco, J. J., and Bora, G. C. (2022). Machine learning in evaluating multispectral active canopy sensor for prediction of corn leaf nitrogen concentration and yield. *Remote Sens.* 14, 120. doi: 10.3390/rs14010120

Barzin, R., Pathak, R., Lotfi, H., Varco, J., and Bora, G. C. (2020). Use of UAS multispectral imagery at different physiological stages for yield prediction and input resource optimization in corn. *Remote Sens.* 12, 2392. doi: 10.3390/rs12152392

Bheemanahalli, R., Krishnan, B. S., Wijewardane, N. K., Samiappan, S., and Reddy, K. R. (2023). "Remote sensing algorithms and their applications in plant phenotyping," in *Translating Physiological Tools to Augment Crop Breeding*. Eds. M.H. Masthigowda, K. Gopalareddy, R. Khobra, G. Singh and G.P. Singh (Singapore: Springer Nature), 337–353. doi: 10.1007/978-981-19-7498-4_15

Bheemanahalli, R., Ramamoorthy, P., Poudel, S., Samiappan, S., Wijewardane, N., and Reddy, K. R. (2022). Effects of drought and heat stresses during reproductive stage on pollen germination, yield, and leaf reflectance properties in maize (Zea mays L.). *Plant Direct* 6, e434. doi: 10.1002/pld3.434

Boyd, D. S., Almond, S., Dash, J., Curran, P. J., and Hill, R. A. (2011). Phenology of vegetation in Southern England from Envisat MERIS terrestrial chlorophyll index (MTCI) data. *Int. J. Remote Sens.* 32, 8421–8447. doi: 10.1080/01431161.2010.542194

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Bresler, E., Dasberg, S., Russo, D., and Dagan, G. (1981). Spatial variability of crop yield as a stochastic soil process. *Soil Sci. Soc Am. J.* 45, 600–605. doi: 10.2136/sssaj1981.03615995004500030032x

Brewer, K., Clulow, A., Sibanda, M., Gokool, S., Naiken, V., and Mabhaudhi, T. (2022). Predicting the chlorophyll content of maize over phenotyping as a proxy for crop health in smallholder farming systems. *Remote Sens.* 14, 518. doi: 10.3390/rs14030518

Carlson, T. N., and Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens. Environ.* 62, 241–252. doi: 10.1016/S0034-4257(97)00104-1

Chivasa, W., Mutanga, O., and Burgueño, 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

Croft, H., Chen, J. M., Wang, R., Mo, G., Luo, S., Luo, X., et al. (2020). The global distribution of leaf chlorophyll content. *Remote Sens. Environ.* 236, 111479. doi: 10.1016/j.rse.2019.111479

Crusiol, L. G. T., Nanni, M. R., Furlanetto, R. H., Sibaldelli, R. N. R., Cezar, E., Mertz-Henning, L. M., et al. (2020). UAV-based thermal imaging in the assessment of water status of soybean plants. *Int. J. Remote Sens.* 41, 3243–3265. doi: 10.1080/01431161.2019.1673914

Dash, J., and Curran, P. J. (2004). The MERIS terrestrial chlorophyll index. *Int. J. Remote Sens.* 25, 5403–5413. doi: 10.1080/0143116042000274015

Dash, J., and Curran, P. J. (2007). Evaluation of the MERIS terrestrial chlorophyll index (MTCI). *Adv. Space Res.* 39, 100–104. doi: 10.1016/j.asr.2006.02.034

Dash, J., Curran, P. J., Tallis, M. J., Llewellyn, G. M., Taylor, G., and Snoeij, P. (2010). Validating the MERIS Terrestrial Chlorophyll Index (MTCI) with ground chlorophyll content data at MERIS spatial resolution. *Int. J. Remote Sens.* 31, 5513–5532. doi: 10.1080/01431160903376340

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

Dhillon, J., Li, X., Bheemanahalli, R., and Reed, V. (2022). Mississippi state and county level yield gap in corn production. *Agric. Environ. Lett.* 7, e20092. doi: 10.1002/ael2.20092

Dong, T., Meng, J., Shang, J., Liu, J., and Wu, B. (2015). Evaluation of chlorophyll-related vegetation indices using simulated Sentinel-2 data for estimation of crop fraction of absorbed photosynthetically active radiation. *IEEE J. Sel. Top. Appl. Earth. Obs. Remote Sens.* 8, 4049–4059. doi: 10.1109/JSTARS.2015.2400134

Fahrentrapp, J., Ria, F., Geilhausen, M., and Panassiti, B. (2019). Detection of Gray mold leaf infections prior to visual symptom appearance using a five-band multispectral sensor. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00628

FAO (2022) *FAOSTAT, production*. Available at: https://www.fao.org/faostat/en/#data/QCL (Accessed January 11, 2022).

Geipel, J., Link, J., and Claupein, W. (2014). Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system. *Remote Sens.* 6, 10335–10355. doi: 10.3390/rs61110335

Gitelson, A. A., Stark, R., Grits, U., Rundquist, D., Kaufman, Y., and Derry, D. (2002). Vegetation and soil lines in visible spectral space: a concept and technique for remote estimation of vegetation fraction. *Int. J. Remote Sens.* 23, 2537–2562. doi: 10.1080/01431160110107806

Gonzalez-Dugo, V., Zarco-Tejada, P. J., and Fereres, E. (2014). Applicability and limitations of using the crop water stress index as an indicator of water deficits in citrus orchards. *Agric. For. Meteorol.* 198–199, 94–104. doi: 10.1016/j.agrformet.2014.08.003

Halliday, S. L., and Wolfe, M. L. (1991). Assessing ground water pollution potential from nitrogen fertilizer using a geographic information system. *J. Am. Water Resour. Assoc.* 27, 237–245. doi: 10.1111/j.1752-1688.1991.tb03128.x

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

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

IFA (2022). *Fertilizer use by crop and country for the 2017-2018 period* (Paris, France: International Fertilizer Association (IFA). Available at: https://www.ifastat.org/consumption/fertilizer-use-by-crop.

Jaynes, D. B., Colvin, T. S., and Kaspar, T. C. (2005). Identifying potential soybean management zones from multi-year yield data. *Comput. Electron. Agric.* 46, 309–327. doi: 10.1016/j.compag.2004.11.011

Kitchen, N. R., Sudduth, K. A., Drummond, S. T., Scharf, P. C., Palm, H. L., Roberts, D. F., et al. (2010). Ground-based canopy reflectance sensing for variable-rate nitrogen corn fertilization. *Agron. J.* 102, 71–84. doi: 10.2134/agronj2009.0114

Kitchen, N. R., Sudduth, K. A., Myers, D. B., Drummond, S. T., and Hong, S. Y. (2005). Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. *Comput. Electron. Agric.* 46, 285–308. doi: 10.1016/j.compag.2004.11.012

Kogan, F., Guo, W., Yang, W., and Harlan, S. (2018). Space-based vegetation health for wheat yield modeling and prediction in Australia. *J. Appl. Remote Sens.* 12, 26002. doi: 10.1117/1.JRS.12.026002

Kross, A., McNairn, H., Lapen, D., Sunohara, M., and Champagne, C. (2015). Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *Int. J. Appl. Earth Obs. Geoinf.* 34, 235–248. doi: 10.1016/j.jag.2014.08.002

Lark, R. M. (1998). Forming spatially coherent regions by classification of multivariate data: an example from the analysis of maps of crop yield. *Int. J. Geogr. Inf. Sci.* 12, 83–98. doi: 10.1080/136588198242021

Li, F., Li, D., Elsayed, S., Hu, Y., and Schmidhalter, U. (2021). Using optimized three-band spectral indices to assess canopy N uptake in corn and wheat. *Eur. J. Agron.* 127, 126286. doi: 10.1016/j.eja.2021.126286

Li, F., Miao, Y., Feng, G., Yuan, F., Yue, S., Gao, X., et al. (2014). Improving estimation of summer maize nitrogen status with red edge-based spectral vegetation indices. *Field Crops Res.* 157, 111–123. doi: 10.1016/j.fcr.2013.12.018

Li, Y., Miao, Y., Zhang, J., Cammarano, D., Li, S., Liu, X., et al. (2022). Improving estimation of winter wheat nitrogen status using random forest by integrating multi-source data across different agro-ecological zones. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.890892

Lichtenthaler, H. (1987). Chlorophylls and carotenoids: pigments of photosynthetic biomembranes. *Meth. Enzymol.* 148C, 350–382. doi: 10.1016/0076-6879(87)48036-1

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. doi: 10.3389/fpls.2022.948249

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

Ma, S., Zhou, Y., Gowda, P. H., Dong, J., Zhang, G., Kakani, V. G., et al. (2019). Application of the water-related spectral reflectance indices: a review. *Ecol. Indic.* 98, 68–79. doi: 10.1016/j.ecolind.2018.10.049

Maestrini, B., and Basso, B. (2018). Predicting spatial patterns of within-field crop yield variability. *Field Crops Res.* 219, 106–112. doi: 10.1016/j.fcr.2018.01.028

Maleki, M., Arriga, N., Barrios, J. M., Wieneke, S., Liu, Q., Peñuelas, J., et al. (2020). Estimation of gross primary productivity (GPP) phenology of a short-rotation plantation using remotely sensed indices derived from Sentinel-2 images. *Remote Sens.* 12, 2104. doi: 10.3390/rs12132104

Maresma, Á., Ariza, M., Martínez, E., Lloveras, J., and Martínez-Casasnovas, J. A. (2016). Analysis of vegetation indices to determine nitrogen application and yield prediction in maize (Zea mays L.) from a standard UAV service. *Remote Sens.* 8, 973. doi: 10.3390/rs8120973

Maresma, A., Chamberlain, L., Tagarakis, A., Kharel, T., Godwin, G., Czymmek, K. J., et al. (2020). Accuracy of NDVI-derived corn yield predictions is impacted by time of sensing. *Comput. Electron. Agric.* 169, 105236. doi: 10.1016/j.compag.2020.105236

Matese, A., Toscano, P., Di Gennaro, S. F., Genesio, L., Vaccari, F. P., Primicerio, J., et al. (2015). Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture. *Remote Sens.* 7, 2971–2990. doi: 10.3390/rs70302971

Maya Gopal, P. S., and Bhargavi, R. (2019). Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms. *Appl. Artif. Intell.* 33, 621–642. doi: 10.1080/08839514.2019.1592343

Myneni, R. B., and Asrar, G. (1994). Atmospheric effects and spectral vegetation indices. *Remote Sens. Environ.* 47, 390–402. doi: 10.1016/0034-4257(94)90106-6

Nleya, T., Chungu, C., and Kleinjan, J. (2016). *"Chapter 5: corn growth and development,"* in iGrow Corn: Best Management Practices, eds. D. E. Clay, G. G. Carlson, and E. Byamukama (South Dakota State University).

Or, D., and Hanks, R. J. (1992). Soil water and crop yield spatial variability induced by irrigation nonuniformity. *Soil Sci. Soc Am. J.* 56, 226–233. doi: 10.2136/sssaj1992.03615995005600010035x

Santana, D. C., Cotrim, M. F., Flores, M. S., Rojo Baio, F. H., Shiratsuchi, L. S., Silva Junior, C. A., et al. (2021). UAV-based multispectral sensor to measure variations in corn as a function of nitrogen topdressing. *Remote Sens. Appl.: Soc Environ.* 23, 100534. doi: 10.1016/j.rsase.2021.100534

Santesteban, L. G., Di Gennaro, S. F., Herrero-Langreo, A., Miranda, C., Royo, J. B., 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

Saravi, B., Nejadhashemi, A. P., Jha, P., and Tang, B. (2021). Reducing deep learning network structure through variable reduction methods in crop modeling. *Artif. Intell. Agric.* 5, 196–207. doi: 10.1016/j.aiia.2021.09.001

Scharf, P. C., Shannon, D. K., Palm, H. L., Sudduth, K. A., Drummond, S. T., Kitchen, N. R., et al. (2011). Sensor-based nitrogen applications out-performed producer-chosen rates for corn in on-farm demonstrations. *Agron. J.* 103, 1683–1691. doi: 10.2134/agronj2011.0164

Schepers, J. S., Francis, D. D., Vigil, M., and Below, F. E. (1992). Comparison of corn leaf nitrogen concentration and chlorophyll meter readings. *Commun. Soil Sci. Plant Anal.* 23, 2173–2187. doi: 10.1080/00103629209368733

Sellers, P. J. (1985). Canopy reflectance, photosynthesis and transpiration. *Int. J. Remote Sens.* 6, 1335–1372. doi: 10.1080/01431168508948283

Servadio, P., Bergonzoli, S., and Verotti, M. (2017). Delineation of management zones based on soil mechanical-chemical properties to apply variable rates of inputs throughout a field (VRA). *Eng. Agric. Environ. Food* 10, 20–30. doi: 10.1016/j.eaef.2016.07.001

Shafian, S., Rajan, N., Schnell, R., Bagavathiannan, M., Valasek, J., Shi, Y., et al. (2018). Unmanned aerial systems-based remote sensing for monitoring sorghum growth and development. *PloS One* 13, e0196605. doi: 10.1371/journal.pone.0196605

Simic Milas, A., Romanko, M., Reil, P., Abeysinghe, T., and Marambe, A. (2018). The importance of leaf area index in mapping chlorophyll content of corn under different agricultural treatments using UAV images. *Int. J. Remote Sens.* 39, 5415–5431. doi: 10.1080/01431161.2018.1455244

Sims, D. A., and Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.* 81, 337–354. doi: 10.1016/S0034-4257(02)00010-X

Singh, B. (2018). Are nitrogen fertilizers deleterious to soil health? *Agron.* 8, 48. doi: 10.3390/agronomy8040048

Tan, C.-W., Wang, D.-L., Zhou, J., Du, Y., Luo, M., Zhang, Y.-J., et al. (2018). Assessment of Fv/Fm absorbed by wheat canopies employing in-situ hyperspectral vegetation indexes. *Sci. Rep.* 8, 9525. doi: 10.1038/s41598-018-27902-3

ten Harkel, J., Bartholomeus, H., and Kooistra, L. (2020). Biomass and crop height estimation of different crops using UAV-based lidar. *Remote Sens.* 12, 17. doi: 10.3390/rs12010017

USDA (2022) *USDA National agricultural statistics service*. Available at: https://quickstats.nass.usda.gov/results/EC30F678-F2B4-3CF2-BA25-497916ADC489?pivot=short_desc (Accessed January 10, 2023).

Ustin, S. L., and Jacquemoud, S. (2020). "How the optical properties of leaves modify the absorption and scattering of energy and enhance leaf functionality," in *Remote Sensing of Plant Biodiversity*. Eds. J. Cavender-Bares, J. A. Gamon and P. A. Townsend (Cham: Springer International Publishing), 349–384. doi: 10.1007/978-3-030-33157-3_14

Vong, C. N., Conway, L. S., Zhou, J., Kitchen, N. R., and Sudduth, K. A. (2021). Early corn stand count of different cropping systems using UAV-imagery and deep learning. *Comput. Electron. Agric.* 186, 106214. doi: 10.1016/j.compag.2021.106214

Wang, R., Chen, J. M., Liu, Z., and Arain, A. (2017). Evaluation of seasonal variations of remotely sensed leaf area index over five evergreen coniferous forests. *ISPRS J. Photogramm. Remote Sens.* 130, 187–201. doi: 10.1016/j.isprsjprs.2017.05.017

Wood, C. W., Reeves, D. W., Duffield, R. R., and Edmisten, K. L. (1992). Field chlorophyll measurements for evaluation of corn nitrogen status. *J. Plant Nutr.* 15, 487–500. doi: 10.1080/01904169209364335

Wu, X., Shen, X., Cao, L., Wang, G., and Cao, F. (2019). Assessment of individual tree detection and canopy cover estimation using unmanned aerial vehicle based light detection and ranging (UAV-LiDAR) data in planted forests. *Remote Sens.* 11, 908. doi: 10.3390/rs11080908

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

Xue, J., and Su, B. (2017). Significant remote sensing vegetation indices: a review of developments and applications. *J. Sens.* 2017, e1353691. doi: 10.1155/2017/1353691

Yang, Y., Timlin, D. J., Fleisher, D. H., Lokhande, S. B., Chun, J. A., Kim, S.-H., et al. (2012). Nitrogen concentration and dry-matter accumulation in maize crop: assessing maize nitrogen status with an allometric function and a chlorophyll meter. *Commun. Soil Sci. Plant Anal.* 43, 1563–1575. doi: 10.1080/00103624.2012.675393

Yao, H., Qin, R., and Chen, X. (2019). Unmanned aerial vehicle for remote sensing applications–a review. *Remote Sens.* 11, 1443. doi: 10.3390/rs11121443

Yuan, W., Li, J., Bhatta, M., Shi, Y., Baenziger, P. S., and Ge, Y. (2018). Wheat height estimation using LiDAR in comparison to ultrasonic sensor and UAS. *Sensors* 18, 3731. doi: 10.3390/s18113731

Zahir, S. A. D. M., 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. Actuators A: Phys.* 338, 113468. doi: 10.1016/j.sna.2022.113468

Zarate-Valdez, J. L., Whiting, M. L., Lampinen, B. D., Metcalf, S., Ustin, S. L., and Brown, P. H. (2012). Prediction of leaf area index in almonds by vegetation indexes. *Comput. Electron. Agric.* 85, 24–32. doi: 10.1016/j.compag.2012.03.009

Zhang, X., Han, L., Dong, Y., Shi, Y., Huang, W., Han, L., et al. (2019). A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral UAV images. *Remote Sens.* 11, 1554. doi: 10.3390/rs11131554

Zhang, S., and Liu, L. (2014). The potential of the MERIS Terrestrial Chlorophyll Index for crop yield prediction. *Remote Sens. Lett.* 5, 733–742. doi: 10.1080/2150704X.2014.963734

Zhao, B., Duan, A., Ata-Ul-Karim, S. T., Liu, Z., Chen, Z., Gong, Z., et al. (2018). Exploring new spectral bands and vegetation indices for estimating nitrogen nutrition index of summer maize. *Eur. J. Agron.* 93, 113–125. doi: 10.1016/j.eja.2017.12.006

Zhao, D., Reddy, K. R., Kakani, V. G., Read, J. J., and Koti, S. (2007). Canopy reflectance in cotton for growth assessment and lint yield prediction. *Eur. J. Agron.* 26, 335–344. doi: 10.1016/j.eja.2006.12.001

Ziadi, N., Brassard, M., Bélanger, G., Claessens, A., Tremblay, N., Cambouris, A. N., et al. (2008). Chlorophyll measurements and nitrogen nutrition index for the evaluation of corn nitrogen status. *Agron. J.* 100, 1264–1273. doi: 10.2134/agronj2008.0016

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 26 July 2023
DOI 10.3389/fpls.2023.1214006

Check for updates

OPEN ACCESS

EDITED BY
Huiling Chen,
Wenzhou University, China

REVIEWED BY
Qingquan Chu,
China Agricultural University, China
Yanbo Huang,
United States Department of Agriculture (USDA), United States

*CORRESPONDENCE
Chunsong Bian
bianchunsong@caas.cn
Jiangang Liu
liujiangang@caas.cn

RECEIVED 28 April 2023
ACCEPTED 06 July 2023
PUBLISHED 26 July 2023

CITATION
Lin Y, Li S, Duan S, Ye Y, Li B, Li G, Lyv D, Jin L, Bian C and Liu J (2023) Methodological evolution of potato yield prediction: a comprehensive review. *Front. Plant Sci.* 14:1214006. doi: 10.3389/fpls.2023.1214006

COPYRIGHT
© 2023 Lin, Li, Duan, Ye, Li, Li, Lyv, Jin, Bian and Liu. 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.

Methodological evolution of potato yield prediction: a comprehensive review

Yongxin Lin[1,2], Shuang Li[1], Shaoguang Duan[1], Yanran Ye[1], Bo Li[3], Guangcun Li[1], Dianqiu Lyv[2], Liping Jin[1], Chunsong Bian[1*] and Jiangang Liu[1*]

[1]State Key Laboratory of Vegetable Biobreeding, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China, [2]College of Agronomy and Biotechnology, Southwest University, Chongqing, China, [3]Seeds Development, Syngenta Jealott's Hill International Research Centre, Bracknell, United Kingdom

Timely and accurate prediction of crop yield is essential for increasing crop production, estimating planting insurance, and improving trade benefits. Potato (*Solanum tuberosum* L.) is a staple food in many parts of the world and improving its yield is necessary to ensure food security and promote related industries. We conducted a comprehensive literature survey to demonstrate methodological evolution of predicting potato yield. Publications on predicting potato yield based on methods of remote sensing (RS), crop growth model (CGM), and yield limiting factor (LF) were reviewed. RS, especially satellite-based RS, is crucial in potato yield prediction and decision support over large farm areas. In contrast, CGM are often utilized to optimize management measures and address climate change. Currently, combined with the advantages of low cost and easy operation, unmanned aerial vehicle (UAV) RS combined with artificial intelligence (AI) show superior potential for predicting potato yield in precision management of large-scale farms. However, studies on potato yield prediction are still limited in the number of varieties and field sample size. In the future, it is critical to employ time-series data from multiple sources for a wider range of varieties and large field sample sizes. This study aims to provide a comprehensive review of the progress in potato yield prediction studies and to provide a theoretical reference for related research on potato.

KEYWORDS

yield prediction, potato, precision agriculture, remote sensing, crop growth model

1 Introduction

Global food security is encountering significant challenges from climate change and increasing resource competition (Godfray et al., 2010). Potato (*Solanum tuberosum* L.), a tuberous crop, is cultivated worldwide due to its stable and high yield, wide adaptability, and complete nutritional composition. Furthermore, it is a pivotal crop for realizing the United Nations (UN) Sustainable Development Goals (SDGs). Yield of potato and other

crops is determined interactively by genotype (G), environment (E), and management practices (M) (Cooper et al., 2021). Analysis and modeling of key parameters can effectively predict crop yield, providing crucial guidance and decision support for various stakeholders, such as farmers, policy makers, and agribusinesses. In addition, these predictions substantially impact optimizing planting structure, optimizing trading policies, allocating resources efficiently, and conducting precision management.

There have been significant strides in both theoretical and practical aspects of predicting potato yield. Early on, such prediction relied on field sampling, whereby the number and weight of potatoes per unit area were measured to calculate the yield (Dyke and Avis, 1953). Other agronomic traits in subsequent studies, including petiole potassium content (Holm and Nylund, 1978), also served as useful indicators for potato yield. Nevertheless, these destructive methods require substantial labor for field sampling and do not provide complete spatial or temporal coverage.

Remote sensing (RS) has emerged as a popular tool in crop phenotyping (Araus and Cairns, 2014), growth monitoring (Liu et al., 2021), and yield prediction (Ma et al., 2021), attributed to it being non-destructive, high-throughput, and having large spatial coverage. In 1974, the Large Area Crop Inventory Experiment (LACIE) program, which showed the possibility of RS for yield prediction for the first time, was used to assess wheat acreage in the United States, Canada, and the former Soviet Union combined with Landsat (MacDonald et al., 1975). In 1977, the LACIE accurately predicted a declining trend in spring wheat production in the Soviet Union with precision of 90% leading to a positive impact on the United States economy (Hill et al., 1980). From 1980 to 1986, multiple departments involved in the LACIE program collaborated in the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing (AgRISTARS) initiative (Doraiswamy et al., 1979). This effort aimed to predict yield for eight crops (not including potato) within the United States and other countries worldwide. Satellite-based potato yield prediction commenced later in the 1980s. Potato acreage was estimated to predict production in Canada by Landsat (Ryerson et al., 1985). In 1987, the European Union proposed the Monitoring Agricultural Resources (MARS) project, which utilized satellite and aerial imagery to continuously monitor the planting area and growth status of several staple crops including potatoes (Van der Velde and Nisini, 2019). It also provided timely prediction results of crop yields for the European Union.

Although satellites have the advantage of covering large areas of farmland, they are greatly affected by the revisit interval and low resolution. To address the growing demand for site-specific crop monitoring and yield prediction among farmers, research into proximal RS technologies such as unmanned aerial vehicle (UAV) and ground-based RS has increased rapidly. In contrast to satellites, UAVs possess tremendous potential for site-specific phenotype acquisition, yield prediction, and precision management, which is attributed to their low cost, convenience, and high spatial resolution (Yang G. et al., 2017). Additionally, ground-based methods are increasingly being utilized for more detailed phenotypic analyses in a variety of specific scenarios. In general, RS is capable of rapidly monitoring fields without damaging them. However, it employs empirical modeling methods most of which lack a robust mechanism.

Crop growth model (CGM) aim to describe the process of potato development before harvest. POTATO (Ng and Loomis, 1984) is the first CGM for potato with a complete mechanism. Between the 1990s and early 2000s, potato CGM became more comprehensive with the incorporation of additional parameters, such as water and nitrogen modules (Tang et al., 2021). During this period, various models were developed, including DSSAT-SUBSTOR (Ritchie et al., 1995) and LINTUL-NPOTATO (Van Delden et al., 2003). Although these models are mechanistic and have high precision, they require a substantial number of input parameters. Furthermore, the calibration and validation of these models rely on ground-truth data, which can be laborious to acquire.

During the 2010s, sensor technology, machine learning (ML), digital image analysis, and data mining techniques developed rapidly. Additionally, mechanisms underlying potato growth and development, yield quality formation principles, and interactions between crop-environment-management measures were better understood, encouraging improvement in theories of potato yield prediction.

Yield prediction requires multidisciplinary knowledge at the intersection of agronomy, meteorology, statistics, economics, and computer science. Several studies have reviewed the advancements made in yield prediction for diverse crops such as rice (dela Torre et al., 2021) and maize (Tandzi and Mutengwa, 2020). In contrast, yield prediction for potato differs from other major crops because its edible part is located belowground. Currently, there is no comprehensive literature survey of potato yield prediction due to complex model types and application scenarios. To tackle these issues, this paper provides an overview of the advancements in and prospects for potato yield prediction. First, we present a summary of the commonly used methodologies and compare them. Second, after thoroughly evaluating the existing methods, we envision the future development of potato yield prediction. This review comprehensively evaluates the progress made in potato yield prediction and provides the corresponding theoretical references.

2 Literature survey

A total of 276 articles including the keywords "potato", "yield or production or output", and "estimat* or forecast* or predict* or simulat*" were identified in the Web of Science™ database (Clarivate Analytics) through January 11, 2023. To encompass a broader scope of relevant studies, we also conducted a literature survey with the abovementioned keywords on the Scopus database and retrieved 152 publications. After eliminating duplicates and irrelevant studies, 160 publications were included in this study. As depicted in Figure 1A, the number of pertinent studies has progressively increased since the 2010s. Furthermore, Figure 1B demonstrates a growing number of annual citations for these publications, indicating increased interest in this research domain.

Through literature review, we classified studies on potato yield prediction into three categories including methods based on RS,

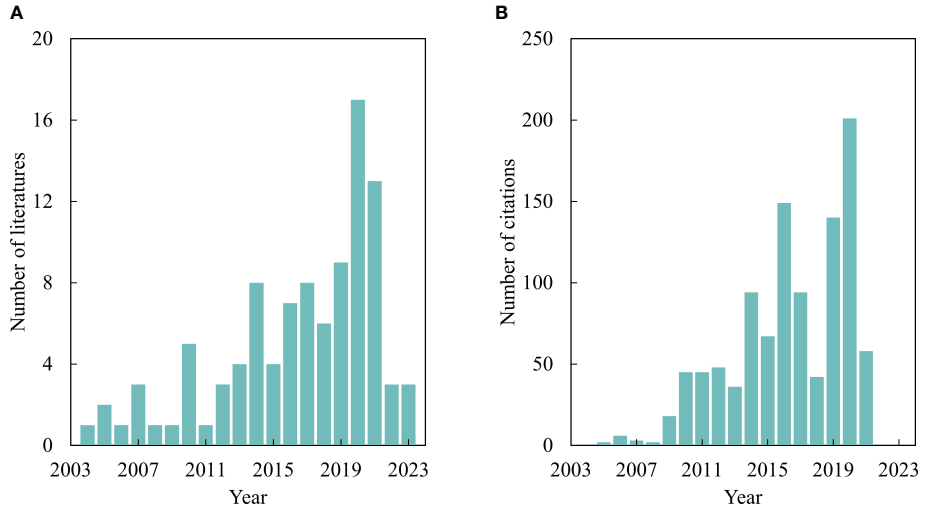


FIGURE 1
Literature counts **(A)** and citations **(B)** for potato yield production since 2003 to 2023.

CGM, and yield limiting factor (LF). Methods with LF include those based on agronomic and environmental parameters. Figure 2 displays the numbers of the three approaches in potato yield prediction studies over the past 50 years. Initially, RS and CGM were less used for potato yield prediction. CGM-based methods have a long history with the key period of research and development occurring in the 1980s and 1990s. Since the 21st century, CGM have been widely applied, with research efforts focused on the parameterization of models under various conditions. The CO_2 response module is integrated with climate models to assess the impact of climate change on future yields. With the development of advanced information technology, RS-based methods have emerged in the study of crop yield prediction, utilizing next-generation sensors, UAVs, and ML algorithms. For instance, in 2020, there were 17 publications relating to potato yield prediction, of which ten were dedicated to RS-based methods. These publications contain RS-based yield prediction at multiple carrying platforms for sensors, ranging from satellites and aerial, to ground-based methods (Figure 3), achieving site-specific yield prediction across multiple spatial scales. Finally, to facilitate comprehension of readers, we have produced a nomenclature (Table 1).

3 Remote sensing for potato yield prediction

Agricultural RS was primarily applied in the resource survey, and in crop growth monitoring, yield prediction, disaster

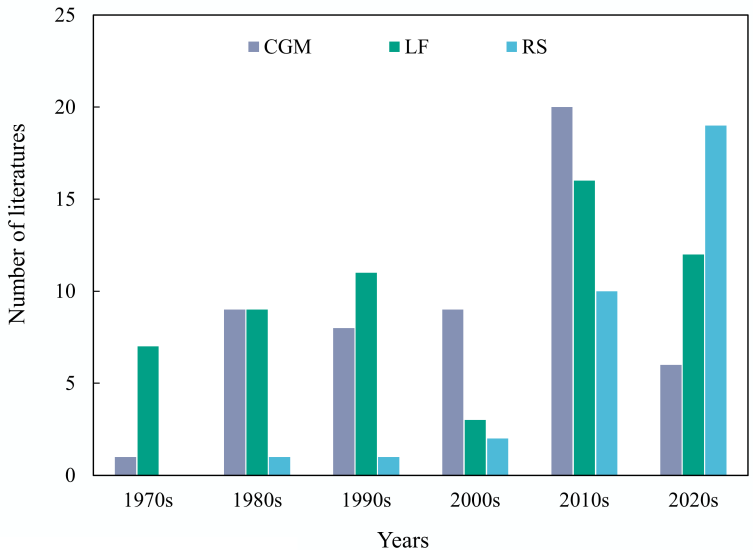


FIGURE 2
Distribution of strategies for potato yield prediction using CGM, RS, and LF-based methods since 1970s to 2020s.

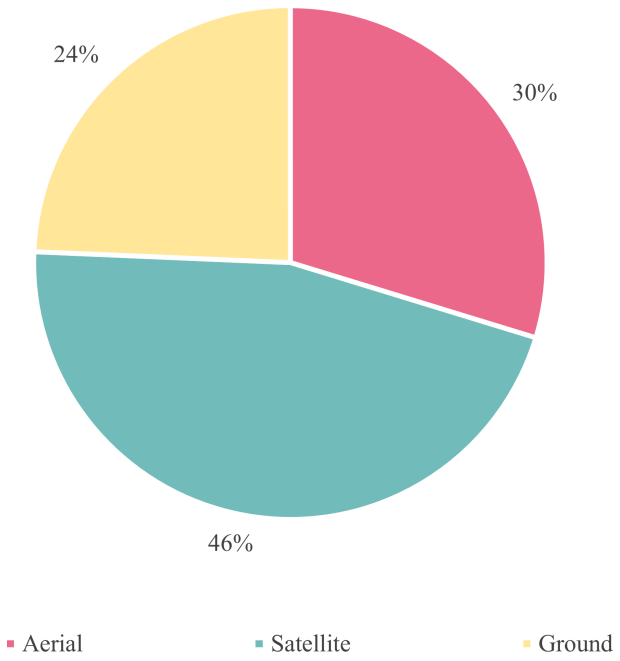


FIGURE 3
Distribution of RS Platforms adopted for potato yield prediction.

estimation, and loss assessment (Weiss et al., 2020). The LACIE program in the 1970s was representative of RS-based yield prediction in other crops; related applications in potatoes have been relatively delayed. In the 1980s, researchers employed satellite imagery to estimate potato acreage and combined it with yield data to estimate production (Ryerson et al., 1985). Nonetheless, during this period, yield data were still obtained through interviewing farmers, rather than direct evaluation of RS imagery. In 1992, potato yield was estimated by combining process-based crop models with leaf area index (LAI) data collected by handheld multispectral sensors (Finke, 1992). However, the spatial coverage of the handheld instruments was incomplete, which made it challenging to capture the yield variability of the entire field. In the early 21st century, several studies were conducted using non-destructive and convenient UAVs and satellites, which provided more comprehensive spatial coverage, for potato yield estimation (Yokobori et al., 2004; Bala and Islam, 2009). Currently, the technology for RS-based potato yield prediction has significantly advanced with the emergence of new-generation platforms, sensors, and advanced algorithms.

In this section RS-based yield prediction methods were reviewed from three perspectives: the acquisition of RS information, the selection of modeling parameters and the evolution of yield prediction models.

3.1 Acquisition of RS information

The RS system is comprised of a platform and integrated sensors. Different types of RS platforms offer unique benefits for specific application scenarios. According to the type of platform, we divided the potato yield prediction method based on RS into satellite-based, aerial-based, and ground-based methods for evaluation.

3.1.1 Satellite-based RS

Primarily, satellites equipped with spectral sensors can obtain ground vegetation spectral information over large areas for the purposes of land resource surveying, crop growth monitoring, and yield prediction (Nakalembe et al., 2021). Since 1972, satellites such as the Landsat-1, which is equipped with a Multispectral Scanner (MSS) containing four spectral bands, have been successfully launched. As a result, humanity began monitoring global resources and environmental factors on a large scale. Landsat imagery was used to estimate potato acreage by Statistics Canada in New Brunswick from 1980 to1982. They found that potato acreage could be estimated accurately using Landsat images, with a coefficient of variation of around 5.5% (Ryerson et al., 1985). A series of weather observation satellites, such as NOAA-6 with the Advanced Very-High-Resolution Radiometer (AVHRR), has been operated by the National Oceanic and Atmospheric Administration (NOAA) since 1979. These satellites have provided an ample amount of RS imagery for accurate potato yield prediction (Akhand et al., 2016).

Even the revisit time of NOAA satellites equipped with AVHRR sensors is 12 hours, the spatial resolution is only 1.1 km. In 1999, the Earth Observing System (EOS) program launched Terra, which carries five specially designed sensors for monitoring environmental and climate change. Terra carries a Moderate Resolution Imaging Spectroradiometer (MODIS) capable of receiving spectral information of 36 bands between 0.4 and 14.4 μm with a spatial resolution of 250-1000 m. The potato yield was estimated using

TABLE 1 Nomenclature: abbreviations and corresponding full names.

Abbreviations	Full Name	Abbreviations	Full Name
ABA	Abscisic Acid	MCARI	Modified Chlorophyll Absorption Ratio Index
AEZ	Agro-Ecological Zone	MCYFS	Mars Crop Yield Forecasting System
AgRISTARS	Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing	ML	Machine Learning
AI	Artificial Intelligence	MLR	Multiple Linear Regression
ANFIS	Adaptive Neuro-Fuzzy Inference System	MME	Multi-Model Ensembles
ANN	Artificial Neural Network	MODIS	Moderate Resolution Imaging Spectroradiometer
APE	Agro-Pastoral Ecotone	MRE	Mean Relative Error
APSIM	Agricultural Production System Simulator Next Generation	MS	Multispectral
AVHRR	Advanced Very-High-Resolution Radiometer	MSE	Mean Squared Error
CC	Canopy Cover	MSS	Multispectral Scanner
CGM	Crop Growth Model	MTY	Marketable Tuber Yield
CGMS	Crop Growth Monitoring System	NDVI	Normalized Difference Vegetation Index
CI	Chlorophyll Index	NOAA	National Oceanic and Atmospheric Administration
CI1	Red-Edge Chlorophyll Index 1	NRCT	Normalized Relative Canopy Temperature
CI2	Red-Edge Chlorophyll Index 2	OLI	Thematic Mapper Plus
CIP	International Potato Center	PAR	Photosynthetically Active Radiation
CT	Computed Tomography	PPI	Potato Productivity Index
CV	Computer Vision	PROSPECT	Leaf Optical Properties Spectra Model
DL	Deep Learning	RF	Random Forest
EnKF	Ensemble Kalman Filter	RS	Remote Sensing
EOS	Earth Observing System	RVI	Ratio Vegetation Index
ETM+	Enhanced Thematic Mapper Plus	SAIL	Scattering By Arbitrarily Inclined Leaves Model
EVI	Enhanced Vegetation Index	SAR	Synthetic Aperture Radar
FAO	Food And Agriculture Organization	SAVI	Soil-Adjusted Vegetation Index
FPAR	Fraction Of Absorbed Photosynthetically Active Radiation	SDGs	Sustainable Development Goals
GA	Genetic Algorithm	SLR	Simple Linear Regression
GF-1	Gaofen-1	SVM	Support Vector Machine
GLUE	Generalized Likelihood Uncertainty Estimation	TCI	Temperature Condition Index
HI	Harvest Index	TIR	Thermal Infrared
HS	Hyperspectral	TM	Thematic Mapper
IIASA	International Institute for Applied Systems Analysis	UAV	Unmanned Aerial Vehicle
LACIE	Large Area Crop Inventory Experiment	UN	United Nations
LAI	Leaf Area Index	VCI	Vegetation Condition Index
LF	Limiting Factor	VI	Vegetation Index
LiDAR	Light Detection and Ranging	WDVI	Weighted Difference Vegetation Index
LSTM	Long Short-Term Memory Networks	WOFOST	World Food Studies

(Continued)

TABLE 1 Continued

Abbreviations	Full Name	Abbreviations	Full Name
LUE	Light Use Efficiency	WP	Water Productivity
MARS	Monitoring Agricultural Resources	4DVAR	Four-Dimensional Variational Data Assimilation

Normalized Difference Vegetation Index (NDVI), LAI, and Fraction of Absorbed Photosynthetically Active Radiation (FPAR) extracted from Terra-MODIS with an average error rate at 15% compared with actual yield by Bala and Islam (2009).

Despite significant enhancements in the resolution of MODIS imaging sensor compared with the NOAA satellites, the mixed image components comprising of different objects such as soil and potato canopy remain challenging to be discriminated. Low-resolution satellite imaging encounters the obstacle in estimating potato yields in relatively smaller regions. Landsat 4 and subsequent satellites equipped with Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), or Operational Land Imager (OLI), featuring a spatial resolution from 30 to 120 meters. The ongoing Landsat 8 and Landsat 9 observation missions each offer a revisit cycle of around 16 days. The joint utilization of both satellites has the potential to halve the revisit period to 8 days. Similarly, the Sentinel-2 of Copernicus Programme captures images of the same region every 10 days with a high spatial resolution ranging from 10 to 60 meters. Furthermore, the successful integration of Sentinel-2A and Sentinel-2B has the capability to reduce the revisit interval to a mere 5 days.

Satellites with greater spatial resolution and shorter revisit cycles showed better potential in predicting potato yield. Al-Gaadi et al. (2016) compared the accuracy of yield estimation for potato using Landsat-8 and Sentinel-2. Landsat-8 exhibited a range of prediction errors, ranging from 7.9% to 13.5%, along with R^2 values ranging from 0.39 to 0.65 at different sites. In contrast, Sentinel-2 demonstrated a lower prediction error, falling between 3.8% and 10.2%, though there was no significant improvement in R^2, which was between 0.47 and 0.65.

Ultra-high-resolution satellites with meter and sub-meter spatial resolution have emerged in recent years providing high-quality RS imaging data for monitoring crop development. RS employed in satellites has recently undergone substantial development, resulting in significant advancements in spatial, spectral, and temporal resolution. Nevertheless, the current cost of high-precision images acquired by commercial satellites remains high, and it is challenging for free satellite imagery at low spatial resolution to provide high-accuracy yield prediction. Moreover, apart from resolution and cost, weather conditions like cloud cover can also limit the data quality of obtained vegetation spectra. Site-specific potato dry matter yield was estimated using GeoEye-1 with an R^2 value of 0.60 (Elmetwalli et al., 2014). Due to the cost reduction in satellite launch, a number of commercial companies have launched small satellites that can be leveraged for Earth observation. PlanetScope, launched by Planet comprising 130 small satellites that can capture daily multispectral images at 3-meter resolution. PlanetScope images were utilized to develop potato yield prediction models in Idaho and applied them to assess yield differences between Norkotah and Russet varieties in Lebanon (Abou Ali et al., 2020). Table 2 displays several satellites and sensors that have been utilized for potato yield prediction in recent times. Meanwhile, we present some cases of potato yield prediction with satellites in Table 3.

For the past decades, RS employed in satellites has undergone substantial development, resulting in significant advancements in spatial, spectral and temporal resolution. Nevertheless, the current cost of high-precision images acquired by commercial satellites remains high, and it is challenging for the free satellite imagery at low spatial resolution to provide high-accuracy yield prediction. Moreover, apart from resolution and cost, weather conditions like cloud cover can also limit the data quality of obtained vegetation spectra.

3.1.2 Aerial-based RS

Aerial-based RS platforms include aerial vehicles at high altitudes and UAVs at low altitudes. A fixed-wing aerial plane Piper Seneca equipped with multispectral cameras was used to capture images of southern Idaho to estimate irrigated potato yield, which can be performed as a flexible and effective tool for yield prediction (Sivarajan, 2011). However, the cost for fuel and professional pilot is high. Recently, UAVs has become an important tool for RS-based yield prediction owing to its advantages of high resolution, high throughput, and low cost (Yang G. et al., 2017). Compared with satellites and manned aircraft, UAVs equipped with high-resolution sensors are able to acquire more detailed vegetation phenotypic information to predict yield. Most of the UAVs for field phenotyping fly at an altitude of below 150 m (Stöcker et al., 2017), and the image resolution can reach the centimeter level. There are several kinds of UAVs used in agriculture, such as multi-rotor UAVs, fixed-wing UAVs and unmanned helicopters. Multi-rotor UAVs are able to hover and turn flexibly in the air (Fu et al., 2020) but with high power usage, which lead to short battery life mostly within 30 minutes. In addition, multi-rotor UAV can carry limited number and types of sensors due to the small payload. Fixed-wing UAVs can fly at high speed with longer battery life, allowing them to cover a large area of farmland in a short period of time. In addition, fixed-wing UAVs with large wings typically have larger payloads which can offer a wider sensor option. However, it is impossible for fixed-wing UAV to capture data in small-scale farms because of the long runways required for takeoff and landing, and the inability to hover and turn flexibly in the air. Multi-rotor UAV is mostly used for potato yield forecasting, which is also for current mapping operations. Although we have not yet found the application of fixed-wing drones in potato yield prediction, they have great potential for large-scale potato field monitoring due to their high speed, long endurance, and large loads.

TABLE 2 Satellites that have previously been used to predict potato yields.

Platform	Sensors	Bands Number	Bands range*	Revisit interval (days)	Spatial Resolution (m/pixel)	Cost	Running state on orbit
NOAA	AVHRR	5-6	VIS, NIR, MIR, TIR	0.5	1100	Free	In progress
Landsat 5	TM	7	VIS, NIR, SWIR, TIR	16	30, 120	Free	Decommissioned in 2013
Landsat 7	ETM+	7	VIS, NIR, SWIR, TIR, PAN	16	15, 30, 60	Free	Decommissioned in 2021
Terra	MODIS	36	VIS, NIR, SWIR, MIR, TIR	1-2	250, 500, 1000	Free	In progress
IRS P6	LISS-3	4	VIS, NIR	24	23.5	Paid	In progress
GeoEye-1	–	4	VIS, NIR	3-5	0.41-1.65	Paid	In progress
Landsat 8, 9	OLI/TIRS	11	VIS, NIR, SWIR, TIR, PAN	16	15, 30, 100	Free	In progress
PlanetScope	PS2, PS2-SD	4-5	VIS, RE, NIR	1-2	3-4	Paid	In progress
Sentinel-2A/2B	MSI	13	VIS, RE, NIR, SWIR	10	10, 20, 60	Free	In progress

*Meaning of the abbreviations. MIR, Mid-Infrared; MW, Microwave; NIR, Near Infrared; PAN, Panchromatic; RE, Red edge; SWIR, Shortwave Infrared; TIR, Thermal Infrared; VIS, Visible Spectrum.

Compared to satellites that carry a fixed number and type of sensors, UAVs can readily change to appropriate sensors to meet specific needs. For example, it is feasible to extract information such as vegetation structure and reflectance from high-resolution RGB images for growth monitoring and biomass estimation. In contrast to digital RGB cameras that can function in the visible range, multispectral (MS) cameras obtain images at multiple spectral bands, including near infrared, which provides supplemented spectral information to estimate yield by calculating vegetation indexes (VIs). With the relatively low price of RGB and MS cameras, researchers often choose affordable small or medium-sized UAVs to conduct field trials. Most MS can only acquire a small amount of spectral information with low spectral resolution in the visible and near-infrared bands. In contrast, hyperspectral (HS) cameras provide higher spectral resolution with more continuous spectral information than MS. The above-mentioned spectral sensors have specific requirements for weather conditions when performing their tasks; in particular MS and HS need to acquire images in clear and cloud-free conditions.

Relative to passive sensors, active sensors can obtain highly accurate phenotypic information, such as plant height and biomass, independent of sunlight (ten Harkel et al., 2020). Light Detection and Ranging (LiDAR) and Synthetic Aperture Radar (SAR) are typical active sensors available on the market today. LiDAR obtains 3D and echo intensity information of vegetation to monitor crop growth based on the backward scattering characteristics of the light feature (Raj et al., 2020). SAR is a high-resolution active microwave imaging detection sensor that can penetrate clouds to obtain crop phenotypic information independent of atmospheric conditions and solar radiation (Lyalin et al., 2018). Active sensors such as LiDAR and SAR have not been applied in potato yield prediction. Due to the high cost of HS imaging sensor, reliable UAVs, such as the DJI Matrice 600 Pro (DJI Technology Co., Shenzhen, China) is preferred. Despite the advantages of UAVs for yield estimation at a large scale, there are still fewer studies on the application of UAV for potato yield prediction comparing to other staple crops. For reference, we summarize some previous studies of potato yield prediction combined with UAVs in Table 4.

3.1.3 Ground-based RS

Ground-based methods provide higher resolution and more angular image data for crop field phenotype. Many ground-based platforms have been developed, such as handheld or bracketed devices, ground carriers, tracks, ropeways, and fixed towers. They have their own unique advantages for different applications. Handheld or bracketed devices are simple and flexible in acquiring data. Tracks, ropeways, and fixed towers provide continuous observation of the same plot with high-precision sensors. However, they can only acquire data of specific plant samples. In contrast, ground carriers can perform data acquisition tasks over relatively larger areas.

According to our literature review, handheld or bracketed devices are still the most applied ground-based platforms for potato yield prediction. Different sensors have been used to predict potato yield by correlating various factors with yield. Zaeen et al. (2020) combined chlorophyll index (CI) and multiple VIs obtained by active sensors (Crop Circle™ and GreenSeeker™) to improve the performance of potato yield prediction. Their results indicated that the 18th and 20th leaf growth stages were the optimal period for data collection. No significant difference in accuracy between active spectral sensor and passive sensor was found for yield prediction in the early season by comparing handheld active sensors (Crop Circle™ and GreenSeeker™) with the portable MS Altum (MicaSense, Seattle, WA, USA) equipped on UAVs (Jasim et al., 2020). To the best of our knowledge, active sensors such as SAR and LiDAR have not been applied in potato yield prediction.

TABLE 3 Summary of satellite-based potato yield prediction.

Testing year	Country	Satellite	Sensor Type	Sensor Model	Cultivars	Parameters	Modelling Method	Accuracy	Sample size	References
2005/06-2006/07	Bangladesh	Terra	MS[1]	MODIS	–	Regional level: NDVI[2], LAI[3], FPAR[4], Field level: mean NDVI	SLR[23]	Regional level: R^2 (NDVI) = 0.79; R^2 (LAI) = 0.87; R^2 (FPAR) = 0.83; Field level: R^2 (mean NDVI) = 0.66	6 (Regional) 50 (Field)	Bala and Islam, 2009
2004-2005	United States	Landsat 5	MS	TM	–	ISAVI[5], cumulative ET[6]	MLR[24]	R^2 (2004) = 0.97; R^2 (2005) = 0.75	9 (2004) 13 (2005)	Sivarajan, 2011
2011	Libya	GeoEye-1	MS	–	–	NDVI	SLR	R^2 = 0.60	24	Elmetwalli et al., 2014
1980-2014	Bangladesh	NOAA	MS	AVHRR	–	VCI[7], TCI[8]	ANN[25]	error % < 10%	24	Akhand et al., 2016
2016	Saudi Arabia	Landsat 8, Sentinel-2A	MS	OLI/TIRS, MSI	–	NDVI, SAVI, cumulative NDVI, cumulative SAVI	SLR	R^2 (Landsat 8) = 0.39-0.65 R^2 (Sentinel-2A) = 0.47-0.65	60	Al-Gaadi et al., 2016
2010/11-2015/16	Bangladesh	Landsat 5, 7, 8	MS	TM, ETM+, OLI/TIRS	–	Mean NDVI	SLR	R^2 = 0.81	6	Newton et al., 2018
2016-2018	Spain	Sentinel-2A, 2B	MS	MSI	–	ARI2[9], CRI2[10], IRECI2[11], LCC[12], NDVI, PSRI[13], WDVI[14], S2 bands[15]	GLM[26], rqlasso[27], LeapBack[28], svmL[29], svmR[30], MARS[31], kknn[32], RF[33], avNNet[34]	The best three algrithoms: R^2 (rqlasso) = 0.90; R^2 (LeapBack) = 0.89; R^2 (svmR) = 0.93	33	Gómez et al., 2019
2012/13-2017/18	India	IRS P6	MS	LISS-3	–	VCI, climate data	Step wise regression	RMSE = 9.8-21.8%	–	Kumar et al., 2019
2017	United States, Lebanon	PlanetScope	MS	PS2, PS2-SD	Russet Burbank, Norkotah	SAVI	OLS35	R^2 (Russet Burbank) = 0.44; R^2 (Norkotah) = 0.57	–	Abou Ali et al., 2020
2004-2018	Mexico	Terra	MS	MODIS	–	NDVI, Harvested last year, climate, irrigation	RF, svmP, svmL, svmR, GLM	R^2 (RF) = 0.757-0.839; R^2 (svmR) = 0.733-0.837; R^2 (svmL) = 0.692-0.863; R^2 (svmP) = 0.717-0.858; R^2 (GLM)	838	Salvador et al., 2020

(Continued)

TABLE 3 Continued

Testing year	Country	Satellite	Sensor Type	Sensor Model	Cultivars	Parameters	Modelling Method	Accuracy	Sample size	References
								= 0.612-0.834		
2016-2019	Spain	Sentinel-2	MS	MSI	Monalisa, Spunta, Rudolf	NDVI, PPI[16], S2 bands	RF, SVM	S2 & PPI: R^2 (RF) = 0.77; R^2 (SVM) = 0.63 S2 & NDVI: R^2 (RF) = 0.66; R^2 (SVM) = 0.64	40	Gómez et al., 2021
2019-2020	United Kingdom	Sentinel-2	MS	MSI	Maris Piper, Amora, Pentland Dell	NDVI, SLAVI[17], NDMI[18], CIG[19]	SLR	R^2 = 0.65, NRMSE = 0.16	94	Mhango et al., 2021
2020	India	Sentinel-2	MS	MSI	–	NDVI	SLR	R^2 = 0.692	50	Singha and Swain, 2022
2016-2018	Belgium	Sentinel-2	MS	MSI	–	NDVI integral, Tmax[20], P[21], SDrz[22]	RF	R^2 (Late potato) = 0.57; R^2 (Early potato) = 0.68	723	Vannoppen and Gobin, 2022

[1] MS, Multispectral; [2] NDVI, Normalized Difference Vegetation Index; [3] LAI, Leaf Area Index; [4] FPAR, Fraction of Absorbed Photosynthetically Active Radiation; [5] ISAVI, three-date Integrated Soil Adjusted Vegetation Index; [6] ET, Evapotranspiration; [7] VCI, Vegetation Condition Index; [8] TCI, Temperature Condition Index; [9] ARI2, Anthocyanin Reflectance Index; [10] CRI2, Carotenoid Reflectance Index; [11] IRECI2, Inverted Red-Edge Chlorophyll Index; [12] LCC, Leaf Chlorophyll Content; [13] PSRI, Plant Senescence Reflectance Index; [14] WDVI, Weighted Difference Vegetation Index; [15] S2 bands, Sentinel-2 bands; [16] PPI, Potato Productivity Index; [17] SLAVI, Specific Leaf Area Vegetation Index; [18] NDMI, Normalized Difference Moisture Index; [19] CIG, Chlorophyll Index Green; [20] Tmax, monthly maximum temperature; [21] P, monthly precipitation; [22] SDrz, daily root-zone soil water depletion; [23] SLR, Single Linear Regression; [24] MLR, Multiple Linear Regression; [25] ANN, Artificial Neural Network; [26] GLM, Generalised Linear Model; [27] rqlasso, Quantile Regression with LASSO penalty; [28] LeapBack, Linear Regression with Backwards Selection; [29] svmL, Support Vector Machine Linear; [30] svmR, Support Vector Machine Radial; [31] svmP, Support Vector Machine Polynomial; [32] MARS, Multivariate adaptive regression splines; [33] RF, Random Forest; [34] kknn, k-Nearest Neighbours; [35] avNNet, Averaged Neural Network.

Proximal handled thermal infrared (TIR) imaging sensors could obtain canopy temperature at higher accuracy, but the accuracy decreased to ±5°C when integrating with UAV due to the impact of environmental conditions for potato yield prediction (Kelly et al., 2019). By capturing imaging data with combining a handheld infrared camera (Ti-32, Fluke Thermography, Glottertal, Germany) and a digital RGB camera (D5100 reflex, Kodak, Tokyo, Japan), an integrated adaptive neuro-fuzzy inference system with a genetic algorithm (ANFIS-GA) was used to predict yield of two potato varieties in a dry crop trial in Egypt (Elsayed et al., 2021). Similarly, we list some studies for potato yield prediction by ground-based RS in the Table 5. Other platforms were not identified for potato yield prediction.

3.2 Selection of RS-based modeling parameters

The spectral, structural, thermal, and textural information of crop canopies are important indicators to explain yield variability (Maimaitijiang et al., 2020). Reflected light from the canopy allows us to estimate the photosynthetic capacity and other crop growth conditions of plants to predict yield. VIs highlight image spectral features to analyze crop phenotypic traits by fusing reflectance information from two or more bands; the normalized difference vegetation index (NDVI) is the most used VI in agriculture RS (Huang et al., 2021). In addition to NDVI, other VIs such as Soil-Adjusted Vegetation Index (SAVI), Ratio Vegetation Index (RVI) and Enhanced Vegetation Index (EVI) are also used in the field of potato yield prediction. Recently, Potato Productivity Index (PPI), calculated based on the two bands at 490 to 945 nm bands considering the key role of water stress on potato tuber development and yield formation, was designed for potato production practices by Gómez et al. (2021). All bands of Sentinel-2, NDVI, PPI, coupled with a random forest (RF) model were adopted to predict potato yield, with the R^2 of 0.77. In addition to moisture, temperature is also an important environmental factor affecting potato tuber development. TIR cameras can generate thermal indices such as normalized relative canopy temperature (NRCT) to monitor temperature change on canopy to monitor drought tolerance (Elsayed et al., 2021). The difference between canopy temperature and air temperature can also reveal the water stress status of potato. The full spectral bands obtained by HS cameras provide more spectral information in visible and near infrared region for potato yield prediction than the spectral index above. Potato yield was predicted by using full-band spectra

TABLE 4 Summary of aerial-based potato yield prediction.

Country	Platform Type	Platform Model	Sensor Type	Sensor Model	Cultivars	Bands number	Bands range	Parameters	Modelling Method	Accuracy	Sample size	References
United States	Plane	Piper Seneca	MS[1]	three Kodak Megaplus 4.2i digital cameras	–	3	G, R, NIR	Integrated SAVI	SLR[11]	R^2 = 0.89	18	Sivarajan, 2011
Japan	UAV	YH300 (Yammer Ltd.)	MS	MSIS, MS2100, DuncanTech, Ltd.	Toyoshiro	4	VIS, NIR	SPAD, stem length	MLR[12]	R^2 = 0.75	10	Yokobori et al., 2004
Japan	UAV	DJI Spreading Wings S900	MS	Micro MCA RGB+3, Tetracam	Toyoshiro	6	425-950 nm	NDVI, height	Alex Net, SLR, MLR	R^2 (NDVI) = 0.18-0.67; R^2 (NDVI & height) = 0.12-1.00; Alex Net: Insufficient precision	23	Tanabe et al., 2019
China	UAV	DJI Matrice 600 Pro	HS[2]	Headwall Nano-Hyperspec (Headwall Photonics Inc., Bolton, MA, USA)	Favorita; Shepody; Zhongshu 5/10/ 18/19	272	400-1000 nm	MCARI[3], CI1[4], CI2[5], MCARI/OSAVI[6], full spectrum, height	PLSR[13], RF[14]	R^2 (PLSR) = 0.81; R^2 (RF) = 0.63	144	Li et al., 2020
United States	UAV	DJI Matrice 600 Pro	HS	Headwall Nano-Hyperspec (Headwall Photonics Inc., Bolton, MA, USA)	6 unknown varieties	273	400-1000 nm	full spectrum	Ridge[15]; OLS[16]; PLSR; SVM[17]; RF; AdaBoost	R^2 (ridge) = 0.63; R^2 (OLS) = 0.13; R^2 (PLSR) = 0.53; R^2 (SVR) = 0.57; R^2 (RF) = 0.51; R^2 (AdaBoost) = 0.45;	96	Sun et al., 2020
United States	UAV	DJI Phantom 4, DJI Inspire 2	MS	NIR + regular camera/senser; Altum multispectral sensor (MicaSense, Seattle, WA, USA)	Russet Burbank; Shepody; Superior	4, 6	VIS, RE, NIR, LWIR	ANTHO[7], GNDVI[8], BNDVI[9]; CHLGR[10], NDVI	GLM[18]	R^2 = 0.34-0.63	288	Jasim et al., 2020
China	UAV	DJI Matrice 600	HS	regular camera, Cubert UHD-185	Zhongshu 3/5	125	450-950 nm	–	PLSR	R^2 = 0.74, NRMSE = 22.37%	42	Wu et al., 2020

(Continued)

10.3389/fpls.2023.1214006

TABLE 4 Continued

Country	Platform Type	Platform Model	Sensor Type	Sensor Model	Cultivars	Bands number	Bands range	Parameters	Modelling Method	Accuracy	Sample size	References
United States	UAV	DJI Inspire 2	MS	GEMS multispectral camera (Sentek Systems LLC, Minneapolis, MN, USA)	–	4	R, G, B, NIR	cultivar information	RF, SVM	R^2 = 0.75-0.79	–	Li et al., 2021

[1] MS, Multispectral; [2] HS, Hyperspectral; [3] MCARI, Modified Chlorophyll Absorption Reflectance Index; [4] CI1, Red-edge Chlorophyll Index 1; [5] CI2, Red-edge Chlorophyll Index 2; [6] OSAVI, Optimised Soil Adjusted Vegetation Index; [7] ANTHO, Anthocyanin; [8] GNDVI, Green Normalized Difference Vegetation Index; [9] BNDVI, Blue Normalized Difference Vegetation Index; [10] CHLGR, Chlorophyll Green; [11] SLR, Single Linear Regression; [12] MLR, Multiple Linear Regression; [13] PLSR, Parcial Least Squares Regression; [14] RF, Random Forest; [15] Ridge, Ridge Regression; [16] OLS, Ordinary Least Squares; [17] SVM, Support Vector Machine; [18] GLM, Generalised Linear Model.

between 400 and 1000 nm (R^2 = 0.81), which had better performance than using plant height and several VIs including Red-edge Chlorophyll Index 1 (CI1), Modified Chlorophyll Absorption Ratio Index (MCARI), Ratio2, and Red-edge Chlorophyll Index 2 (CI2) (R^2 = 0.69) (Li et al., 2020).

It is essential to select the best period to estimate yield because of the great variation in the predicted performance of different growth periods. The weight of values at different growing periods were determined for estimating yield of potato using a handheld hyperspectral camera (Luo et al., 2020). They believed that the tuber expansion period (about 70 days after planting in this study) was the best period for potato yield prediction with an adjusted R^2 = 0.83. In addition, it has also been noted that 90 days after planting is satisfactory for potato yield prediction (Li et al., 2020). This may be caused by the great variation in species and environmental conditions selected for different studies. Meanwhile, many studies have used time series data rather than single period data to predict yields (Fernandes et al., 2017; Aghighi et al., 2018). This could be due to the more comprehensive information on crop growth and development contained within the time-series data. For potato yield prediction, three-date Integrated SAVI (ISAVI) is a better predictor of yield than single-period SAVI (Sivarajan, 2011). Time-series data can also be applied to advanced algorithms such as Long Short-Term Memory networks (LSTM) and three-dimensional Convolutional Neural Network (3D-CNN).

Compared with spectral features, canopy traits such as LAI, plant height, and canopy cover (CC) can reflect the light use efficiency on the canopy. VIs combined with structural parameters such as plant height and LAI provides a better prediction of potato yield (Sharma et al., 2017; Tanabe et al., 2019). In contrast to passive RS, active RS comes with its own radiation source and reflects the characteristics of the ground by transmitting and receiving electromagnetic waves. Their applications are less affected by ambient light and the electromagnetic wave wavelength and emission mode can be set according to different land features, allowing them to obtain the vegetation spatial structure parameters more accurately. In addition to spectral and structural parameters, adding texture features can potentially impact the performance of yield prediction (Ma et al., 2022). However, there are no publications investigating potato yield prediction with texture features extracted from image analysis.

It is also difficult to fully represent the crop growth status by RS data alone. Integrating RS parameters with other indicators of agronomy and meteorology is an effective way to improve yield prediction capability. NDVI combined with plant height provides improved estimation accuracy of potato yield compared with using NDVI alone (Tanabe et al., 2019). Combining soil parameters, such as moisture, conductivity, and nutritional parameters, with NDVI obtained from handheld instruments, potato yield was predicted by Support Vector Machine (SVM) and the determination coefficient of different datasets ranged from 0.54 to 0.72 (Abbas et al., 2020). Likewise, RS information combined with meteorological parameters provides a good prediction of yield with the determination of coefficients ranging from 0.76 to 0.86 in winter and summer growing seasons (Salvador et al., 2020).

TABLE 5 Summary of ground-based potato yield prediction.

Country	Platform Type	Sensor Type	Sensor Model	Cultivars	Bands number	Bands range (nm)	Parameters	Modelling Method	Accuracy	Sample size	References
Germany	Handheld	MS[1]	hand-held multi-spectral radiometer (CROPSCAN, Inc.)	–	2	670, 870	LAI[4], model input parameters	assimilation of LAI and LEACHN model	r = 0.63	36	Finke, 1992
Canada	Handheld	HS[2]	FieldSpec HandHeld spectroradiometer (Analytical Spectral Devices [ASD] Inc., Boulder, CO)	Russet Burbank	200	325-1075	CI1[5]	SLR[13]	r = 0.64	40	Morier et al., 2015
United States	Handheld	MS	GreenSeeker™ (Trimble Navigation Limited, Sunnyvale, CA, USA); Holland Scientific Crop Circle™ ACS 430 (Holland Scientific, Inc., Lincoln, NE, USA)	Russet Burbank	GreenSeeker: 2; Crop Circle: 3	GreenSeeker: 660, 770; Crop Circle: 650, 730, 760	NDVI[6], proprietor-proxy LAI	nonlinear regression	R^2 (GreenSeeker) = 0.60; R^2 (Crop Circle) = 0.64	144	Sharma et al., 2017
United States	Handheld	MS	GreenSeeker™;Holland Scientific Crop Circle™ ACS 430	Russet Burbank, Shepody, Superior	GreenSeeker: 2; Crop Circle: 3	GreenSeeker: 660, 770; Crop Circle: 650, 730, 760	NDVI; NDVI, NDRE[7], CHLRE[8], LAI	nonlinear regression	R^2_{adj} (GreenSeeker) = 0.57; R^2_{adj} (GreenSeeker) = 0.36	288	Jasim et al., 2020
China	Handheld	HS	USB 2000 spectrometer (Ocean Optics, Inc., Dunedin, Florida, United States)	Shepody	1630	350-1100	OCW-based CI	SLR	R^2 = 0.8225; RMSE = 0.2257	27	Luo et al., 2020
Canada	Handheld	MS	FieldScout CM 1000 NDVI Meter (Spectrum Technologies, Aurora, USA)	Russet Burbank	2	660, 840	NDVI, soil data	SLR; EN[14]; k-NN[15]; SVM[16]	R^2 (SVR) = 0.54-0.72; R^2 (SLR) = 0.53-0.70; R^2 (EN) = 0.49-0.64; R^2 (k-NN) = 0.53-0.64	479	Abbas et al., 2020
Egypt	Handheld	Thermal, RGB[3]	handheld infrared thermal camera (Ti-32; Fluke Thermography, Glottertal, Germany); 14-megapixel digital camera (Kodak D5100 reflex; Tokyo, Japan)	Bellini, Arizona	–	7500-14000, 400-700	color Indices, NRTC[10]	SMLR[17]; ANFIS[18]-GA[19]	R^2 (SMLR) = 0.73; R^2 (ANFIS-GA) = 0.80	48	Elsayed et al., 2021
Peru	Handheld	Thermal, RGB	FLIR thermal camera (Model E60, FLIR Systems Inc., Täby, Sweden); digital camera D5300 (Nikon, Thailand)	Unica	–	7500-13000, 400-700	CC[11], dT[12], model input parameters	assimilation of dT and SOLANUM model	R^2 = 0.91-0.99	8	Ninanya et al., 2021

[1] MS, Multispectral; [2] HS, Hyperspectral; [3] RGB, red-green-blue; [4] LAI, Leaf Area Index; [5] CI1, Red-edge Chlorophyll Index 1; [6] NDVI, Normalized Difference Vegetation Index; [7] NDRE, Normalized Difference Red-edge; [8] CHLRE, Chlorophyll Red-edge; [9] OCW, Optimal Combination Weighting Method; [10] NRTC, Normalized Relative Canopy Temperature; [11] CC, Canopy Cover; [12] dT, Canopy temperature minus air temperature; [13] SLR, Single Linear Regression; [14] EN, Elastic Net; [15] k-NN, k-nearest neighbor; [16] SVM, Support Vector Machine; [17] SMLR, Stepwise Multiple Linear Regression; [17] ANFIS, Adaptive Neuro-fuzzy Inference System; [19] GA, Genetic Algorithm.

10.3389/fpls.2023.1214006

3.3 Evolution of RS-based yield prediction methods

Empirical modeling methods, such as Simple Linear Regression (SLR), Multiple Linear Regression (MLR), and Artificial Neural Network (ANN), are mostly used for current RS-based potato yield prediction. LR clearly shows the relationship between one or more explanatory variables and yield. SLR can build a linear relationship between a single parameter and the yield. Introducing more variables including VIs, agronomic parameters, and meteorology by MLR can improve model performance. However, since the relationship between the variables of the dataset is not linear in most real-life scenarios, a non-linear approach is necessary.

With the development of artificial intelligence (AI), ML models have been increasingly applied to RS-based potato yield prediction. An ANN model was constructed with variables including Vegetation Condition Index (VCI) and Temperature Condition Index (TCI) captured by NOAA-AVHRR between 1980-2014 (Akhand et al., 2016). Percentage error was calculated as lower than 10% to quantify the difference between actual and predicted yield. In addition, there are significant differences in prediction accuracy between various ML algorithms. Six ML algorithms (PLSR, Parcial Least Squares Regression; RF, Random Forest; Ridge, Ridge Regression; OLS, Ordinary Least Squares; SVM, Support Vector Machine; GLM: Generalised Linear Model) were compared for potato yield prediction at different irrigation levels using a Headwall nano-hyperspec imager, and ridge regression showed the highest accuracy with R^2 of 0.63 (Sun et al., 2020). Yield prediction performance of several ML algorithms were compared using Sentinel-2 images and svmRadial got the highest accuracy (R^2 = 0.93) (Gómez et al., 2019). It is also worth noting that varietal differences can significantly impact predicted results. ML combined with cultivar information and UAV-based images was used to improve potato yield prediction (Li et al., 2021). The results showed that RF and SVM models using only RS data yield poor estimation (R^2 = 0.48-0.51) but had significantly improved performance (R^2 = 0.75-0.79) when variety information was included. The use of ML algorithms combined with high spatial resolution images and cultivar information can significantly improve yield prediction for different potato varieties than approaches without variety information.

The AlexNet algorithm, proposed in 2012 as the first deep learning (DL) model, generates both low- and high-level features of data through a multilayer neural network as the input of fully connected layers before a classification task (Krizhevsky et al., 2017). Compared to conventional ML with handcrafted features, which reaches a bottleneck in model performance with increasing the size of training dataset, DL can further improve model performance by enlarging the training dataset due to the huge amount of generated features. The performance of MLR and AlexNet to assess potato yield was compared and concluded that the DL algorithm was superior (Tanabe et al., 2019). However, the accuracy of the proposed model was still not high enough to meet the requirement in practice, which encouraged the investigation of more complex DL networks. With the rapid development of DL, other networks including LSTM have been widely applied in yield prediction in recent years (Muruganantham et al., 2022). Many DL-based studies have been conducted in other crops (Tian et al., 2021; Liu et al., 2022). However, there is still a lack of application of such DL methods applied to potato yield prediction.

Different from the empirical models that do not have a complete mechanism of crop development, the physical model of RS considering spectra, radiation, and scattering, are deterministic based on the laws of physics. The PROSAIL, a combination of PROSPECT (leaf optical PROperties SPECTra model) and SAIL (Scattering by Arbitrarily Inclined Leaves model), considering leaf angle, canopy structure and biochemical properties of vegetation is widely used for to estimate chlorophyll content, LAI and biomass (Berger et al., 2018). A mechanistic physical model, such as the radiative transfer model, requires a thorough understanding of vegetation structure characteristics and radiative transfer theory. Probably due to high complexity of the models, physical models have not yet been applied to potato yield prediction.

Compared with the mechanistic models based on RS, the semi-empirical models allow a compromise by estimating intermediate variables or simplifying the model. The commonly used semi-empirical approach is the light use efficiency (LUE) model, which calculates dry matter yield by estimating total primary productivity and combining it with harvest index (Monteith, 1972). According to the principle of assimilates accumulation and distribution, yield could be calculated as the product of photosynthetically active radiation (PAR), fraction of absorbed PAR (FPAR), LUE, and harvest index (HI). Calculation formula is shown below:

$$Yield = PAR \times FPAR \times LUE \times HI$$

4 Crop growth model for potato yield prediction

CGM delineate crop growth and development as a function of environmental factors, such as climatic, soil, and management parameters, predicated upon the physiological and ecological tenets of crops (Raymundo et al., 2014). The mechanistic simulation of potato development is an efficacious tool for predicting potato yield.

4.1 Evolution of potato CGM

The potato growth model has evolved from establishing fundamental principles to widespread application and continuous optimization. The original CGM was established by de Wit at Wageningen University during the 1960s (Bouman et al., 1996). Development of potato CGM began in the late 1970s when researchers designed a model based on physiological characteristics and field experiments. During this stage, the model simulated the accumulation and distribution of assimilates through potential light and thermal conditions, thereby simulating the process of potato yield formation. A simulation of potato growth

was conducted by utilizing temperature, photoperiod, and soil moisture during specific time intervals (Sands et al., 1979). Light interception calculated from canopy cover was used to estimate potential potato yield (Van der Zaag, 1984). Similarly, it was posited that fundamental data such as the time of sowing and harvest, soil and air temperature, and solar radiation can be used to estimate the maximum dry matter yield of potatoes (Marshall et al., 1984). The POTATO model, crafted by Ng and Loomis (1984), is the pioneer mechanistic model for comprehensively delineating the morphology and physiology of potato. During this period, models primarily focused on productivity simulations without considering the effects of environmental stress on actual yield.

In the 1990s, soil water and nitrogen dynamic modules were successively incorporated into potato CGM. For instance, SUBSTOR-potato (Griffin et al., 1993), a sub-module of DSSAT, is utilized to simulate potato growth, with its water and nitrogen dynamics module derived from CERES (Crop Environment Research Synthesis). During the same period, numerous studies explored the optimization of potato irrigation and fertilization management schemes using CGM. Additionally, some research employed models to assess the impact of climate change on potato production.

Currently, models are being widely calibrated and validated across different regions to suit the needs of potato growth simulation. Additionally, CO_2 response modules have been added to the CGM (Wolf, 2002), which have been extensively used for decision support and climate change response studies (Raymundo et al., 2018; Tang et al., 2021). The uncertainty in potato models caused by model structure, model input, and model parameters has already attracted the attention of researchers (Ojeda et al., 2020). This has been extensively studied in other crops (Bert et al., 2007; Wang et al., 2020). In addition, CGM utilize data of specific samples, which cannot reflect the spatial heterogeneity of large-scale farmland. Combination of the high-throughput and full-coverage advantages of RS with the complete mechanism of CGM makes the assimilation of RS and CGM an effective way to achieve continuous spatiotemporal monitoring of potato growth dynamics. For instance, in recent years, AquaCrop has emerged to simulate crop yield using CC as an intermediate variable, which is closely related to RS. However, there is still limited research on predicting potato yield using assimilation techniques. Table 6 illustrates the application of CGM for predicting potato yield in recent years.

4.2 Representative CGM

After more than 40 years of evolution, dozens of potato growth models have been built. The principles of CGM have certain commonalities. Most models include basic crop growth, meteorology, soil, and management modules. The models also have their own focuses and have formed their own schools in various parts of the world and in different application fields. In this subsection, we will introduce some common potato CGM and systematically evaluate their applications over decades.

The potential productivity of a crop can be derived by simulating the net photosynthesis and the percentage of assimilates apportioned to the tubers (Weir et al., 1984). Multiple CGM employ this underlying principle while integrating environmental modules such as soil and climate to simulate yield. Among the early potato models, POTATO stands out as a light-driven model, which completely simulates the growth and development of the crop. Nonetheless, this model is still an oversimplified representation of the crop's growth. Potato yield was effectively modeled by modifying the POTATO model through adjusting the photosynthetic capacity of potatoes on cloudy days (Ewing et al., 1990). In contrast to POTATO, which results in overestimated yields, NPOTATO offers more accurate yield simulation (Wolf, 2002).

SUBSTOR-Potato, a light-driven model derived from CERES, is a more comprehensive and widely used model. For improved accuracy, SUBSTOR-Potato 2.0 added water and nitrogen simulation modules (Griffin et al., 1993). Over decades of research and experimentation, researchers have identified some shortcomings in the model. As the number of studies increases, the model is constantly being refined. In a Canadian study, SUBSTOR-Potato was applied to simulate yield, but an underestimation of 15% occurred due to incorrectly simulated soil moisture content (Mahdian and Gallichand, 1997). Similarly, the model predictions may still underestimate yields under extreme weather conditions. Data from 87 field experiments was synthesized and it was proposed that it is necessary to improve SUBSTOR-Potato to capture the effects of increased atmospheric CO_2 concentration and temperature rise on crop growth (Raymundo et al., 2017). In DSSAT version 4.7, this problem was solved by modifying the response function (Raymundo et al., 2018). However, this version of DSSAT still neglects the impacts of pests and diseases on yield loss caused by quality degradation (Tooley et al., 2021).

LINTUL-Potato, which is based on the light interception and utilization model, carefully considers the influence of temperature and daylength on potato yield formation (Kooman and Haverkort, 1995). Temperature is significant in seedling emergence, light energy utilization, canopy morphogenesis, tuber bulking, and yield formation, while photoperiod has a considerable effect on light energy utilization and potato tuberization (Snyder and Ewing, 1989). By assessing the effect of freezing on yield, the simulated result of LINTUL-Potato showed that an increase in the cold tolerance of potatoes from -1°C to -2°C and -3°C led to respective increases in average yield of 26% and 40% (Hijmans et al., 2003). In addition, numerous models have been derived from LINTUL-Potato that are tailored to various scenarios. Van Delden et al. (2003) simulated nitrogen dynamics and potato yield under different organic nitrogen management strategies in the Netherlands using LINTUL-NPOTATO. Similarly, the LINTUL-Potato model was optimized for simulating yield of potatoes with different genotypes in the Andes Mountains by the International Potato Center (CIP). The revised model known as SOLANUM (Condori et al., 2010) showed acceptable results (R^2>0.88). LINTUL-POTATO-DSS is an enhanced version of LINTUL-Potato that reduces the potential for errors during computation by utilizing fewer parameters (Haverkort et al., 2015).

Simulating the formation and distribution of photosynthetic assimilation products is essential in the CGM. Moreover, moisture

10.3389/fpls.2023.1214006

TABLE 6 Summary of CGM-based potato yield prediction.

Model	Country	Testing time	Varieties	Sites	Accuracy	Sample size	References
APSIM-Potato	Australia	2012/13	Russet Burbank, Moonlight	2	NRMSE = 15.4%	–	Borus et al., 2018
APSIM-Potato	China	1986-2020	–	1	R^2 = 0.92, NRMSE = 14.48%	10	Luo et al., 2022
Aquacrop	Argentina	2009-2010	Spunta	1	R^2 = 0.56	25	Casa et al., 2013
AquaCrop	Denmark	2013-2015	Folva	1	R^2 = 0.63-0.98, NRMSE = 7.3-21%	–	Razzaghi et al., 2017
AquaCrop	Ethiopia	2012	Jalene	2	E = 0.84-0.96, NRMSE=3.49-8%	–	Yibrah et al., 2015
AquaCrop	Iran	2010	Kuzima	1	R^2 = 0.9, NRMSE = 9.21%	–	Afshar et al., 2014
CropSystVB-CSPotato	United States	2001-2002	Ranger Russet	1	–	–	Alva, 2010
Daisy	Poland	2000-2006	Triada	1	RRMSE = 15.4%	–	Mazurczyk et al., 2007
Hamer-model	United Kingdom	3-15 years	–	5	average error % = 15.8	115	Ejieji and Gowing, 2000
Infocrop-potato	India	8-18 years	–	2	NRMSE = 8-10%	–	Govindakrishnan et al., 2007
LINTUL-NPOTATO	Netherlands	1996-1999	Eersteling, Bintje, Junior, Agria	2	R^2 = 0.865, RMSE = 1.08 Mg ha^{-1}	18	Van Delden et al., 2003
LINTUL-POTATO-DSS	South Africa	2013-2014	Innovator	10	R^2 = 0.939, 0.635	6, 9	Machakaire et al., 2016
LPOTCO	Ireland, Germany, Sweden, Finland, UK, Belgium, Italy	1998-1999	Bintje	20	R^2 = 0.65	–	Wolf, 2002
MacKerron and Waister (1985) model, Versteeg and Van Keulen (1986) model	India	1992/1993	Kufri Chandermukhi	1	error % = 4.1-25.7%	2	Prihar et al., 1995
MoDrY	Poland	1971-1996 (excluding 1984)	–	1	R^2 = 0.64, MRE = 12.40%	25	Zyromski et al., 2013
Sands-model	Australia	–	Exton, Sebago, Kennebec, Delaware, Sequoia	9	–	–	Hackett et al., 1979
SSM-iCrop2	Iran	2017-2018	Sante, Arinda, Agria, Marfona	4	r = 0.80, RMSE = 543 g m^{-2}	20	Dadrasi et al., 2020
SUBSTOR-Potato	Argentina	1979-1983, 1987-1991	Huinkul, Kennebec, Mailen and Spunta	4	R^2 = 0.915	24	Travasso et al., 1996
SUBSTOR-Potato	Canada	1992-1993	Kennebec	8, 12	error % = 4-15%	–	Mahdian and Gallichand, 1997
SUBSTOR-Potato	Czech	1994-2002	Rosara	1	R^2 = 0.97 (4 years)	4	Šťastná et al., 2010
SUBSTOR-Potato	Uganda, Burundi, Peru, India, USA	1980, 2002-2010	Asante, Amarilis, Kufri Bahar, Kathadin	5	RRMSE = 28.1%	26	Kleinwechter et al., 2016
SUBSTOR-Potato, AquaCrop	China	2018-2019	Zihuabai	1	DM: R^2 = 0.37-0.68; FM: R^2 = 0.37-0.72	12	Wang et al., 2023

dynamics are critical in determining potato yield. Some models utilize the transpiration or evapotranspiration of the crop or soil as a driver to simulate crop growth and yield formation processes. These water-driven models can assist in the development of rational irrigation practices for efficient utilization of limited water resources. The AquaCrop model (Steduto et al., 2009), developed by the Food and Agriculture Organization (FAO) of the UN, is an example of a water-driven model that calculates biomass as the product of water productivity (WP) and cumulative evapotranspiration, multiplied by a harvest index to determine yield:

$$Yield = B \times HI = WP \times \sum Tr \times HI$$

where Y is final crop yield ($kg{\cdot}m^{-2}$), B is biomass ($kg{\cdot}m^{-2}$), HI is harvest index (%), Tr is transpiration (mm), and WP is water production efficiency ($kg{\cdot}m^{-2}{\cdot}mm^{-1}$).

Compared to other models that require numerous input parameters, the AquaCrop model is relatively simple and demands fewer input parameters. Furthermore, AquaCrop employs CC rather than LAI to depict the canopy structure, which allows for direct use of RS data with this model (Steduto et al., 2009; Sun et al., 2017). AquaCrop was employed to achieve better simulation results for potato tuber yield under varying irrigation conditions (R^2 = 0.98, NRMSE=0.046) (Razzaghi et al., 2017). However, it should be noted that AquaCrop demonstrated limited efficacy in simulating each indicator at higher or lower irrigation levels (Jin et al., 2019).

Another class of water-driven model integrates a CGM with a hydrological model to describe the impacts of alterations in crop water management on potato respiration and yield. A combined SWAP-WOFOST model was employed to evaluate productivity and recommended the inclusion of capillary rise and recirculation in the model to enhance the precision of potato yield prediction (Kroes et al., 2018).

The improvement of crop yields is heavily reliant on the use of fertilizers, particularly nitrogen fertilizers. Nonetheless, excessive nitrogen application can inflate production costs, harm the environment, and pose risks to human health (Zhang et al., 2015). Precise management of nitrogen fertilizer can diminish pollution while also reducing expenses. Researchers recognized the significance of accurate nitrogen management several decades ago and integrated a nitrogen simulation module into the potato model. DAISY, a one-dimensional soil-plant-atmosphere system model, can simulate crop production, soil water balance, carbon and nitrogen cycles, and so on (Plauborg et al., 2022). DAISY was employed to simulate root abscisic acid (ABA) synthesis, stomatal conductance, transpiration and yield under water-saving irrigation conditions in potato crops (Plauborg et al., 2010). Potato yield under different split-N fertigation regimes was simulated and it was observed that prolonged N fertigation consistently increased yield (Zhou et al., 2018).

The APSIM-Potato model is part of the Agricultural Production System Simulator Next Generation (APSIM) family. APSIM simulates potato development and yield formation based on radiation, temperature, photoperiod, soil water, and nitrogen balance in daily increments (Keating et al., 2003). Many studies conducted in Australia and China have focused on water, nitrogen, sowing management, and strategies for coping with climate change (Tang et al., 2021; Li et al., 2022). APSIM-Potato, however, requires additional parameters to improve model performance (Borus et al., 2018).

The World Food Studies (WOFOST) model, developed based on the SUCROS model from Wageningen University, incorporates water and soil simulation modules to primarily simulate regional-scale crop growth and yield changes (Van Diepen et al., 1989). A study that simulated the yield of early potatoes under water-limited conditions indicated that simulation results of WOFOST are sensitive to water deficits (Kulig et al., 2020). Subsequent versions of WOFOST have included a CO_2 response module to better simulate the effects of climate change on potato yields.

4.3 Assimilation methods

The utilization of RS technology enables high-throughput and non-destructive acquisition of crop phenotype data in the field. However, it falls short in simulating the crop yield formation process and lacks a strong mechanistic foundation. Mechanistic CGM simulate crop growth and development as well as yield formation processes, but they use specific samples, an aspect that is lacking in spatial expansion. The assimilation of RS and CGM can leverage the advantages of both to enhance the prediction accuracy of various crop canopy state variables and yields at regional and national scales. Despite several studies being conducted on RS and CGM assimilation for other crops, limited research has been conducted in potato. LAI acquired by Gaofen-1 (GF-1) satellite data was employed as the assimilated variable coupled with DSSAT-SUBSTOR with the SCE-UA optimization algorithm for regional potato yield prediction (Duan, 2019). The mean relative error (MRE) was only 6.17%, 9.45% lower than that of unassimilated RS data. Quiroz et al. (2017) estimated single-point potato yield using CC and the weighted difference vegetation index (WDVI) corrected crop growth model SOLANUM. Current data assimilation algorithms such as Ensemble Kalman Filter (EnKF) and Four-Dimensional Variational Data Assimilation (4DVAR), have emerged, which could lead to further research progress in assimilation studies of RS and CGM. Due to the lack of application of assimilation methods, the technical gaps might increase between yield prediction of potato and other crops.

5 Methods based on yield limiting factor

In the past, agronomists predicted potato yields by the "visual method" using basic conditions of local agricultural production and the growth of potatoes. Considering the impact of yield-enhancing technical measures and the yearly climate on yields, they assessed potato yields per unit area visually and by experience. However, this method relies on the investigators' experience with crop growth and

development patterns and yield formation rules, which is highly subjective. When new situations arise, such as the adoption of new technologies, the promotion of superior varieties, or when crops suffer severe losses due to abnormal disasters, the judgment of crop growth status often exceeds the investigators' experience. This method thus often results in large errors in yield estimates and is rarely used in current agricultural production. Instead, quantitative yield estimation models based on LFs have become an important method for predicting potato yields.

5.1 Agronomic parameters-based methods

Methods employing traditional agronomic parameters were applied in potato yield prediction before the advent of RS and CGM. Earlier studies destructively sampled tubers to record potato weight and number of potatoes to estimate yield directly (Dyke and Avis, 1953). In addition, canopy parameters could be linked to yield by reflecting the growth status. The first category is structure indicators that can directly reflect photosynthetic capacity, such as LAI and leaf number. At the canopy structure level, planting density and number of leaves were used as yield indicators and achieved a reduction of mean squared error (MSE) of 9% (Singh et al., 2020). In addition, some physiological and biochemical indicators have been adopted. The highest correlation between yield and chlorophyll content (expressed as SPAD) has been revealed with an R^2 value of 0.663 (Meng et al., 2021). The nutritional status of the 4th leaf, as measured by the Mg DRIS index (Mgi) and N DRIS index (Ni) during the onset of tuberization, have demonstrated potential as yield predictors. Moreover, N content in stems has shown a strong correlation with marketable tuber yield (MTY), while the Ca:N ratio in stems has displayed the highest correlation with MTY.

Traditional regression methods often have inadequate simulation performance, while some innovative methods provide better prediction results. For instance, a Canadian study demonstrated that using a three-input multiple-layer perceptron (MLP) network with cumulative LAI, maximum LAI, and cumulative rainfall achieved a higher accuracy in yield estimation than MLR and SUBSTOR (Fortin et al., 2011).

Methods based on agronomic parameters often require tedious field sampling. In addition to the plant itself, environmental and management factors, among others, can affect yield. Therefore, the ability of a method to simulate different cultivation and management conditions varies.

5.2 Environmental parameters-based methods

Environmental factors affecting crop growth, such as meteorology, soil, pests, and diseases, could be considered yield indicators. The concept of using meteorological data to replace destructive sampling for predicting potato yield was proposed in a report in 1929 by an unknown author in the American Potato Journal. In addition, the agro-ecological zone (AEZ) model, jointly developed by the FAO of the UN and the International Institute for Applied Systems Analysis (IIASA), predicted the yield potential of different potato farming areas based on statistical data in China from 1961 to 1997 (Cai et al., 2006). In a study examining the relationship between meteorology and yield variability over multiple years, models were constructed with 35 years of data based on MLR, stepwise regression, and BP neural networks, with MREs of 6.715%, 7.811%, and 4.479% (Yang S. et al., 2017). Piekutowska et al. (2021) constructed MLR and ANN models using yield data and meteorological and management data from 2010 to 2017, with the ANN model estimating potato yield more accurately (R^2 = 0.86).

Furthermore, soil indicators have been utilized as predictors of potato yield in some studies. For instance, ANN and MLR models were constructed that incorporated soil infiltration resistance, organic matter, microbial load, and tillage system, resulting in superior prediction of potato yield, with R^2 values of 0.951 and 0.894, respectively (Abrougui et al., 2019). Yield was simulated based on soil apparent conductivity and achieved an R^2 between 0.57 and 0.66 (Frąckowiak et al., 2020). However, yield loss due to pests and diseases also constitutes an essential component of yield prediction theory. It was reported that the rate of yield loss caused by 64 potato cyst nematode eggs per gram of soil ranged from 8.5% to 56% and 9% to 58%, for two experimental sites (Hajihassani et al., 2013).

Several environmental indicators can be used as yield indicators because their variability greatly affects potato yield. However, as with methods based on agronomic parameters, these methods do not reflect the full range of potato yield formation.

5.3 Input-output model

The approach based on input-output theory considers various agricultural inputs, including human labor, energy, fertilizer, irrigation, pesticides, and so on (Chen and Wang, 2010). Some of the input-output model for yield prediction were developed by quantifying the effects of different energy inputs on yield by combining economic mathematical models such as the Cobb-Douglas function. Potato production was estimated in Iran using input-output theory combined with ANN and ANFIS, with correlation coefficients of 0.925 and 0.987, respectively (Khoshnevisan et al., 2014). Farm potato production was compared in Iran combined with various inputs, such as manpower, machinery, diesel, fertilizer, farmyard manure, pesticides, electricity, irrigation water, and seeds (Hamedani et al., 2015). In addition, empirical statistical models consider the yield of previous years as a crucial indicator for yield estimation.

This type of model generally employs a questionnaire to acquire the different forms of energy inputs. However, the most significant aspect of questionnaire approach might be the representativeness of the respondents and the authenticity of the survey data. Additionally, there are some differences in total energy consumption and potato yield among various production models (Al-Hamed and Wahby, 2016). For instance, the energy ratio, energy productivity, and net energy of large-scale farms (>3 ha)

were considerably higher than those of smaller farms (Khoshnevisan et al., 2014). These factors constrain the applicability of prediction methods at different scales.

6 Discussion

This study revealed the research progress of estimating potato yield since 1953. A systematic review of different methods provides an important reference for understanding potato yield prediction applications. In this section, application scenarios, and advantages and disadvantages of various methods are systematically discussed. Additionally, the trend of potato yield prediction is prospected.

6.1 Application scenarios for yield prediction

Different application scenarios necessitate the use of matching yield prediction methods. Yield prediction could be used to create maps of potato yield layout in various regions and to identify the most suitable planting areas within the planting structure layout. Precise pre-harvest yield prediction at the international trade level is beneficial for early adjustment of trade policies to manage evolving international trade dynamics. Furthermore, in large-scale potato production, pre-harvest yield estimation enables development of timely marketing and storage plans to ensure economic benefits. We present several application scenarios and discuss the applicable yield prediction methods below.

6.1.1 Optimal allocation of resources

Potato yield forecasts at the regional and national level can support decision-making in planning growing areas and international trade. Satellite RS is the most intuitive and effective method to meet these demands because it can monitor the extent of potato cultivation over large areas while obtaining various parameters for assessing vegetation growth and predicting final yields. In contrast to traditional statistical survey methods, satellite RS can provide accurate and real-time maps of potato yield distribution. Combining yield mapping for other crops and yield projections for different geographic spaces under future climate change conditions can identify more suitable planting areas for each crop. The integration of satellite RS technology and CGM can facilitate the prediction of crop yields over large areas with greater accuracy and precision. Satellite-based RS could predict yield about two months before harvest to allow earlier development of trade strategies and ensure economic development and food security.

From a farm perspective, potato yield prediction provides decision support for marketing and storage strategies. High-resolution satellites could be used to predict potato yield for large plantings. However, it is preferable to use UAVs to acquire high-resolution image data for slightly smaller farmlands. Companies in smart agriculture could provide these special services. In addition, long-term yield forecasting services, such as MARS Crop Yield Forecasting System (MCYFS), are necessary for all stakeholders.

6.1.2 Precision management

Estimating potential yield through crop models and simulating potato productivity under different environmental conditions can offer suggestions to potato farmers to reduce yield differentials, such as implementing better management strategies for irrigation, fertilization, and sowing period (Deguchi et al., 2016; Li et al., 2022). UAV RS provides a fast and non-destructive way to access field phenotypes and is an important tool for decision support in field management. Various yield estimation models can also be integrated with intelligent management systems to achieve accurate field management.

The image processing and model building for UAV remote sensing and CGM can be challenging and require specialized skills. However, many growers may not have the necessary professional background to conduct yield prediction using these methods. To address this issue, smart agriculture companies with expertise can provide data collection, analysis, and integrated delivery services. For instance, a user interface based on the input-output model was developed using C-sharp that enables direct potato yield prediction (Al-Hamed and Wahby, 2016). In addition, the CGM could provide decision support for precision management by simulating the potato growth process under different management and climate conditions. However, it is still necessary to use different varieties to calibrate CGM in different regions. Due to the complexity of CGM, some simplified models, such as AquaCrop, might be suitable.

As smartphones become more prevalent, one approach to site-specific yield forecasting is to deliver dependable yield projections and decision-making support by incorporating multiple sources of big data into mobile devices and designing an intuitive user interface for potato farmers. It is critical that the yield estimation technique is practical and comprehensible. It should provide unambiguous information as accurately as possible while remaining applicable to commercial farming practices. Dispelling user misunderstandings is crucial in facilitating replication. Conducting pilot studies and demonstrating the effects of yield estimation in specific regions is an effective means of promotion.

6.1.3 Responding to climate change

Assessing the impact of climate change on potato growth and yield is beneficial for selecting the most suitable varieties and management strategies under climate change conditions. However, the impact of climate change on potato production varies due to differences in cultivation areas, growing seasons, and cultivation management practices (Bender and Sentelhas, 2020; Yagiz et al., 2020).

There are many studies pointing to a possible decline in potato yields under future climate change conditions. Applying future climate change scenarios to current potato cropping systems using an improved SUBSTOR-potato, it was pointed to a small decline in global tuber production by 2055 (-2% to -6%) and a large decline by 2085 (-2% to -26%)(Raymundo et al., 2018). DSSAT was used to simulate the yields of barley and potato under future climate change conditions (Holden et al., 2003). Non-irrigated potato tuber production in Ireland is projected to decline in 2055 and 2075 due to water shortages. A study of yield changes under future

climate change scenarios for a variety of crops in Mexico using AquaCrop suggests that Mexican potato yields will decline (Arce Romero et al., 2020).

Interestingly, other studies have produced the opposite result. The possible negative effects of increased temperatures and reduced water availability for potato are offset by the positive effects of increased CO_2 levels on water use efficiency and crop productivity (Haverkort et al., 2013). By mid-century, potato yield variability and productivity will increase in Belgium with greater variability between climate change models than spatial variability (Vanuytrecht et al., 2016). The differences in prediction results could be attributed to the different CGM and study areas used in these investigations. In addition, different climate change scenarios also produce different simulation results (Pushpalatha et al., 2021). Despite extensive research on the impact of climate change on potato yields, validation of previous climate change impact assessment results on yields is still lacking.

According to the prediction results, it is possible to implement corresponding countermeasures, such as changing the planting area or management practices. For example, optimizing sowing and irrigation strategies can improve agro-pastoral ecotone (APE) potato yield and water productivity (Tang et al., 2018). Delayed sowing and selection of medium-maturing potato varieties are important ways to cope with warm and dry climates in the APE of northern China (Li et al., 2019). WOFOST has been integrated into the Crop Growth Monitoring System (CGMS) to assess the impacts of climate change. The impact of climate change on potential productivity of potato was studied in West Bengal using the WOFOST crop growth simulation model (Dua and Sharma, 2017). They minimized the impact of climate change by selecting appropriate varieties and changing planting dates to design management strategies.

6.2 Comparison of different prediction strategies for predicting potato yield

Various prediction methods have their own characteristics. In this section, yield prediction strategies are compared and the knowledge gaps between potato and other crops are presented.

6.2.1 RS-based methods

RS-based methods could replace other destructive sampling methods by obtaining crop phenotype through non-destructive methods to build yield prediction models. Morier et al. (2015) inferred nitrogen stress and estimated potato yield combined with a handheld hyperspectral sensor (FieldSpec handheld spectroradiometer), which obtained similar results to destructive methods. In addition, there are significant variations between different platforms and sensors for predicting potato yield.

Satellite RS can be used to monitor large areas of field with its wide coverage. However, the constraints of cloud cover and revisit cycles might limit the availability of images. Additionally, the low spatial resolution of satellite results in a blended image element composed of soil and vegetation is challenging to process. Yield prediction using satellites faces significant challenges in many regions due to small plot sizes, mixed cropping, and intercropping. Advanced satellites and image processing methods might solve these problems. Satellites used for potato yield estimation are detailed in Table 2. Currently, satellites equipped with spectral sensors are still prevalent, and there is a lack of active RS satellites, such as the C-SAR-equipped GF-3 with high resolution of 1 m, applied to related research. In addition, hyperspectral satellites have not been applied to predict potato yield.

Compared to satellites, proximal RS is more suitable for site-specific rather than regional yield prediction. UAVs are a promising solution for precision agriculture management, given their high resolution, low cost, and flexibility. According to our literature survey, studies on potato estimation with UAV RS have employed a variety of sensors, but there are few applications of UAV-based active sensors. Fixed-wing drones are also applied less in potato yield prediction, although their applications in more rapid monitoring of large areas of farms cannot be ignored. Ground-based RS can capture phenotypes with higher resolution. However, handheld instruments have incomplete spatial coverage and are time-consuming for sampling. Additionally, fixed ground-based phenotyping devices have limited coverage and come with high costs.

Potato yield formation can be described as the production of photosynthetic assimilates multiply by the harvest index (Khan, 2012). The growth dynamics of the aboveground canopy can directly reflect canopy light energy interception and thus affect the formation of photosynthetic assimilation products. Therefore, many practical applications in the studies we investigated have been conducted to predict potato yield by directly obtaining aboveground phenotypes from passive optical sensors. However, they can only obtain information from the top of the canopy and are susceptible to clouds and light. Active sensors that used controlled sources of radiation, such as SAR and LiDAR, are not affected by weather and clouds. Another method is to detect the underground tubers directly by sensors with penetrating ability. Longer wavelengths are more penetrating because the object absorbs less of the wave. There has been a study conducted on the use of sound waves for detecting sweet potatoes in sandy soil (Iwase et al., 2015). However, the research is still in its preliminary stages. Computed tomography (CT) can be used to obtain potato tuber phenotypes non-destructively using X-ray to penetrate the soil for precise yield prediction. Nevertheless, CT is mostly used in indoor pot experiments (Ferreira et al., 2010), and it is difficult to apply it in open conditions due to the associated radiation dangers and the requirements of a receiver after penetrating the tuber. Ground penetrating radar (GPR) is an effective means of non-destructive detection of underground targets. Although GPR is still at the research phase in the detection of underground tubers in potato with limited throughput, this type of tool has great potential for future potato yield prediction (Cheng et al., 2022). Currently, there are few studies comparing active and passive sensors for potato yield prediction, and more research is needed to evaluate the predictive performance of both.

There are some differences in potato RS-based yield estimation compared to other crops. For example, computer vision (CV) is

widely used in intelligent yield prediction to identify and count fruits for further harvesting and marketing decisions (Ramu and Priyadarsini, 2021), but potato tubers located underground are difficult to observe directly by conventional optical sensors. Sound waves and CT could obtain tuber phenotype without the effect of soil, but most studies are limited to potted plants. In complex field conditions, there is a lack of research on obtaining potato tuber phenotype directly for predicting yield. Furthermore, the leaves of potatoes are compound, and the propagation path of light between the leaves is different from that of rice and other field crops. It is necessary to improve or establish a dedicated RS-based yield prediction methods for potatoes, taking into account the inherent characteristics of potato.

Compared to the mechanistic model, most RS-based methods build empirical models by observing aboveground phenotype directly. In addition, these complex relationships between crop growth and external environmental factors, including meteorological conditions, soil nutrients, and field management, remain difficult to accurately capture through RS.

Despite significant progress in the development of theoretical models and operational systems for RS yield estimation, the limitations of current RS technology prevent accurate and quantitative reflection of the underlying mechanisms of crop development and yield formation.

6.2.2 CGM-based methods

CGM simulate yield with comprehensive information of variety, management, meteorological parameters, and soil. CGM thus have a complete mechanism to simulate the crop growth process, which is their most significant property. In addition, along with yield prediction, CGM can simulate other elements such as water and nitrogen dynamics of crops to provide support for precision management. Finally, combining CGM with climate change models could predict the yield potential under different regions, cultivars, and cultivation strategies under future climate change conditions to make optimal decisions.

However, most CGM make accurate simulations based on numerous input parameters, which is often difficult to achieve in potato production. In addition, CGM require laborious field trials to calibrate and validate the models for different varieties in different regions. Somewhat simplified CGM, such as AquaCrop and LINTUL-POTATO-DSS, which significantly reduce the number of input parameters, have also been developed in recent years. Additionally, the establishment of a database platform of shared variety parameters helps to enhance the applicability of the models.

Compared to large scale RS methods, CGM cannot reflect the variation of yield in different spaces by collecting data from specific samples. Assimilation of RS with CGM could extend the model application to a regional scale. In addition, although some models use a constant HI, the effectiveness of such an approach is not always satisfactory. Techniques for assimilating external observations into the model to continually adjust certain state variables and attributes can be used to enhance model performance. RS data can offer prompt updates on crop or environmental conditions, allowing for periodic updates of model simulations during the simulation process (Hao et al., 2021).

Currently, DSSAT-SUBSTOR is still the most widely used model for predicting potato yield. However, due to differences in the modeling principles of individual models, these models often exhibit different simulation results. Uncertainty due to differences in model structure could be reduced by multi-model ensembles (MME), which could improve the prediction performance compared to a single model (Martre et al., 2015).

In addition to yield, commercial attributes of potato are also very important. Tuber yield and tuber size (expressed as number of tubers per 10 kg) were simulated using LINTUL-POTATO-DSS (Machakaire et al., 2016). However, there were few studies that simulated commercial indicators such as potato tuber size. Additionally, unpredictable extreme weather events remain a concern, and even with the ability to anticipate and forecast potential risks, current capacity to cope with them is limited.

6.2.3 Methods based on yield limiting factor

Yield prediction based on physiological and biochemical agronomic indicators, require frequent manual sampling, but achieving full spatial coverage is often challenging. In addition, many agronomic parameter-based approaches employ one or a few parameters for empirical modeling, which might lead to weaker generalizability.

Meteorological parameter-based methods predict potato yields under specific climatic conditions, without laborious field sampling. However, multi-year historical yield and climate data is difficult to obtain in reality. As with the agronomic parameter-based approach, which employs an empirical approach to modeling, this method also oversimplifies. Simulation results are not reliable with drastic changes in weather conditions. Providing farmers with suitable agricultural insurance may be a way to mitigate the adverse effects of unexpected weather conditions. With the growing scope of agricultural informatization, developing a precise and comprehensive agricultural information data platform containing a range of meteorological and yield data is crucial to achieve multi-year yield prediction.

Yield prediction methods based on input-output theory necessitate multiple input parameters, which are frequently derived from interviews with growers. Meanwhile, farmer surveys are also labor-intensive. Ensuring the accuracy of the data collected can be challenging (Fermont and Benson, 2011).

6.3 Uncertainties of yield prediction

Uncertainty is a range centered on the true value, and the larger the range the greater the uncertainty (Lu, 2004). Model uncertainty emerges due to necessary simplification of the real physical process. Differences between the theoretical and real values due to various factors such as assumptions made during model construction, boundary conditions, and the difficulty of reflecting them in calculations at the current state of technology are considered model uncertainty (Xing and Guo, 2006). Potato yield formation is a complex system determined by a combination of cultivation management practices, climatic conditions, soil conditions, and varieties. The yield simulation process could be influenced by any

changes in the above conditions. At the same time, different parameter selections, model inputs, and model structures can lead to uncertainty when constructing potato yield prediction models. It is important to address sources of uncertainty and improve the adaptability of model prediction accuracy and prediction confidence for accurate and reliable potato yield prediction (Ma et al., 2021).

The causes of uncertainty include inherent limitations in predictability (e.g., future greenhouse gas emissions) and deficiencies in forecasting skills (e.g., flaws in model design) (Challinor et al., 2013). Model structure, inputs, and parameters are the three main sources of uncertainty (Wallach et al., 2016). Although the empirical statistical model is computationally simple, there is uncertainty in the functional form and coefficients of different modeling approaches (Wu et al., 2014). Any simple model-based or empirical regression-based inversion is unlikely to produce stable inversion results, and accounting for errors in ground truth and sensor data can improve the estimation accuracy of the parameters in the model (Fermont and Benson, 2011). The mechanistic model simplifies the real growth state of the crop due to assumptions made in the construction process, boundary conditions, and the difficulty of responding in the calculations at the current level of technology. Therefore, the uncertainty caused by the model structure needs to be considered in yield prediction.

Moreover, the accuracy and representativeness of the yield prediction model largely depends on the accuracy of the input data, including weather, soil, and management information. The uncertainty of these input data, especially in large-scale applications, can lead to significant errors in the prediction results. Therefore, it is necessary to further improve the accuracy of input data acquisition and data processing methods to improve the accuracy of yield prediction methods. In terms of RS, its integration with other methods, such as ground-based sensors and machine learning algorithms, can also improve the accuracy and practicality of yield prediction models. For example, RS data noise and environmental stress can increase the prediction uncertainty (Ma et al., 2021).

Uncertainty in model parameters, on the other hand, is a bias in simulation results due to deviations in the selection of parameters. For example, the impact of parameter value uncertainty on spring wheat phenology prediction uncertainty was quantified, and the relative contribution of model structure-driven and parameter value-driven uncertainty to overall prediction uncertainty were assessed (Alderman and Stanfill, 2017).

Few studies address all three sources of uncertainty simultaneously. For CGM, most studies have been devoted to the resolution of model input uncertainty. Sensitivity analysis is the most used uncertainty analysis method to determine which model inputs are more important for simulation results (Matott et al., 2009). In addition, other methods such as Monte Carlo analysis, Bayesian methods, and Generalized Likelihood Uncertainty Estimation (GLUE) can be used for model uncertainty analysis. Currently, most uncertainty assessments focus on the three staple crops wheat, maize, and rice, with less research on potato. Yield forecasting is the most concentrated area of uncertainty research because of its importance and because the results are influenced by many factors (Chapagain et al., 2022).

6.4 Fusing multi-source information

The yield estimation performance of models with a single source of estimation information are often inferior to estimation models that combine data from multiple sources. Previous studies have shown that combining multiple image feature parameters can improve yield prediction accuracy. Incorporating various parameters in RS yield estimation models can improve model performance. The crop model AquaCrop was combined with an economic model for optimizing irrigation management at the farm level (Garcia-Vila and Fereres, 2012). It is significant to use field trials, simulations, and deep learning models to study changing sowing dates to mitigate the effects of climate change (Dewedar et al., 2021). By combining deep learning algorithms with multi-source imagery including LiDAR and optical sensors, crop detection can be significantly improved (Prins and Van Niekerk, 2020). Coupling crop models with RS data is now a common method for yield estimation. RS images can be used to invert the LAI and applied to the crop model. However, the inconsistent relationship of LAI with most VIs and saturation problems lead to uncertainty in yield estimation. Other canopy variables, such as CC can be used in the AquaCrop model (Steduto et al., 2009) for yield prediction instead of LAI.

7 Conclusion

In this paper, methodologies for potato yield prediction and its evolution were comprehensively reviewed. The advantages and disadvantages of various strategies for potato yield prediction were compared. Moreover, the uncertainties of models and multi-source data fusion for yield prediction were discussed, providing a foundation for future studies.

Currently, potato yield prediction on large farmlands commonly employs RS and CGM. RS-based methods obtain farmland image information quickly and comprehensively by making full use of the unique advantages of RS platforms and sensors, however it is difficult to reflect the intrinsic mechanism of crop growth. CGM-based methods simulate the yield formation process of the crop, but their operation is challenging due to the lack of spatial expansion. In addition, methods based on agronomic parameters, meteorological parameters, and input-output theory are also widely used in the field of potato yield prediction.

With the progress in RS platforms, sensor technologies, and AI algorithms, UAVs and satellites equipped with advanced sensors have become mainstream tools for field monitoring and yield prediction at regional scales, which can be used for resource allocation and trade decisions. In addition, with the incorporation of modules, including water, nitrogen, meteorology, and economics, mechanisms of CGM have been more comprehensive. Combined with the comparison and improvement of multi-model ensembles,

potato models are continuously improved in terms of precision management.

Multi-source and time-series data have great potential for future yield prediction, despite the current study using a limited number of varieties and sample sizes for potato yield prediction. In the future, it is necessary to pay attention to large time-series data studies with multiple varieties and large sample sizes.

Author contributions

YL, JL and BL conducted the literature survey and drafted the article; CB, GL, SD, YY, SL, DL and LJ gave valuable comments to the manuscript and carried out critical revisions. JL and CB obtained the funding to support this study. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by National Natural Science Foundation of China (32001485), Breeding new varieties for advantageous agricultural industries in Ningxia - Digital breeding system for potato in China (2019NYYZ01-4) and Key scientific and technological projects of Heilongjiang province in China (2021ZXJ05A05-03) awarded to Jiangang Liu, and China Agriculture Research System (CARS-09-P12) awarded to CB.

Acknowledgments

We are grateful to the reviewers for their valuable comments and recommendations.

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

Abbas, F., Afzaal, H., Farooque, A. A., and Tang, S. (2020). Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy* 10 (7), 1046. doi: 10.3390/agronomy10071046

Abou Ali, H., Delparte, D., and Griffel, L. (2020). "From pixel to yield: Forecasting potato productivity in Lebanon and Idaho," in *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*. Eds. T. R. Jordan and K. Schuckman (Switzerland: ISPRS), 1–7.

Abrougui, K., Gabsi, K., Mercatoris, B., Khemis, C., Amami, R., and Chehaibi, S. (2019). Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). *Soil Tillage Res.* 190, 202–208. doi: 10.1016/j.still.2019.01.011

Afshar, A., Afsharmanesh, G. R., Adeli, M., and Malekian, A. (2014). Assessment of AquaCrop model in the simulation of potato yield and water use efficiency under different water regimes. *J. Biol. Environ. Sci.* 8 (23), 79–86.

Aghighi, H., Azadbakht, M., Ashourloo, D., Shahrabi, H. S., and Radiom, S. (2018). Machine learning regression techniques for the silage maize yield prediction using time-series images of Landsat 8 OLI. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 11 (12), 4563–4577. doi: 10.1109/JSTARS.2018.2823361

Akhand, K., Nizamuddin, M., Roytman, L., and Kogan, F. (2016). "Using remote sensing satellite data and artificial neural network for prediction of potato yield in Bangladesh," in *Remote Sensing and Modeling of Ecosystems for Sustainability XIII*. Eds. W. W. Gao and N. B. Chang (California: SPIE), 52–66.

Alderman, P. D., and Stanfill, B. (2017). Quantifying model-structure- and parameter-driven uncertainties in spring wheat phenology prediction with Bayesian analysis. *Eur. J. Agron.* 88, 1–9. doi: 10.1016/j.eja.2016.09.016

Al-Gaadi, K. A., Hassaballa, A. A., Tola, E., Kayad, A. G., Madugundu, R., Alblewi, B., et al. (2016). Prediction of potato crop yield using precision agriculture techniques. *PloS One* 11 (9), e0162219. doi: 10.1371/journal.pone.0162219

Al-Hamed, S. A., and Wahby, M. F. (2016). Prediction of potato yield based on energy inputs using artificial neural networks and C-sharp under Saudi Arabia conditions. *Biosci. Biotechnol. Res. Asia* 13 (2), 631–644. doi: 10.13005/bbra/2079

Alva, A. K. (2010). Enhancing sustainable nutrient and irrigation management for potatoes. *J. Crop Improv.* 24 (3), 281–297. doi: 10.1080/15427528.2010.487742

Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19 (1), 52–61. doi: 10.1016/j.tplants.2013.09.008

Arce Romero, A., Monterroso Rivas, A. I., Gómez Díaz, J. D., Palacios Mendoza, M. A., Navarro Salas, E. N., López Blanco, J., et al. (2020). Crop yield simulations in Mexican agriculture for climate change adaptation. *Atmósfera* 33 (3), 215–231. doi: 10.20937/ATM.52430

Bala, S. K., and Islam, A. S. (2009). Correlation between potato yield and MODIS-derived vegetation indices. *Int. J. Remote Sens.* 30 (10), 2491–2507. doi: 10.1080/01431160802552744

Bender, F. D., and Sentelhas, P. C. (2020). Assessment of regional climate change impacts on Brazilian potato tuber yield. *Int. J. Plant Prod.* 14 (4), 647–661. doi: 10.1007/s42106-020-00111-7

Berger, K., Atzberger, C., Danner, M., D'Urso, G., Mauser, W., Vuolo, F., et al. (2018). Evaluation of the PROSAIL model capabilities for future hyperspectral model environments: A review study. *Remote Sens.* 10 (1), 85. doi: 10.3390/rs10010085

Bert, F. E., Laciana, C. E., Podestá, G. P., Satorre, E. H., and Menéndez, A. N. (2007). Sensitivity of CERES-Maize simulated yields to uncertainty in soil properties and daily solar radiation. *Agric. Syst.* 94 (2), 141–150. doi: 10.1016/j.agsy.2006.08.003

Borus, D., Parsons, D., Boersma, M., Brown, H., and Mohammed, C. (2018). Improving the prediction of potato productivity: APSIM-Potato model parameterization and evaluation in Tasmania, Australia. *Aust. J. Crop Sci.* 12 (1), 32–43. doi: 10.21475/ajcs.18.12.01.pne570

Bouman, B., Van Keulen, H., Van Laar, H., and Rabbinge, R. (1996). The 'School of de Wit'crop growth simulation models: a pedigree and historical overview. *Agric. Syst.* 52 (2-3), 171–198. doi: 10.1016/0308-521X(96)00011-X

Cai, C., van Velthuizen, H., Fischer, G., and Prieler, S. (2006). Analyses of potato yield potential by chinese farming system zoning based on AEZ model. *Chin. Potato J.* 20 (4), 207–211. doi: 10.3969/j.issn.1672-3635.2006.04.005

Casa, A., Ovando, G., Bressanini, L., and Martinez, J. (2013). Aquacrop model calibration in potato and its use to estimate yield variability under field conditions. *Atmos. Clim. Sci.* 3 (3), 397–407. doi: 10.4236/acs.2013.33041

Challinor, A. J., Smith, M. S., and Thornton, P. (2013). Use of agro-climate ensembles for quantifying uncertainty and informing adaptation. *Agric. For. Meteorol.* 170, 2–7. doi: 10.1016/j.agrformet.2012.09.007

Chapagain, R., Remenyi, T. A., Harris, R. M. B., Mohammed, C. L., Huth, N., Wallach, D., et al. (2022). Decomposing crop model uncertainty: A systematic review. *Field Crops Res.* 279, 108448. doi: 10.1016/j.fcr.2022.108448

Chen, X., and Wang, H. (2010). Some important applications of input-occupancy-output technique. *Chin. J. Manage.* 7 (12), 1737–1740+1748. doi: 10.3969/j.issn.1672-884X.2010.12.001

Cheng, Q., van Verre, W., Podd, F. J., Daniels, D. J., and Peyton, A. J. (2022). "Ground penetrating radar antenna alignment for potato detection," in *2022 23rd International Radar Symposium (IRS)* (Gdansk: IEEE), 54–56.

Condori, B., Hijmans, R. J., Quiroz, R., and Ledent, J.-F. (2010). Quantifying the expression of potato genetic diversity in the high Andes through growth analysis and modeling. *Field Crops Res.* 119 (1), 135–144. doi: 10.1016/j.fcr.2010.07.003

Cooper, M., Voss-Fels, K. P., Messina, C. D., Tang, T., and Hammer, G. L. (2021). Tackling G× E× M interactions to close on-farm yield-gaps: creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity. *Theor. Appl. Genet.* 134, 1625–1644. doi: 10.1007/s00122-021-03812-3

Dadrasi, A., Torabi, B., Rahimi, A., Soltani, A., and Zeinali, E. (2020). Parameterization and evaluation of a simple simulation model (SSM-iCrop2) for potato (Solanum tuberosum L.) growth and yield in Iran. *Potato Res.* 63 (4), 545–563. doi: 10.1007/s11540-020-09456-y

Deguchi, T., Iwama, K., and Haverkort, A. J. (2016). Actual and potential yield levels of potato in different production systems of Japan. *Potato Res.* 59 (3), 207–225. doi: 10.1007/s11540-016-9322-z

dela Torre, D. M. G., Gao, J., and Macinnis-Ng, C. (2021). Remote sensing-based estimation of rice yields using various models: A critical review. *Geo Spatial Inf. Sci.* 24 (4), 580–603. doi: 10.1080/10095020.2021.1936656

Dewedar, O., Plauborg, F., El-Shafie, A., and Marwa, A. (2021). Response of potato biomass and tuber yield under future climate change scenarios in Egypt. *J. Water Land Dev.* 49 (IV-VI), 139–150. doi: 10.24425/jwld.2021.137106

Doraiswamy, P. C., Hodges, T., and Phinney, D. E. (1979) *Crop yield literature review for agristars cropscorn, soybeans, wheat, barley, sorghumrice, cotton, and sunflowers*. Available at: https://ntrs.nasa.gov/api/citations/19800015257/downloads/19800015257.pdf.

Dua, V. K., and Sharma, J. (2017). Forecasting impact of climate change on potato productivity in west Bengal and adaptation strategies. *Indian J. Hortic.* 74 (4), 533–540. doi: 10.5958/0974-0112.2017.00103.7

Duan, D. (2019). *Regional potato production estimation based on data assimilation of remote sensing information and DSSAT-SUBSTOR model.* (Beijing: Chinese Academy of Agricultural Sciences).

Dyke, G. V., and Avis, P. R. D. (1953). A survey of maincrop potatoes I. Estimates of yield 1948-50. *J. Agric. Sci.* 43 (4), 450–455. doi: 10.1017/S0021859600057920

Ejieji, C. J., and Gowing, J. W. (2000). A dynamic model for responsive scheduling of potato irrigation based on simulated water-use and yield. *J. Agric. Sci.* 135, 161–171. doi: 10.1017/S0021859699008102

Elmetwalli, A. H., Fouda, T. Z., and Ali, E. Y. (2014). High resolution satellite imagery to detect stress in potato. *Acta Hortic.* 1038, 97–104. doi: 10.17660/ActaHortic.2014.1038.10

Elsayed, S., El-Hendawy, S., Khadr, M., Elsherbiny, O., Al-Suhaibani, N., Alotaibi, M., et al. (2021). Combining thermal and RGB imaging indices with multivariate and data-driven modeling to estimate the growth, water status, and yield of potato under different drip irrigation regimes. *Remote Sens.* 13 (9), 1679. doi: 10.3390/rs13091679

Ewing, E. E., Heym, W. D., Batutis, E. J., Snyder, R. G., Ben Khedher, M., Sandlan, K. P., et al. (1990). Modifications to the simulation model POTATO for use in New York. *Agric. Syst.* 33 (2), 173–192. doi: 10.1016/0308-521X(90)90079-6

Fermont, A., and Benson, T. (2011). Estimating yield of food crops grown by smallholder farmers: a review in the Uganda context. *IFPRI Discussion Papers (1097).*

Fernandes, J. L., Ebecken, N. F. F., and Esquerdo, J. C. D. M. (2017). Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble. *Int. J. Remote Sens.* 38 (16), 4631–4644. doi: 10.1080/01431161.2017.1325531

Ferreira, S. J., Senning, M., Sonnewald, S., Keßling, P.-M., Goldstein, R., and Sonnewald, U. (2010). Comparative transcriptome analysis coupled to X-ray CT reveals sucrose supply and growth velocity as major determinants of potato tuber starch biosynthesis. *BMC Genomics* 11 (1), 1–17. doi: 10.1186/1471-2164-11-93

Finke, P. A. (1992). Integration of remote sensing data in the simulation of spatially variable yield of potatoes. *Soil Technol.* 5 (3), 257–270. doi: 10.1016/0933-3630(92)90026-W

Fortin, J. G., Anctil, F., Parent, L.-E., and Bolinder, M. A. (2011). Site-specific early season potato yield forecast by neural network in Eastern Canada. *Precis. Agric.* 12 (6), 905–923. doi: 10.1007/s11119-011-9233-6

Frąckowiak, K., Potarzycki, J., Grzebisz, W., and Szczepaniak, W. (2020). Potato nutritional status at the onset of tuberisation – a yield prediction tool. *Plant Soil Environ.* 66 (2), 86–92. doi: 10.17221/533/2019-PSE

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 (3), 508. doi: 10.3390/rs12030508

Garcia-Vila, M., and Fereres, E. (2012). Combining the simulation crop model AquaCrop with an economic model for the optimization of irrigation management at farm level. *Eur. J. Agron.* 36 (1), 21–31. doi: 10.1016/j.eja.2011.08.003

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food security: the challenge of feeding 9 billion people. *Science* 327 (5967), 812–818. doi: 10.1126/science.1185383

Gómez, D., Salvador, P., Sanz, J., and Casanova, J. L. (2019). Potato yield prediction using machine learning techniques and Sentinel 2 data. *Remote Sens.* 11 (15), 1745. doi: 10.3390/rs11151745

Gómez, D., Salvador, P., Sanz, J., and Casanova, J. L. (2021). New spectral indicator Potato Productivity Index based on Sentinel-2 data to improve potato yield prediction: a machine learning approach. *Int. J. Remote Sens.* 42 (9), 3430–3448. doi: 10.1080/01431161.2020.1871102

Govindakrishnan, P. M., Singh, J. P., Lal, S. S., and Sushma, P. (2007). A methodology for pre-harvest prediction of mean potato yield at regional scale using infocrop-potato model. *Potato J.* 34 (1/2), 125–126.

Griffin, T. S., Johnson, B. S., and Ritchie, J. T. (1993). *A simulation model for potato growth and development: Substor-potato Version 2.0.* (Michigan: IBSNAT Project).

Hackett, C., Sands, P. J., and Nix, H. A. (1979). A model of the development and bulking of potatoes (Solanum tuberosum L.) II. Prediction of district commercial yields. *Field Crops Res.* 2 (C), 333–347. doi: 10.1016/0378-4290(79)90032-7

Hajihassani, A., Ebrahimian, E., and Hajihasani, M. (2013). Estimation of yield damage in potato caused by iranian population of *Globodera Rostochiensis* with and without aldicarb under greenhouse conditions. *Int. J. Agric. Biol.* 15 (2), 352–356.

Hamedani, S. R., Liaqat, M., Shamshirband, S., Al-Razgan, O. S., Al-Shammari, E. T., and Petkovic, D. (2015). Comparative study of soft computing methodologies for energy input-output analysis to predict potato production. *Am. J. Potato Res.* 92 (3), 426–434. doi: 10.1007/s12230-015-9453-9

Hao, S., Ryu, D., Western, A., Perry, E., Bogena, H., and Franssen, H. J. H. (2021). Performance of a wheat yield prediction model and factors influencing the performance: A review and meta-analysis. *Agric. Syst.* 194, 103278. doi: 10.1016/j.agsy.2021.103278

Haverkort, A. J., Franke, A. C., Engelbrecht, F. A., and Steyn, J. M. (2013). Climate change and potato production in contrasting south african agro-ecosystems 1. Effects on land and water use efficiencies. *Potato Res.* 56 (1), 31–50. doi: 10.1007/s11540-013-9230-4

Haverkort, A., Franke, A., Steyn, J. M., Pronk, A., Caldiz, D., and Kooman, P. (2015). A robust potato model: LINTUL-Potato-DSS. *Potato Res.* 58 (4), 313–327. doi: 10.1007/s11540-015-9303-7

Hijmans, R. J., Condori, B., Carrillo, R., and Kropff, M. J. (2003). A quantitative and constraint-specific method to assess the potential impact of new agricultural technology: the case of frost resistant potato for the Altiplano (Peru and Bolivia). *Agric. Syst.* 76 (3), 895–911. doi: 10.1016/s0308-521x(02)00081-1

Hill, J. D., Strommen, N. D., Sakamoto, C. M., and Leduc, S. K. (1980). LACIE - application of meteorology for united-states and foreign wheat assessment. *J. Appl. Meteorol.* 19 (1), 22–34. doi: 10.1175/1520-0450(1980)019<0022:Laomfu>2.0.Co;2

Holden, N. M., Brereton, A. J., Fealy, R., and Sweeney, J. (2003). Possible change in Irish climate and its impact on barley and potato yields. *Agric. For. Meteorol.* 116 (3-4), 181–196. doi: 10.1016/S0168-1923(03)00002-9

Holm, D. G., and Nylund, R. E. (1978). Use of mineral element content of potato petioles for predicting yield potential. *Am. Potato J.* 55 (6), 291–305. doi: 10.1007/BF02852071

Huang, S., Tang, L., Hupy, J. P., Wang, Y., and Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *J. For. Res.* 32 (1), 1–6. doi: 10.1007/s11676-020-01155-1

Iwase, J., Sato, Y., Comparini, D., Masi, E., Mancuso, S., and Kawano, T. (2015). Non-invasive acoustic sensing of tuberous roots of sweet potato (Ipomoea batatas) growing belowground. *Adv. Hortic. Sci.* 29 (4), 176–180.

Jasim, A., Zaeen, A., Sharma, L. K., Bali, S. K., Wang, C., Buzza, A., et al. (2020). Predicting phosphorus and potato yield using active and passive sensors. *Agriculture* 10 (11), 564. doi: 10.3390/agriculture10110564

Jin, J., Huang, J., and Gui, L. (2019). Optimization of potato irrigation system based on AquaCrop model. *Acta Agric. Boreali-occident. Sin.* 28 (08), 1250–1258. doi: 10.7606/j.issn.1004-1389.2019.08.006

Keating, B. A., Carberry, P. S., Hammer, G. L., Probert, M. E., Robertson, M. J., Holzworth, D., et al. (2003). An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agron.* 18 (3-4), 267–288. doi: 10.1016/S1161-0301(02)00108-9

Kelly, J., Kljun, N., Olsson, P.-O., Mihai, L., Liljeblad, B., Weslien, P., et al. (2019). Challenges and best practices for deriving temperature data from an uncalibrated UAV thermal infrared camera. *Remote Sens.* 11 (5), 567. doi: 10.3390/rs11050567

Khan, M. S. (2012). *Assessing genetic variation in growth and development of potato*. (Wageningen: Wageningen University).

Khoshnevisan, B., Rafiee, S., Omid, M., and Mousazadeh, H. (2014). Prediction of potato yield based on energy inputs using multi-layer adaptive neuro-fuzzy inference system. *Measurement* 47, 521–530. doi: 10.1016/j.measurement.2013.09.020

Kleinwechter, U., Gastelo, M., Ritchie, J., Nelson, G., and Asseng, S. (2016). Simulating cultivar variations in potato yields for contrasting environments. *Agric. Syst.* 145, 51–63. doi: 10.1016/j.agsy.2016.02.011

Kooman, P., and Haverkort, A. (1995). "Modelling development and growth of the potato crop influenced by temperature and daylength: LINTUL-POTATO," in *Potato ecology and modelling of crops under conditions limiting growth* (Wageningen: Springer), 41–59.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60 (6), 84–90. doi: 10.1145/3065386

Kroes, J., Supit, I., Van Dam, J., Van Walsum, P., and Mulder, M. (2018). Impact of capillary rise and recirculation on simulated crop yields. *Hydrol. Earth Syst. Sci.* 22 (5), 2937–2952. doi: 10.5194/hess-22-2937-2018

Kulig, B., Skowera, B., Klimek-Kopyra, A., Kołodziej, S., and Grygierzec, W. (2020). The use of the WOFOST model to simulate water-limited yield of early potato cultivars. *Agronomy* 10 (1), 81. doi: 10.3390/agronomy10010081

Kumar, P., Dubey, S., Kimothi, M. M., Neetu,, Mamatha, S., and Ray, S. S. (2019). "Analysis of remote sensing-based assessment of potato statistics and its comparison with government estimates," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (New Delhi: ISPRS Archives) XLII-3/W6, 299–305.

Li, D., Miao, Y., Gupta, S. K., Rosen, C. J., Yuan, F., Wang, C., et al. (2021). Improving potato yield prediction by combining cultivar information and UAV remote sensing data using machine learning. *Remote Sens.* 13 (16), 3322. doi: 10.3390/rs13163322

Li, Y., Wang, J., Fang, Q., Hu, Q., Zhang, J., Pan, Z., et al. (2022). Optimal planting dates for diverse crops in Inner Mongolia. *Field Crops Res.* 275, 108365. doi: 10.1016/j.fcr.2021.108365

Li, Y., Wang, J., Tang, J., Huang, M., Bai, H., Wang, N., et al. (2019). Coupling impacts of planting date and cultivar on potato yield. *Chin. J. Eco-Agric.* 27 (2), 296–304. doi: 10.13930/j.cnki.cjea.180707

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

Liu, N., Townsend, P. A., Naber, M. R., Bethke, P. C., Hills, W. B., and Wang, Y. (2021). Hyperspectral imagery to monitor crop nutrient status within and across growing seasons. *Remote Sens. Environ.* 255, 112303. doi: 10.1016/j.rse.2021.112303

Liu, Y., Wang, S., Chen, J., Chen, B., Wang, X., Hao, D., et al. (2022). Rice yield prediction and model interpretation based on satellite and climatic indicators using a transformer method. *Remote Sens.* 14 (19), 5045. doi: 10.3390/rs14195045

Lu, S. (2004). *Geographic information system* (Beijing: Higher Education Press).

Luo, S., He, Y., Li, Q., Jiao, W., Zhu, Y., and Zhao, X. (2020). Nondestructive estimation of potato yield using relative variables derived from multi-period LAI and hyperspectral data based on weighted growth stage. *Plant Methods* 16 (1), 150. doi: 10.1186/s13007-020-00693-3

Luo, X., Sun, Y., Liu, L., Wang, L., Yang, L., and Gao, X. (2022). Simulation of response of potato growth and yield to drought stress in the singlecropping region in northern China: A case of Wuchuan County. *Arid Land Geogr.* 45 (3), 867–878. doi: 10.12118/j.issn.1000-6060.2021.350

Lyalin, K. S., Biryuk, A. A., Sheremet, A. Y., Tsvetkov, V. K., and Prikhodko, D. V. (2018). "UAV synthetic aperture radar system for control of vegetation and soil moisture," in *2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2018*. Ed. S. Shaposhnikov (Moscow and St. Petersburg: IEEE), 1673–1675.

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. doi: 10.3389/fpls.2022.925986

Ma, Y., Zhang, Z., Kang, Y., and Özdoğan, M. (2021). Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sens. Environ.* 259, 112408. doi: 10.1016/j.rse.2021.112408

MacDonald, R., Hall, F., and Erb, R. (1975). "The use of Landsat data in a large area crop inventory experiment (LACIE)," in *LARS Symposia* (Indiana: IEEE), Paper 46.

Machakaire, A. T. B., Steyn, J. M., Caldiz, D. O., and Haverkort, A. J. (2016). Forecasting yield and tuber size of processing potatoes in south africa using the LINTUL-Potato-DSS model. *Potato Res.* 59 (3), 195–206. doi: 10.1007/s11540-016-9321-0

Mahdian, M. H., and Gallichand, J. (1997). Estimating potato yield with the SUBSTOR model in Quebec. *Can. Agric. Eng.* 39 (3), 157–164.

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

Marshall, B., Heilbron, T., MacKerron, D., and Waister, P. (1984). "Modelling potential dry-matter production," in *Proceedings of the 9th Triennial Conference of the European Association for Potato Research*. 340–341.

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J. W., Rötter, R. P., et al. (2015). Multimodel ensembles of wheat growth: many models are better than one. *Global Change Biol.* 21 (2), 911–925. doi: 10.1111/gcb.12768

Matott, L. S., Babendreier, J. E., and Purucker, S. T. (2009). Evaluating uncertainty in integrated environmental models: A review of concepts and tools. *Water Resour. Res.* 45, W06421. doi: 10.1029/2008WR007301

Mazurczyk, W., Nowacki, W., and Takac, J. (2007). Comparison of the production and environmental effects of different potato cultivation systems based on the simulation experiment: the Daisy model. *Acta Scientiarum Polonorum Agricultura* 6 (3), 27–34.

Meng, X., Han, Z., Zhang, S., Sun, J., Xv, F., Xv, S., et al. (2021). Dynamic changes of chlorophyll in potato leaves and their correlation with yield. *J. Jilin Agric. Sci.* 46 (03), 79–81.

Mhango, J. K., Harris, E. W., Green, R., and Monaghan, J. M. (2021). Mapping potato plant density variation using aerial imagery and deep learning techniques for precision agriculture. *Remote Sens.* 13 (14), 2705. doi: 10.3390/rs13142705

Monteith, J. L. (1972). Solar radiation and productivity in tropical ecosystems. *J. Appl. Ecol.* 9 (3), 747–766. doi: 10.2307/2401901

Morier, T., Cambouris, A. N., and Chokmani, K. (2015). In-Season nitrogen status assessment and yield estimation using hyperspectral vegetation indices in a potato crop. *Agron. J.* 107 (4), 1295–1309. doi: 10.2134/agronj14.0402

Muruganantham, P., Wibowo, S., Grandhi, S., Samrat, N. H., and Islam, N. (2022). A systematic literature review on crop yield prediction with deep learning and remote sensing. *Remote Sens.* 14 (9), 1990. doi: 10.3390/rs14091990

Nakalembe, C., Becker-Reshef, I., Bonifacio, R., Hu, G., Humber, M. L., Justice, C. J., et al. (2021). A review of satellite-based global agricultural monitoring systems available for Africa. *Global Food Secur.* 29, 100543. doi: 10.1016/j.gfs.2021.100543

Newton, I. H., Islam, A. M. T., Islam, A. S., Islam, G. T., Tahsin, A., and Razzaque, S. (2018). Yield prediction model for potato using landsat time series images driven vegetation indices. *Remote Sens. Earth Sys. Sci.* 1 (1), 29–38. doi: 10.1007/s41976-018-0006-0

Ng, N., and Loomis, R. S. (1984). *Simulation of growth and yield of the potato crop* Vol. 4 (Wageningen: Centre for Agricultural Publishing and Documentation, P.O. Box), 6700 AA.

Ninanya, J., Ramirez, D. A., Rinza, J., Silva-Diaz, C., Cervantes, M., Garcia, J., et al. (2021). Canopy temperature as a key physiological trait to improve yield prediction under water restrictions in potato. *Agronomy* 11 (7), 1436. doi: 10.3390/agronomy11071436

Ojeda, J. J., Rezaei, E. E., Remenyi, T. A., Webb, M. A., Webber, H. A., Kamali, B., et al. (2020). Effects of soil- and climate data aggregation on simulated potato yield and irrigation water requirement. *Sci. Total Environ.* 710, 135589. doi: 10.1016/j.scitotenv.2019.135589

Piekutowska, M., Niedbala, G., Piskier, T., Lenartowicz, T., Pilarski, K., Wojciechowski, T., et al. (2021). The application of multiple linear regression and artificial neural network models for yield prediction of very early potato cultivars before harvest. *Agronomy* 11 (5), 885. doi: 10.3390/agronomy11050885

Plauborg, F., Abrahamsen, P., Gjettermann, B., Mollerup, M., Iversen, B. V., Liu, F., et al. (2010). Modelling of root ABA synthesis, stomatal conductance, transpiration and potato production under water saving irrigation regimes. *Agric. Water Manage.* 98 (3), 425–439. doi: 10.1016/j.agwat.2010.10.006

Plauborg, F., Motarjemi, S. K., Nagy, D., and Zhou, Z. (2022). Analysing potato response to subsurface drip irrigation and nitrogen fertigation regimes in a temperate environment using the Daisy model. *Field Crops Res.* 276, 108367. doi: 10.1016/j.fcr.2021.108367

Prihar, S. S., Arora, V. K., Singh, G., and Singh, R. (1995). Estimating potato-tuber yield in a subtropical environment with simple radiation-based models. *Exp. Agric.* 31 (1), 65–73. doi: 10.1017/S0014479700025023

Prins, A. J., and Van Niekerk, A. (2020). Regional mapping of vineyards using machine learning and LiDAR data. *Int. J. Appl. Geospat. Res.* 11 (4), 1–22. doi: 10.4018/IJAGR.2020100101

Pushpalatha, R., Santhosh Mithra, V., Sunitha, S., Goerge, J., Nedunchezhiyan, M., Mamatha, K., et al. (2021). Impact of climate change on the yield of tropical root and tuber crops vs. rice and potato in India. *Food Secur* 14 (2), 495-508. doi: 10.1007/s12571-021-01226-z

Quiroz, R., Loayza, H., Barreda, C., Gavilan, C., Posadas, A., and Ramirez, D. A. (2017). Linking process-based potato models with light reflectance data: Does model complexity enhance yield prediction accuracy? *Eur. J. Agron.* 82, 104–112. doi: 10.1016/j.eja.2016.10.008

Raj, T., Hanim Hashim, F., Baseri Huddin, A., Ibrahim, M. F., and Hussain, A. (2020). A survey on LiDAR scanning mechanisms. *Electronics* 9 (5), 741. doi: 10.3390/electronics9050741

Ramu, K., and Priyadarsini, K. (2021). "A review on crop yield prediction using machine learning methods," in *Proceedings - 2nd International Conference on Smart Electronics and Communication, ICOSEC 2021* (Trichy: IEEE). 1239–1245. doi: 10.1109/ICOSEC51865.2021.9591764

Raymundo, R., Asseng, S., Cammarano, D., and Quiroz, R. (2014). Potato, sweet potato, and yam models for climate change: A review. *Field Crops Res.* 166, 173–185. doi: 10.1016/j.fcr.2014.06.017

Raymundo, R., Asseng, S., Prassad, R., Kleinwechter, U., Concha, J., Condori, B., et al. (2017). Performance of the SUBSTOR-potato model across contrasting growing conditions. *Field Crops Res.* 202, 57–76. doi: 10.1016/j.fcr.2016.04.012

Raymundo, R., Asseng, S., Robertson, R., Petsakos, A., Hoogenboom, G., Quiroz, R., et al. (2018). Climate change impact on global potato production. *Eur. J. Agron.* 100, 87–98. doi: 10.1016/j.eja.2017.11.008

Razzaghi, F., Zhou, Z., Andersen, M. N., and Plauborg, F. (2017). Simulation of potato yield in temperate condition by the AquaCrop model. *Agric. Water Manage.* 191, 113–123. doi: 10.1016/j.agwat.2017.06.008

Ritchie, J. T., Griffin, T. S., and Johnson, B. S. (1995). *SUBSTOR: functional model of potato growth, development and yield* (Wageningen: Wageningen Pers), 401–435.

Ryerson, R. A., Dobbins, R. N., and Thibault, C. (1985). Timely crop area estimates from landsat. *Photogramm. Eng. Remote Sens.* 51 (11), 1735–1743.

Salvador, P., Gómez, D., Sanz, J., and Casanova, J. L. (2020). Estimation of potato yield using satellite data at a municipal level: A machine learning approach. *ISPRS Int. J. Geo- nf.* 9 (6), ijgi9060343. doi: 10.3390/ijgi9060343

Sands, P. J., Hackett, C., and Nix, H. A. (1979). A model of the development and bulking of potatoes (Solanum Tuberosum L.) I. Derivation from well-managed field crops. *Field Crops Res.* 2 (C), 309–331. doi: 10.1016/0378-4290(79)90031-5

Sharma, L. K., Bali, S. K., Dwyer, J. D., Plant, A. B., and Bhowmik, A. (2017). A case study of improving yield prediction and sulfur deficiency detection using optical sensors and relationship of historical potato yield with weather data in Maine. *Sensors* 17 (5), 1095. doi: 10.3390/s17051095

Singh, N. O., Singh, N. G., Singh, L. N., Kumar, S., and Paull, A. K. (2020). Forecasting model of potato yield from farmers' fields in Manipur. *Int. J. Agric. Stat. Sci.* 16 (1), 401–405.

Singha, C., and Swain, K. C. (2022). "Evaluating the NDVI based Rice and Potato Yield Prediction map Using GIS Geostatistical Environment," in *2nd International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies, ICAECT 2022* (Bhilai: IEEE). doi: 10.1109/ICAECT54875.2022.9807981

Sivarajan, S. (2011). *Estimating yield of irrigated potatoes using aerial and satellite remote sensing*. (Utah: Utah State University).

Snyder, R. G., and Ewing, E. E. (1989). Interactive effects of temperature, photoperiod, and cultivar on tuberization of potato cuttings. *HortScience* 24 (2), 336–338. doi: 10.21273/HORTSCI.24.2.336

Šťastná, M., Toman, F., and Dufková, J. (2010). Usage of SUBSTOR model in potato yield prediction. *Agric. Water Manage.* 97 (2), 286–290. doi: 10.1016/j.agwat.2009.09.015

Steduto, P., Hsiao, T. C., Raes, D., and Fereres, E. (2009). Aquacrop-the FAO crop model to simulate yield response to water: I. concepts and underlying principles. *Agron. J.* 101 (3), 426–437. doi: 10.2134/agronj2008.0139s

Stöcker, C., Bennett, R., Nex, F., Gerke, M., and Zevenbergen, J. (2017). Review of the current state of UAV regulations. *Remote Sens.* 9 (5), 459. doi: 10.3390/rs9050459

Sun, C., Feng, L., Zhang, Z., Ma, Y., Crosby, T., Naber, M., et al. (2020). Prediction of end-of-season tuber yield and tuber set in potatoes using in-season UAV-based hyperspectral imagery and machine learning. *Sensors* 20 (18), 5293. doi: 10.3390/s20185293

Sun, S., Zhang, L., Chen, Z., and Sun, J. (2017). Advances in AquaCrop model research and application. *Sci. Agric. Sin.* 50 (17), 3286–3299. doi: 10.3864/j.issn.0578-1752.2017.17.004

Tanabe, D., Ichiura, S., Nakatsubo, A., Kobayashi, T., and Katahira, M. (2019). "Yield prediction of potato by unmanned aerial vehicle," in *TAE 2019 - Proceeding of 7th International Conference on Trends in Agricultural Engineering 2019* (Prague: TAE). 540–546.

Tandzi, L. N., and Mutengwa, C. S. (2020). Estimation of Maize (Zea mays L.) Yield Per Harvest Area: Appropriate methods. *Agronomy* 10 (1), 29. doi: 10.3390/agronomy10010029

Tang, J., Wang, J., Fang, Q., Wang, E., Yin, H., and Pan, X. (2018). Optimizing planting date and supplemental irrigation for potato across the agro-pastoral ecotone in North China. *Eur. J. Agron.* 98, 82–94. doi: 10.1016/j.eja.2018.05.008

Tang, J., Xiao, D., Wang, J., Fang, Q., Zhang, J., and Bai, H. (2021). Optimizing water and nitrogen managements for potato production in the agro-pastoral ecotone in North China. *Agric. Water Manage.* 253, 106945. doi: 10.1016/j.agwat.2021.106945

ten Harkel, J., Bartholomeus, H., and Kooistra, L. (2020). Biomass and crop height estimation of different crops using UAV-based lidar. *Remote Sens.* 12 (1), 17. doi: 10.3390/rs12010017

Tian, H., Wang, P., Tansey, K., Zhang, J., Zhang, S., and Li, H. (2021). An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain, PR China. *Agric. For. Meteorol.* 310, 108629. doi: 10.1016/j.agrformet.2021.108629

Tooley, B. E., Mallory, E. B., Porter, G. A., and Hoogenboom, G. (2021). Predicting the response of a potato-grain production system to climate change for a humid continental climate using DSSAT. *Agric. For. Meteorol.* 307, 108452. doi: 10.1016/j.agrformet.2021.108452

Travasso, M. I., Caldiz, D. O., and Saluzzo, J. A. (1996). Yield prediction using the SUBSTOR-potato model under Argentinian conditions. *Potato Res.* 39 (3), 305–312. doi: 10.1007/BF02360922

Van Delden, A., Schröder, J. J., Kropff, M. J., Grashoff, C., and Booij, R. (2003). Simulated potato yield, and crop and soil nitrogen dynamics under different organic nitrogen management strategies in The Netherlands. *Agric. Ecosyst. Environ.* 96 (1-3), 77–95. doi: 10.1016/S0167-8809(03)00012-4

Van der Velde, M., and Nisini, L. (2019). Performance of the MARS-crop yield forecasting system for the European Union: Assessing accuracy, in-season, and year-to-year improvements from 1993 to 2015. *Agric. Syst.* 168, 203–212. doi: 10.1016/j.agsy.2018.06.009

Van der Zaag, D. (1984). Reliability and significance of a simple method of estimating the potential yield of the potato crop. *Potato Res.* 27, 51–73. doi: 10.1007/BF02356197

Van Diepen, C. A., Wolf, J., van Keulen, H., and Rappoldt, C. (1989). WOFOST: a simulation model of crop production. *Soil Use Manage.* 5 (1), 16–24. doi: 10.1111/j.1475-2743.1989.tb00755.x

Vannoppen, A., and Gobin, A. (2022). Estimating yield from NDVI, weather data, and soil water depletion for sugar beet and potato in Northern Belgium. *Water* 14 (8), 1188. doi: 10.3390/w14081188

Vanuytrecht, E., Raes, D., and Willems, P. (2016). Regional and global climate projections increase mid-century yield variability and crop productivity in Belgium. *Reg. Environ. Change* 16 (3), 659–672. doi: 10.1007/s10113-015-0773-6

Wallach, D., Thorburn, P., Asseng, S., Challinor, A. J., Ewert, F., Jones, J. W., et al. (2016). Estimating model prediction error: Should you treat predictions as fixed or random? *Environ. Modell. Softw.* 84, 529–539. doi: 10.1016/j.envsoft.2016.07.010

Wang, H., Cheng, M., Liao, Z., Guo, J., Zhang, F., Fan, J., et al. (2023). Performance evaluation of AquaCrop and DSSAT-SUBSTOR-Potato models in simulating potato growth, yield and water productivity under various drip fertigation regimes. *Agric. Water Manage.* 276, 108076. doi: 10.1016/j.agwat.2022.108076

Wang, B., Feng, P., Liu, D. L., O'Leary, G. J., Macadam, I., Waters, C., et al. (2020). Sources of uncertainty for wheat yield projections under future climate are site-specific. *Nat. Food* 1 (11), 720–72+. doi: 10.1038/s43016-020-00181-w

Weir, A., Bragg, P., Porter, J., and Rayner, J. (1984). A winter wheat crop simulation model without water or nutrient limitations. *J. Agric. Sci.* 102 (2), 371–382. doi: 10.1017/S0021859600042702

Weiss, M., Jacob, F., and Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* 236, 111402. doi: 10.1016/j.rse.2019.111402

Wolf, J. (2002). Comparison of two potato simulation models under climate change. I. Model calibration and sensitivity analyses. *Clim. Res.* 21 (2), 173–186. doi: 10.3354/cr021173

Wu, X., Washaya, P., Liu, L., Li, K., Shao, Y., Meng, L., et al. (2020). Rice yield estimation based on spaceborne SAR: A review from 1988 to 2018. *IEEE Access* 8, 157462–157469, 9179784. doi: 10.1109/ACCESS.2020.3020182

Wu, X., Xiao, Q., Wen, J., Liu, Q., Peng, J., and Li, X. (2014). Advances in uncertainty analysis for the validation of remote sensing products: Take leaf area index for example. *Natl. Remote Sens. Bull.* 18 (05), 1011–1023. doi: 10.11834/jrs.20143332

Xing, K., and Guo, H. (2006). Review on uncertainty analysis methods in environment model. *Environ. Sci. Technol.* 05), 112–115+121. doi: 10.19672/j.cnki.1003-6504.2006.05.045

Yagiz, A. K., Cakici, M., Aydogan, N., Omezli, S., Yerlikaya, B. A., Ayten, S., et al. (2020). Exploration of climate change effects on shifting potato seasons, yields and water use employing NASA and national long-term weather data. *Potato Res.* 63 (4), 565–577. doi: 10.1007/s11540-020-09457-x

Yang, S., Liu, J., Liang, J., Yang, C., QIN, Y., Xv, X., et al. (2017). Potato climate yield prediction model based on BP neural network. *J. Agric.* 7 (04), 29–33. doi: 10.11923/j.issn.2095-4050.cjas16080007

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. doi: 10.3389/fpls.2017.01111

Yibrah, G., Araya, B., and Amsalu, N. (2015). Performance of aquacrop model in simulating tuber yield of potato (Solanum tuberosum L.) under various water availability conditions in Mekelle Area, Northern Ethiopia. *J. Nat. Sci. Res.* 5 (5), 123–130.

Yokobori, J., Niwa, K., Sugiura, R., Noguchi, N., and Chiba, Y. (2004). "Variable management for uniform potato yield using remote sensing images with unmanned helicopter," in *Proceedings of the International Conference on Automation Technology for Off-road Equipment, ATOE 2004* (Kyoto: MSABE). 447–454.

Zaeen, A. A., Sharma, L., Jasim, A., Bali, S., Buzza, A., and Alyokhin, A. (2020). In-season potato yield prediction with active optical sensors. *Agrosyst. Geosci. Environ.* 3 (1), e20024. doi: 10.1002/agg2.20024

Zhang, X., Davidson, E. A., Mauzerall, D. L., Searchinger, T. D., Dumas, P., and Shen, Y. (2015). Managing nitrogen for sustainable development. *Nature* 528 (7580), 51–59. doi: 10.1038/nature15743

Zhou, Z., Plauborg, F., Liu, F., Kristensen, K., and Andersen, M. N. (2018). Yield and crop growth of table potato affected by different split-N fertigation regimes in sandy soil. *Eur. J. Agron.* 92, 41–50. doi: 10.1016/j.eja.2017.10.001

Zyromski, A., Szulczewski, W., Biniak-Pierog, M., and Zmuda, R. (2013). Application of the MoDrY model for the estimation of potato yielding. *Int. J. Plant Prod.* 7 (3), 505–516. doi: 10.22069/IJPP.2013.1116

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 September 2023
DOI 10.3389/fpls.2023.1208404

Check for updates

OPEN ACCESS

EDITED BY
Daobilige Su,
China Agricultural University, China

REVIEWED BY
Thorsten Kraska,
University of Bonn, Germany
Prashant Kaushik,
Yokohama Ueki, Japan

*CORRESPONDENCE
Zhenjiang Zhou
zhenjiangz@zju.edu.cn
Si Chen
chensi6@zju.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 19 April 2023
ACCEPTED 30 August 2023
PUBLISHED 18 September 2023

CITATION
He Y, Wang Z, Sun S, Zhu L, Li Y, Wang X, Shi J, Chen S, Qi D, Peng J and Zhou Z (2023) Using crop intercepted solar radiation and vegetation index to estimate dry matter yield of Choy Sum. *Front. Plant Sci.* 14:1208404. doi: 10.3389/fpls.2023.1208404

COPYRIGHT
© 2023 He, Wang, Sun, Zhu, Li, Wang, Shi, Chen, Qi, Peng 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.

Using crop intercepted solar radiation and vegetation index to estimate dry matter yield of Choy Sum

Yiyin He[1,2†], Zhao Wang[1,2†], Sashuang Sun[1,2], Lijun Zhu[1,2], Yu Li[3], Xiaoxiao Wang[3], Jiang Shi[4], Si Chen[5*], Dunchang Qi[1], Junxiang Peng[6] and Zhenjiang Zhou[1,2*]

[1]College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China, [2]Huanan Industrial Technology Research Institute of Zhejiang University, Guangzhou, China, [3]Agricultural Experiment Station, Zhejiang University, Hangzhou, China, [4]Hangzhou Academy of Agricultural Sciences, Hangzhou, China, [5]College of Water Resources and Environmental Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou, China, [6]Department of Agricultural Research for Northern Sweden, Swedish University of Agricultural Sciences, Umeå, Sweden

An accurate assessment of vegetable yield is essential for agricultural production and management. One approach to estimate yield with remote sensing is via vegetation indices, which are selected in a statistical and empirical approach, rather than a mechanistic way. This study aimed to estimate the dry matter of Choy Sum by both a causality-guided intercepted radiation-based model and a spectral reflectance-based model and compare their performance. Moreover, the effect of nitrogen (N) rates on the radiation use efficiency (*RUE*) of Choy Sum was also evaluated. A 2-year field experiment was conducted with different N rate treatments (0 kg/ha, 25 kg/ha, 50 kg/ha, 100 kg/ha, 150 kg/ha, and 200 kg/ha). At different growth stages, canopy spectra, photosynthetic active radiation, and canopy coverage were measured by RapidScan CS-45, light quantum sensor, and camera, respectively. The results reveal that exponential models best match the connection between dry matter and vegetation indices, with coefficients of determination (R^2) all below 0.80 for normalized difference red edge (NDRE), normalized difference vegetation index (NDVI), red edge ratio vegetation index (RERVI), and ratio vegetation index (RVI). In contrast, accumulated intercepted photosynthetic active radiation (*Aipar*) showed a significant linear correlation with the dry matter of Choy Sum, with root mean square error (*RMSE*) of 9.4 and R^2 values of 0.82, implying that the *Aipar*-based estimation model performed better than that of spectral-based ones. Moreover, the *RUE* of Choy Sum was significantly affected by the N rate, with 100 kg N/ha, 150 kg N/ha, and 200 kg N/ha having the highest *RUE* values. The study demonstrated the potential of *Aipar*-based models for precisely estimating the dry matter yield of vegetable crops and understanding the effect of N application on dry matter accumulation of Choy Sum.

KEYWORDS

Choy Sum, canopy coverage, radiation use efficiency, dry matter, N fertilization

1 Introduction

Choy Sum (*Brassica rapa* var. *parachinensis*) is one of the most productive and consumed vegetables in Asia (Kok et al., 1991; Zou et al., 2021). Accurate estimation of its dry matter yield is vital for the assessment of crop performance and decision-making during the growth season. Laboratory analysis of fresh or dry plants is conventionally performed to determine yield, but the process is time- and labor-intensive. Furthermore, plants are typically sampled on limited spots, which may not accurately represent dry matter yield across the complete region. Therefore, a quick and precise method for determining the dry matter yield of Choy Sum is required.

Due to the non-destructive, accurate, and timely access to crop information, the spectral-based remote sensing method has been used extensively in agriculture to estimate the crop agronomy parameters including nitrogen concentration, leaf area index, canopy coverage, and biomass (Mulla, 2013; Pereira et al., 2020). One typical approach is to use statistically selected vegetation indices to estimate crop dry matter of wheat (Erdle et al., 2011; Elsayed et al., 2021), rice (Gnyp et al., 2014), potato (Liu et al., 2022), etc. However, vegetation indices are prone to saturation at dense canopies, resulting in decreased accuracy of the estimation model (Wang et al., 2012; Gnyp et al., 2014). This is mainly due to the fact that spectral penetration at a specific wavelength is constrained on dense canopies, a mechanistic method that takes into account that plant growth is needed to address the problem of saturation. Thus, a method based on canopy intercepted solar radiation and its use efficiency is proposed in this study.

As plant dry matter (DM) is the production of photosynthetic activity, which uses sun radiation, it is determined by accumulated intercepted photosynthetic active radiation (*Aipar*) and the efficiency of using it (Leblon et al., 1991; Andersen et al., 1996). This can be expressed in the form of DM = *RUE* × *Aipar*, where *RUE* is the radiation use efficiency. *RUE* is a paramount parameter for crop growth models. Among all crop management factors, nitrogen (N) application rate can affect crop dry matter via its effects on *Aipar*, *RUE*, or both (Shah et al., 2004; Zhou et al., 2016). *RUE* is primarily controlled by crop net photosynthesis (Monteith, 1994), which in turn has a close dependency on leaf N concentration (Lawlor et al., 1989). Therefore, N fertilization affects *RUE* by influencing the photosynthesis rate (Muchow and Sinclair, 1994). Despite the fact that numerous studies have been conducted on the impact of N on *RUE* or *Aipar* (Chakwizira et al., 2018; Wang et al., 2020; Cafaro La Menza et al., 2022), very few of them have focused on the vegetable crop. Thus, the impact of N rates on the *RUE* of Choy Sum needs to be investigated. In addition, the potential of using the *Aipar*-based method to estimate the dry matter of vegetable crops is still unknown.

Therefore, the purposes of this study were to 1) explore the capacity of *Aipar* to estimate the dry matter yield of Choy Sum and 2) investigate the effect of N rates on the radiation use efficiency of Choy Sum.

2 Materials and methods

2.1 Field experiments

A 2-year field experiment was carried out from 2021 to 2022 at the Qiyuan farm site (30°100E, 119°480N), Hangzhou City, China. The experimental site has a typical subtropical monsoon climate condition and is characterized by abundant precipitation with mild seasonal temperature variation. The yearly mean annual temperature and total precipitation are 18.7°C and 1,930 mm, respectively. The predominant soil is characterized as loamy, having a pH of 7.9, total N content of 1.45 g/kg, hydrolytic N content of 199 mg/kg, and organic matter of 24.1 g/kg.

In 2021, Choy Sum was transplanted on April 15 and harvested on May 24. In 2022, Choy Sum was transplanted on April 1 and harvested on April 25. The average temperature was 22.9°C and the total precipitation was 147.0 mm in the growth season of 2021, while the average temperature was 18.1°C and the total precipitation was 121.7 mm in the growth season of 2022. Pumpkin and sweet potato were the previous crops in both seasons. Each plot had a size of 5 m × 5 m. The plant density was 0.15 m × 0.15 m and 0.3 m × 0.3 m in 2021 and 2022, respectively. Six treatments were included in each experiment differentiated by varying N rates of 0 kg/ha, 25 kg/ha, 50 kg/ha, 100 kg/ha, 150 kg/ha, and 200 kg/ha (defined as N_0, N_{25}, N_{50}, N_{100}, N_{150}, and N_{200}, respectively). The experimental design utilized in the field was a completely randomized block design, which consisted of three replications. Urea was used to apply N fertilizer at transplantation, while 30 kg/ha of phosphorus (P) and 90 kg/ha of potassium (K) were applied simultaneously. The administration of irrigation, pest control, and disease management were conducted in accordance with regional best practices.

2.2 Sample collection and data acquisition

A total of 90 and 54 samples were collected in 2021 and 2022, respectively (Table 1). Each plot had a sub-plot of 0.5 m × 0.7 m

TABLE 1 Field experimental design in 2021 and 2022.

Year	Transplanting date	Harvest date	Replicates	N rates (kg N/ha)	Sampling time (days after transplanting)
2021	April 15	May 24	3	0, 25, 50, 100,150, 200	15–21–24–27–31
2022	April 1	April 25	3	0, 25, 50, 100,150, 200	15–20–25

marked for sampling, and plants within the sub-plot area were collected and stored for subsequent analysis. Digital images and spectral measurements were taken prior to destructive sampling. All samples were desiccated in the oven at 105° C for 30 minutes and then at 75° C for 48 hours to measure dry matter, which refers to the dry weight of above-ground plant material per unit area. Subsequently, plant N concentration (*PNC*, %) of dried samples was determined by the Kjeldahl method (KDN-B, Shanghai Xinjia Electronic, Co., Ltd., Shanghai, China).

A hand-held field spectrometer (RapidScan CS-45; Holland Scientific, Lincoln, NE, USA) was used to measure canopy spectral reflectance at 670 nm (defined as Red), 730 nm (defined as RedEdge), and 780 nm (defined as near infrared (NIR)) wavelength and provides normalized difference red edge (NDRE) and normalized difference vegetation index (NDVI) (Aranguren et al., 2019). For each plot, a representative row was selected for spectral data measurement. The instrument was maintained approximately 1 m above the crop canopy during the scanning procedure. The four most commonly used vegetation indices named NDRE, NDVI, red edge ratio vegetation index (RERVI) (Erdle et al., 2011), and ratio vegetation index (RVI) (Zhou et al., 2017) were calculated as follows.

$$\mathrm{NDRE} = (\mathrm{NIR} - \mathrm{RedEdge}) / (\mathrm{NIR} + \mathrm{RedEdge}), \quad (1)$$

$$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{Red}) / (\mathrm{NIR} + \mathrm{Red}), \quad (2)$$

$$\mathrm{RERVI} = \mathrm{NIR} / \mathrm{RedEdge}, \quad (3)$$

$$\mathrm{RVI} = \mathrm{NIR} / \mathrm{Red}. \quad (4)$$

A digital camera (D5600, Nikon, Tokyo, Japan) was used to capture crop images, which were held above the crop canopy at a distance of 1 m. Images were converted to HSI three-color space (hue, saturation, and intensity channels) using Matlab (The MathWorks, Inc., Natick, MA, USA) programming. The threshold of H values was used to generate binary pictures. Those pixels with H values between 0.2 and 0.4 were given the value 1 (green vegetation), while other pixels without vegetation were put to 0 (soil background). Canopy coverage was calculated as the proportion of the number of pixels with a value of 1 to the whole pixel number in the binary image.

2.3 Radiation interception measurement by light quantum sensor

Photosynthetic active radiation was measured by a light quantum sensor (model Li-190R, Li-Cor Inc., Lincoln, NE, USA), which was installed near the experimental site. The radiation sensor was connected to a data logger (model CR3000 Series), which was used to record incident photosynthetic active radiation every 10 minutes. Daily intercepted photosynthetic active par (*Ipar*) and accumulated *Ipar* of plant canopy were calculated as follows:

$$Ipar = I_0 Fipar, \quad (5)$$

$$Aipar = \sum Ipar\, dt, \quad (6)$$

where I_0 is photosynthetic active radiation retrieved directly from the sensor. Fipar is the fraction of photosynthetic active radiation, which was substituted by observed canopy coverage in this study. *Aipar* is accumulated intercepted photosynthetic active radiation, with t as the growth period under consideration.

2.4 Statistical analysis

The coefficient of determination (R^2) was utilized to assess the general fit of the regression equation. Moreover, root mean square error (*RMSE*) indicates the performance of regression models. They were calculated as follows:

$$R^2 = \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(P_i - \bar{P})^2}, \quad (7)$$

$$RMSE = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N}(P_i - O_i)^2}, \quad (8)$$

where N is the number of observations, P_i is the observed value, $\bar{P}$ is the mean of observed values, and O_i is the predicted value.

The Tukey–Kramer method was applied to test the significant differences of *PNC*, DM, and *RUE* under different N treatments. The nominal alpha value of 0.05 was applied to determine R^2 and significance levels. In this study, OriginPro 2023 (OriginLab Corp., Northampton, MA, USA) was used to draw all graphs, and SPSS 14.0 software (SPSS Inc., Chicago, IL, USA) was used to perform statistical analysis.

3 Results

3.1 Effect of N rates on plant N concentration and dry matter

Table 2 displays the variation in PNC at different growth stages. In 2021, the PNC ranged from 3.02% to 5.71%, while in 2022, it ranged from 3.14% to 4.99%. In both seasons, the *PNC* of Choy Sum declined with the progress of crop growth. In 2021, the *PNC* of N_0 and N_{25} treatments was significantly lower than that of the other treatments (N_{100}, N_{150}, and N_{200}). The results in 2022 showed a similar pattern to 2021, although *PNC* of N_{25} only showed a significantly lower value until 25 days after transplanting.

The dry matter of Choy Sum varied from 4.30 g/m^2 to 93.12 g/m^2 in 2021 and varied from 3.87 g/m^2 to 20.57 g/m^2 in 2022 (Table 3). In 2021, a notable difference was detected in the dry matter between low N treatments (N_0 and N_{25}) and high N treatments (N_{50}, N_{100}, N_{150}, and N_{200}) at all growth stages. In 2022, there was also an obvious difference in the amount of dry matter between high N treatments (N_{100}, N_{150}, and N_{200}) and low N treatments (N_0 and N_{25}) at 16, 20, and 25 days after transplanting.

10.3389/fpls.2023.1208404

TABLE 2 The variation in plant N concentration (*PNC*, %) of Choy Sum under different N rates in 2021 and 2022.

Year	N rates (kg/ha)	Days after transplanting				
		14	20	23	26	33
2021	0	3.50 ± 0.11 c	3.57 ± 0.30 b	3.80 ± 0.12 b	3.02 ± 0.17 b	3.12 ± 0.35 b
	25	4.75 ± 0.05 b	4.50 ± 1.28 ab	3.92 ± 0.10 b	3.76 ± 0.88 ab	3.27 ± 0.42 b
	50	5.71 ± 0.12 a	5.12 ± 0.56 a	5.30 ± 0.17 a	3.99 ± 0.62 ab	4.08 ± 0.79 ab
	100	5.58 ± 0.09 a	4.97 ± 0.75 a	5.46 ± 0.36 a	4.79 ± 0.91 a	4.43 ± 0.29 ab
	150	5.51 ± 0.12 a	5.15 ± 0.37 a	5.23 ± 0.33 a	4.67 ± 0.37 a	4.86 ± 0.74 a
	200	5.70 ± 0.20 a	4.39 ± 0.74 a	5.64 ± 0.39 a	4.44 ± 1.02 a	5.31 ± 0.33 a
Year	N rates (kg/ha)	Days after transplanting				
		16	20	–	25	–
2022	0	3.97 ± 0.47 b	4.22 ± 0.09 b	–	3.70 ± 0.71 a	–
	25	4.94 ± 0.19 a	4.68 ± 0.28 ab	–	3.27 ± 0.21 b	–
	50	4.77 ± 0.06 a	4.52 ± 0.09 ab	–	3.14 ± 0.35 b	–
	100	4.99 ± 0.09 a	4.93 ± 0.22 a	–	3.43 ± 0.37 a	–
	150	4.84 ± 0.16 a	3.55 ± 0.45 c	–	3.57 ± 0.14 a	–
	200	5.11 ± 0.20 a	4.32 ± 0.23 ab	–	3.70 ± 0.49 a	–

Values with a distinct letter after them are significant at the 5% level using Tukey–Kramer method.

3.2 Estimation of dry matter using vegetation indices and *Aipar*

Dry matter was estimated by two different models in this study, i.e., spectral vegetation index-based and *Aipar*-based models. Dry matter and vegetation indices were observed to be closely related. Their relationship can be best fitted by exponential models (Figure 1), while the performances of linear models were relatively worse than those of the exponential ones. The coefficient of determination of all exponential models had considerable R^2 of 0.74, 0.76, 0.75, and 0.76 for NDRE,

TABLE 3 The variation in plant dry matter (g/m^2) of Choy Sum under different N rates in 2021 and 2022.

Year	N rates (kg/ha)	Days after transplanting				
		14	20	23	26	33
2021	0	4.30 ± 0.96 b	11.79 ± 3.51 b	14.60 ± 3.02 b	19.15 ± 1.78 c	28.59 ± 1.75 c
	25	6.30 ± 0.79 ab	19.96 ± 2.21 ab	21.86 ± 2.81 b	38.80 ± 6.10 b	45.01 ± 13.28 bc
	50	7.33 ± 0.90 ab	30.24 ± 8.71 a	36.33 ± 5.71 a	52.36 ± 8.81 ab	65.90 ± 18.81 abc
	100	8.00 ± 1.31 a	33.32 ± 3.66 a	38.88 ± 5.01 a	66.14 ± 6.86 a	68.97 ± 8.22 ab
	150	6.80 ± 1.74 ab	26.76 ± 8.06 ab	40.75 ± 3.91 a	53.44 ± 11.36 ab	74.67 ± 17.33 ab
	200	7.03 ± 0.50 ab	33.45 ± 5.63 a	39.01 ± 7.26 a	59.03 ± 3.08 a	93.12 ± 15.74 a
Year	N rates (kg/ha)	Days after transplanting				
		16	20	–	25	–
2022	0	3.87 ± 0.23 d	5.70 ± 0.95 c	–	7.57 ± 0.40 b	–
	25	6.47 ± 2.02 bc	9.30 ± 2.11 bc	–	14.63 ± 2.11 ab	–
	50	5.67 ± 0.98 c	10.90 ± 2.86 ab	–	16.26 ± 2.21 ab	–
	100	6.57 ± 0.25 bc	10.73 ± 2.14 ab	–	19.27 ± 4.48 a	–
	150	7.63 ± 1.82 ab	11.80 ± 3.50 ab	–	16.90 ± 1.54 a	–
	200	8.00 ± 2.31 a	13.10 ± 2.65 a	–	20.23 ± 7.95 a	–

Values with a distinct letter after them are significant at the 5% level using Tukey–Kramer method.

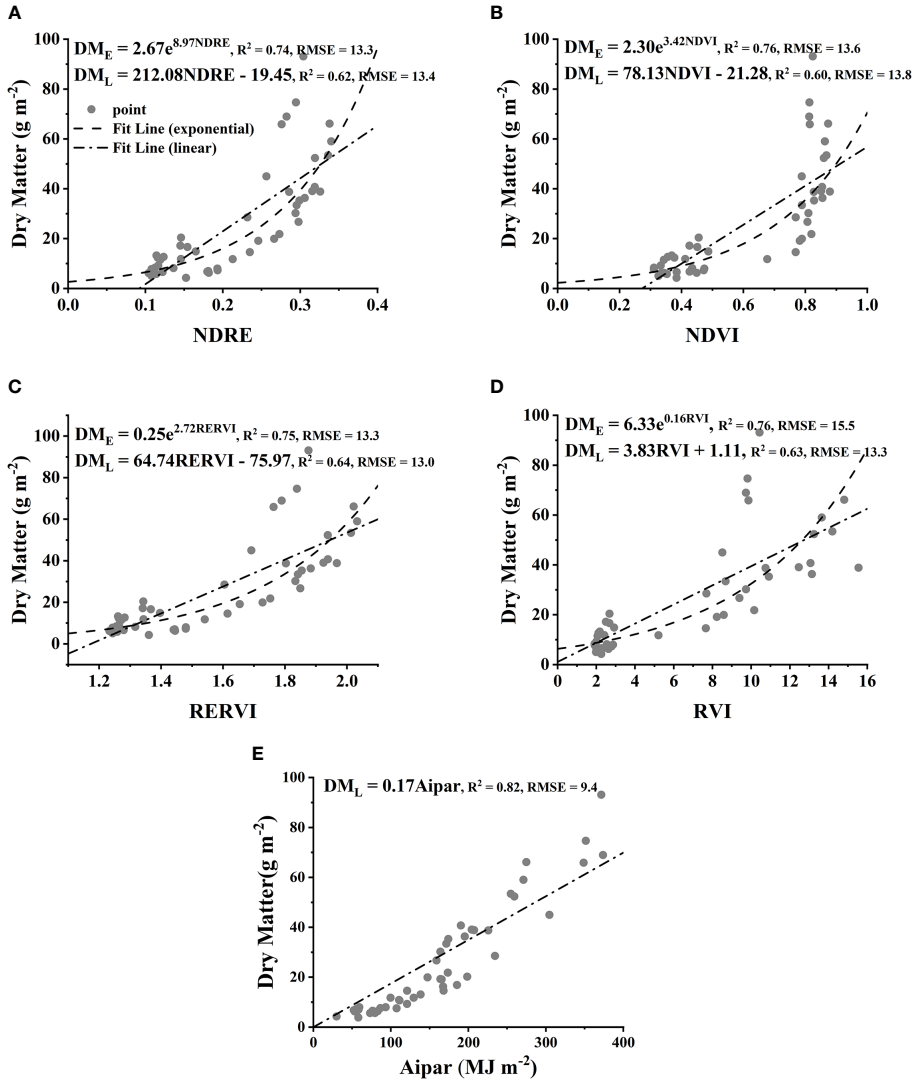


FIGURE 1
The relation between dry matter (g/m^2) and NDRE **(A)**, NDVI **(B)**, RERVI **(C)**, RVI **(D)**, and *Aipar* **(E)** across 2021 and 2022 seasons. The number of samples is 48, and each value is the mean of three replicates. E and L stand for exponential (E) and linear (L) models, respectively. All fitted relationship functions were tested to be highly significant. NDRE, normalized difference red edge; NDVI, normalized difference vegetation index; RERVI, red edge ratio vegetation index; RVI, ratio vegetation index.

NDVI, RERVI, and RVI, respectively. In contrast, the correlation between dry matter and *Aipar* was significantly linear (R^2 = 0.82).

3.3 The relationship between dry matter and accumulated intercepted photosynthetic active radiation

A linear relationship between DM and *Aipar* was detected across different N rates in both 2021 and 2022 (Figure 2; Table 4). The linear relationship was derived from the equation of DM = *RUE* × *Aipar*, where *RUE* was treated as the slope of the linear function, and a comparison of *RUE* of different N treatments was also conducted. In both 2021 and 2022, the *RUE* of N_0 and N_{250} was substantially lower than the *RUE* of other treatments. N_{200} obtained the highest *RUE* of 0.21 in 2021, while N_{100} obtained the highest *RUE* of 0.11 in 2022. Significant differences in *RUE* between N_{100}, N_{150}, and N_{200} were not detected. The R^2 values of all regression models were higher than 0.95.

4 Discussion

4.1 Estimation of dry matter by spectral- and *Aipar*-based models

In this study, both *Aipar*-based and spectral reflectance-based models were adopted to estimate the dry matter yield of Choy Sum, and the performances of these two types of models were comprehensively compared. Four commonly used vegetation indices were extracted from the active field spectrometer to build estimation models of Choy Sum yield.

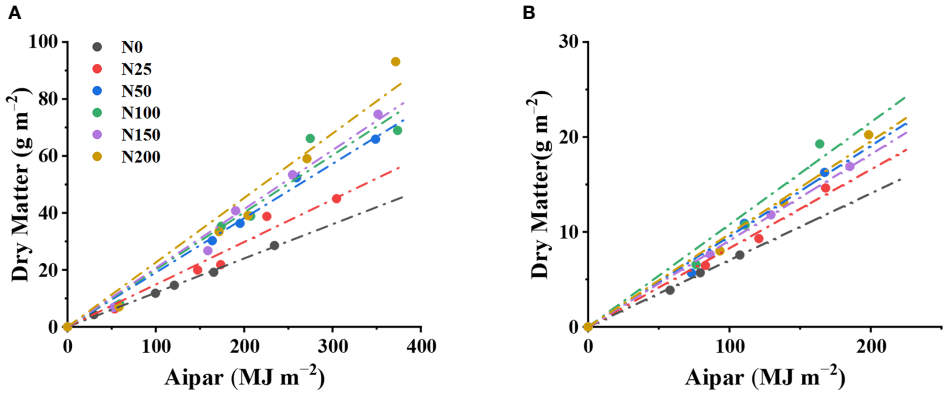


FIGURE 2
The linear correlation between dry matter and accumulated intercepted photosynthetic active radiation (*Aipar*) across different N levels in 2021 **(A)** and 2022 **(B)**.

The relationships between vegetation indices and dry matter yield were best fitted by exponential models, which outperformed the linear ones, suggesting saturation problems occurred when dry matter approached a certain high level (Figure 1). This phenomenon agreed with the findings of other studies, where saturation often occurred for high biomass or LAI values in the late growth stage. For instance, it occurs when biomass reached 3 t/ha or NDVI approached 0.95 for rice (Gnyp et al., 2014), or NDVI reached 0.8 for wheat (Erdle et al., 2011), while it occurred when NDVI was approximately 0.8 for Choy Sum in this study. Nonetheless, dry matter can still be satisfactorily estimated by vegetation index-based models over the two seasons (Figure 1).

The saturation issue could be well addressed using the *Aipar*-based model, as *Aipar* increased proportionally to dry matter, even when the canopy approached a high level (Figure 1); this is due to the fact that the parameter of *Aipar* integrates not only canopy coverage but also actual sun radiation. The constraint of the study is that *Fipar* was not directly measured; alternatively, canopy coverage was used as a substitution for *Fipar*. This could lead to a deviation of the true *RUE* values, although Haverkort reported that canopy coverage could be used as an approximation of *Fipar* on potato crops (Haverkort et al., 1991; Chakwizira et al., 2015). Many studies calculated *Fipar* from remote sensing or leaf area index since there exist close relationships and turned out to be relatively accurate (Zhou et al., 2017; Chakwizira et al., 2018; Peng et al., 2021). Nonetheless, the *Fipar* estimation method in this study has rapid access at a lower cost, despite some precision sacrificed.

Furthermore, since only 2 years' data in one site were contained in this study, and there existed many factors of environmental variation, the specific model may not be so universal over a large scale of time and space but would be a similar trend.

TABLE 4 Radiation use efficiency (*RUE*) of Choy Sum in 2021 and 2022.

Year	N rates (kg/ha)	*RUE*
2021	0	0.12 d
	25	0.14 c
	50	0.19 b
	100	0.20 ab
	150	0.20 ab
	200	0.21 a
2022	0	0.07 b
	25	0.08 ab
	50	0.10 a
	100	0.11 a
	150	0.09 a
	200	0.10 a

For a specific year, values with a distinct letter after them are significant at the 5% level using Tukey–Kramer method.

4.2 Effect of N rates on *RUE*

In both 2021 and 2022, significantly lower *RUE* was detected in non-nitrogen (N_0) and low-nitrogen treatments (N_{25}) than in the high N treatments of N_{50}, N_{100}, N_{150}, and N_{200}, among which *RUE* showed no difference. This indicated that the accumulation of dry matter was caused by the increase of not only *Aipar* but also *RUE*. When the N rate was above 50 kg N/ha, *RUE* will not be a restriction factor for DM accumulation, implying DM increase was entirely caused by *Aipar* increase resulting from N application. Crops have different strategies to cope with N deficiency; some crops maintain chlorophyll content to keep constant radiation use efficiency, i.e., with *RUE* unaffected. The research in potatoes showed that *RUE* remained unchanged at different N levels (Millard and Marshall, 1986; Bangemann et al., 2014), as potatoes tend to reduce the leaf area instead of photosynthesis efficiency per unit leaf area when N is insufficient (Vos and van der Putten, 1998); a similar phenomenon is also demonstrated on beet crops (Chakwizira et al., 2018). In contrast, other studies showed that *RUE* increased significantly with N application and leveled off when the N rate reached to medium level on corn (Muchow and Davis, 1988; Kaur et al., 2012; Kar and

Kumar, 2016). In the case of Choy Sum, increased leaf N content was mainly used for thylakoid accumulation in photosynthetic cells and synthesis of carboxylases, rather than increasing leaf area to capture light (Khairun et al., 2016), corroborating well with the discovery of this research. In conclusion, N application significantly increases the *RUE* of Choy Sum when the N rate is below 50 kg N/ha. However, beyond this threshold, further increases in the N rate did not affect *RUE*.

5 Conclusions

This study evaluated the performance of *Aipar* and spectral-based methods for estimating the dry matter yield of Choy Sum over two growing seasons. The results showed that the relationships between DM of Choy Sum and vegetation indices were best fitted by exponential models due to the saturation problems in the case of high biomass, while *Aipar* had a significant linear relationship with DM ($R^2 = 0.82$) even when the canopy approached a high level. This suggests that *Aipar* could be used as a better candidate for dry matter yield estimation. With regard to the effect of N application on the *RUE* of Choy Sum, the *RUE* of crops applied with low N (0–25 kg N/ha) was significantly lower than crops applied with high N (50–200 kg N/ha). Therefore, when applied with a N rate above 50 kg N/ha, dry matter yield production of Choy Sum will not be constrained by reduced *RUE*.

Data availability statement

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

Author contributions

Conceptualization, ZW, YH, and ZZ; methodology, ZW; software, ZW and JP; formal analysis, SS and LZ; investigation, JS, SS, SC, DQ, and ZZ; data curation, ZZ; writing—original draft, ZW and YH; writing—review and editing, SC, JS, YL, XW, and ZZ; supervision, ZZ; funding acquisition, ZZ. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the Science and Technology Department of Guangdong Province (grant 2019B020216001), the National Key Research and Development Program of China (grant 2019YFE0125500-02), and the National Natural Science Foundation of China (grant 52109069).

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.2023.1208404/full#supplementary-material

References

Andersen, M., Heidmann, T., and Plauborg, F. (1996). The effects of drought and nitrogen on light interception, growth and yield of winter oilseed rape. *Acta Agriculturae Scandinavica B-Plant Soil Sci.* 46 (1), 55–67. doi: 10.1080/09064719609410947

Aranguren, M., Castellon, A., and Aizpurua, A. (2019). Crop sensor-based in-season nitrogen management of wheat with manure application. *Remote Sens.* 11 (9). doi: 10.3390/rs11091094

Bangemann, L.-W., Sieling, K., and Kage, H. (2014). The effect of nitrogen and late blight on crop growth, solar radiation interception and yield of two potato cultivars. *Field Crops Res.* 155, 56–66. doi: 10.1016/j.fcr.2013.09.022

Cafaro La Menza, N., Arkebauer, T. J., Lindquist, J. L., Monzon, J. P., Knops, J. M. H., Graef, G., et al. (2022). Decoupling between leaf nitrogen and radiation use efficiency in vegetative and early reproductive stages in high-yielding soybean. *J. Exp. Bot.* 74 (1), 352–363. doi: 10.1093/jxb/erac408

Chakwizira, E., Meenken, E. D., George, M. J., and Fletcher, A. L. (2015). Can we use photography to estimate radiation interception by a crop canopy? *Plant Biol.* 17 (2), 574–582. doi: 10.1111/plb.12240

Chakwizira, E., Teixeira, E., Meenken, E., Michel, A. J., and Maley, S. (2018). Radiation use efficiency and biomass partitioning to storage roots in fodder beet crops. *Eur. J. Agron.* 92, 63–71. doi: 10.1016/j.eja.2017.10.002

Elsayed, S., Barmeier, G., and Schmidhalter, U. (2021). Passive reflectance sensing and digital image analysis allows for assessing the biomass and nitrogen status of wheat in early and late tillering stages. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.670027

Erdle, K., Mistele, B., and Schmidhalter, U. (2011). Comparison of active and passive spectral sensors in discriminating biomass parameters and nitrogen status in wheat cultivars. *Field Crops Res.* 124 (1), 74–84. doi: 10.1016/j.fcr.2011.06.007

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

Haverkort, A., Uenk, D., Veroude, H., and Van de Waart, M. (1991). Relationships between ground cover, intercepted solar radiation, leaf area index and infrared reflectance of potato crops. *Potato Res.* 34 (1), 113–121. doi: 10.1007/Bf02358105

Kar, G., and Kumar, A. (2016). Radiation utilization efficiency and surface energy exchange of winter maize (Zea mays L.) under different irrigation regimes. *J. Agrometeorol.* 18 (2), 190–195. doi: 10.54386/jam.v18i2.934

Kaur, A., Bedi, S., Gill, G., and Kumar, M. (2012). Effect of nitrogen fertilizers on radiation use efficiency, Crop growth and yield in some maize (Zea mays L) genotypes. *Maydica* 57 (1), 75–82.

Khairun, N., Teh, C., and Hawa, Z. (2016). Simultaneous effects of water and nitrogen stress on the vegetative and yield parameters of choy sum (Brassica chinensis var. parachinensis). *Pertanika J. Trop. Agric. Sci.* 39 (1), 101–116.

Kok, P. T., Keng, H., Avadhani, P., and Foo, J. T. (1991). *Guide to common vegetables* (Singapore Science Centre).

Lawlor, D., Kontturi, M., and Young, A. (1989). Photosynthesis by flag leaves of wheat in relation to protein, ribulose bis phosphate carboxylase activity and nitrogen supply. *J. Exp. Bot.* 40 (1), 43–52. doi: 10.1093/jxb/40.1.43

Leblon, B., Guerif, M., and Baret, F. (1991). The use of remotely sensed data in estimation of PAR use efficiency and biomass production of flooded rice. *Remote Sens. Environ.* 38 (2), 147–158. doi: 10.1016/0034-4257(91)90076-I

Liu, Y., Feng, H. K., Yue, J. B., Fan, Y. G., Jin, X. L., Zhao, Y., et al. (2022). Estimation of potato above-ground biomass using UAV-based hyperspectral images and machine-learning regression. *Remote Sens.* 14 (21). doi: 10.3390/rs14215449

Millard, P., and Marshall, B. (1986). Growth, nitrogen uptake and partitioning within the potato (Solatium tuberosum L.) crop, in relation to nitrogen application. *J. Agric. Sci.* 107 (2), 421–429. doi: 10.1017/S0021859600087220

Monteith, J. L. (1994). Validity of the correlation between intercepted radiation and biomass. *Agric. For. meteorol.* 68 (3-4), 213–220. doi: 10.1016/0168-1923(94)90037-X

Muchow, R., and Davis, R. (1988). Effect of nitrogen supply on the comparative productivity of maize and sorghum in a semi-arid tropical environment II. Radiation interception and biomass accumulation. *Field Crops Res.* 18 (1), 17–30. doi: 10.1016/0378-4290(88)90056-1

Muchow, R., and Sinclair, T. (1994). Nitrogen response of leaf photosynthesis and canopy radiation use efficiency in field-grown maize and sorghum. *Crop Sci.* 34 (3), 721–727. doi: 10.2135/cropsci1994.0011183X003400030022x

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosyst. Eng.* 114 (4), 358–371. doi: 10.1016/j.biosystemseng.2012.08.009

Peng, J.X., Manevski, K., Korup, K., Larsen, R., and Andersen, M. N. (2021). Random forest regression results in accurate assessment of potato nitrogen status based on multispectral data from different platforms and the critical concentration approach. *Field Crops Res.* 268. doi: 10.1016/j.fcr.2021.108158

Pereira, L. S., Paredes, P., Melton, F., Johnson, L., Wang, T., Lopez-Urrea, R., et al. (2020). Prediction of crop coefficients from fraction of ground cover and height. Background and validation using ground and remote sensing data. *Agric. Water Manage.* 241. doi: 10.1016/j.agwat.2020.106197

Shah, S., McKenzie, B., Gaunt, R., Marshall, J. W., and Frampton, C. (2004). Effect of early blight (Alternaria solani) and different nitrogen inputs on radiation interception, radiation use efficiency, and total dry matter production in potatoes (Solanum tuberosum) grown in Canterbury, New Zealand. *New Z. J. Crop Hortic. Sci.* 32 (3), 263–272. doi: 10.1080/01140671.2004.9514305

Vos, J., and van der Putten, P. E. L. (1998). Effect of nitrogen supply on leaf growth, leaf nitrogen economy and photosynthetic capacity in potato. *Field Crops Res.* 59 (1), 63–72. doi: 10.1016/S0378-4290(98)00107-5

Wang, X., Christensen, S., Svensgaard, J., Jensen, S. M., and Liu, F. (2020). The effects of cultivar, nitrogen supply and soil type on radiation use efficiency and harvest index in spring wheat. *Agronomy* 10 (9). doi: 10.3390/agronomy10091391

Wang, W., Yao, X., Yao, X., Tian, Y., Liu, X., 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

Zhou, Z. J., Andersen, M. N., and Plauborg, F. (2016). Radiation interception and radiation use efficiency of potato affected by different N fertigation and irrigation regimes. *Eur. J. Agron.* 81, 129–137. doi: 10.1016/j.eja.2016.09.007

Zhou, Z. J., Plauborg, F., Thomsen, A. G., and Andersen, M. N. (2017). A RVI/LAI-reference curve to detect N stress and guide N fertigation using combined information from spectral reflectance and leaf area measurements in potato. *Eur. J. Agron.* 87, 1–7. doi: 10.1016/j.eja.2017.04.002

Zou, L., Tan, W. K., Du, Y., Lee, H. W., Liang, X., Lei, J., et al. (2021). Nutritional metabolites in Brassica rapa subsp. chinensis var. parachinensis (choy sum) at three different growth stages: Microgreen, seedling and adult plant. *Food Chem.* 357. doi: 10.1016/j.foodchem.2021.129535

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 September 2023
DOI 10.3389/fpls.2023.1233892

Check for updates

OPEN ACCESS

EDITED BY
Jiangang Liu,
Chinese Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Robert John French,
Department of Primary Industries and Regional Development of Western Australia (DPIRD), Australia
Sayantan Sarkar,
Texas A&M AgriLife Research, United States

*CORRESPONDENCE
Arron H. Carter
ahcarter@wsu.edu

RECEIVED 02 June 2023
ACCEPTED 29 August 2023
PUBLISHED 18 September 2023

CITATION
Herr AW and Carter AH (2023) Remote sensing continuity: a comparison of HTP platforms and potential challenges with field applications.
Front. Plant Sci. 14:1233892.
doi: 10.3389/fpls.2023.1233892

COPYRIGHT
© 2023 Herr and Carter. 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.

Remote sensing continuity: a comparison of HTP platforms and potential challenges with field applications

Andrew W. Herr and Arron H. Carter*

Department of Crop and Soil Sciences, Washington State University, Pullman, WA, United States

In an era of climate change and increased environmental variability, breeders are looking for tools to maintain and increase genetic gain and overall efficiency. In recent years the field of high throughput phenotyping (HTP) has received increased attention as an option to meet this need. There are many platform options in HTP, but ground-based handheld and remote aerial systems are two popular options. While many HTP setups have similar specifications, it is not always clear if data from different systems can be treated interchangeably. In this research, we evaluated two handheld radiometer platforms, Cropscan MSR16R and Spectra Vista Corp (SVC) HR-1024i, as well as a UAS-based system with a Sentera Quad Multispectral Sensor. Each handheld radiometer was used for two years simultaneously with the unoccupied aircraft systems (UAS) in collecting winter wheat breeding trials between 2018-2021. Spectral reflectance indices (SRI) were calculated for each system. SRI heritability and correlation were analyzed in evaluating the platform and SRI usability for breeding applications. Correlations of SRIs were low against UAS SRI and grain yield while using the Cropscan system in 2018 and 2019. Dissimilarly, the SVC system in 2020 and 2021 produced moderate correlations across UAS SRI and grain yield. UAS SRI were consistently more heritable, with broad-sense heritability ranging from 0.58 to 0.80. Data standardization and collection windows are important to consider in ensuring reliable data. Furthermore, practical aspects and best practices for these HTP platforms, relative to applied breeding applications, are highlighted and discussed. The findings of this study can be a framework to build upon when considering the implementation of HTP technology in an applied breeding program.

KEYWORDS

high throughput phenotyping, plant breeding, spectral reflectance indices, unoccupied aircraft system - UAS, field based phenotyping, remote sensing

1 Introduction

Severe weather and climate change are creating new challenges in maintaining and improving global food production. Plant breeding is an important tool in adapting to these difficulties (Brown et al., 2015). However, the plant breeding process is not immune to extreme or unpredictable environmental conditions, impacting selection efficiency and genetic gain (Xiong et al., 2022). Despite steady increases in genetic gain, global cereal crop demand is projected to surpass production by 2050 (Ray et al., 2013). A practical method for increasing crop production is through the release of new cultivars by increasing genetic gain. On average, it takes 7-12 years to release a new winter wheat cultivar (Carver, 2009). Thus, it is imperative to look at new methods that increase genetic gain, decrease cycle time, and improve grain yield in an era of climate change and extreme environmental variables (Xiong et al., 2022). Recent advancements in genomic technologies have provided breeders with large amounts of data to utilize genomic and marker-assisted selections. However, genomic data has limited use without the backing of phenotypic data, creating a new bottleneck in the industry, and limiting cultivar development efficiency (Mir et al., 2019). One proposed solution to this limitation is implementing high throughput phenotyping (HTP) methods associated with established breeding strategies (Reynolds et al., 2020).

The development of HTP results from advancements in imaging sensors, image processing technology, and an understanding of secondary phenotypic traits (Pauli et al., 2016; Mir et al., 2019). Despite these advancements, unoccupied aircraft systems (UAS) technology can be fastidious and resource intensive for simple plant breeding applications. Most HTP strategies require the purchase of expensive specialized equipment and tedious data standardization and processing pipelines. There is a continued need to identify and adapt HTP technology to better aid the breeder while maintaining cost-effectiveness (Reynolds et al., 2020).

There are three primary options in field-based HTP. Satellites allow for the high throughput collection of field-scale images yet are limited by image frequencies and resolution, critical factors in plot-level research applications. Alternatively, ground-based and handheld systems provide high resolution imaging across a wide range of frequencies but can be challenging to handle and capture larger-scale, multi-plot images quickly. UAS provide a "goldilocks" ratio of utility, temporal frequency, and spatial resolution (Araus and Cairns, 2014; Song et al., 2021). Rotocopters are a versatile platform that allows for high throughput, high resolution image capture. Due to power usage and battery capacity, the limitations of the platform arise in payload capacity and flight time (Sankaran et al., 2015b; Xie and Yang, 2020).

The spectral reflectance data collected from handheld radiometers and UAS cameras has minimal uses in its raw form. Spectral reflectance indices (SRI) are used to evaluate target features and remove image noise creating a practical, standardized trait value (Myneni et al., 1995; Xue and Su, 2017). Vegetation indices are developed by evaluating the reflectance value of the plant canopy at specific light bands associated with photosynthetic mechanisms. Normalized Difference Vegetation Index (NDVI) is a prevalent index used to evaluate plant health by evaluating contrast in the maximum absorption of red in the leaf through chlorophyll pigmentation and the maximum reflectance of near-infrared due to leaf cellular structure (Rouse, 1974). Normalized Difference Red-edge Index (NDRE), another popular standard vegetation index, works similarly to NDVI but replaces red with red-edge absorption relative to NIR (Gitelson and Merzlyak, 1996). The vegetation index used depends on the crop, growth stage, and target trait. These factors influence reflectance values and relative index effectiveness (Wientjes et al., 2017; Lozada et al., 2020; Herr et al., 2023). Vegetation indices have many applications in capturing routine trait estimates like plot quality, biotic, and abiotic stress (Sankaran et al., 2015a; Guo et al., 2021; Sarkar et al., 2022; Sapkota et al., 2023), as well as previously infeasible traits like chlorophyll content and nitrogen content (Xie and Yang, 2020; Yin et al., 2022). Unlike vegetation indices, water indices such as Normalized Water Index (NWI) use infrared range reflectance to evaluate stomatal conductance and overall photosynthetic efficiency (Babar et al., 2006). Water indices can evaluate and predict relative water content, leaf osmotic potential, stomatal conductance, and canopy temperature(Gutierrez et al., 2010; Bal et al., 2021; Visitacion et al., 2022).

For most breeders working with cereal crops, grain yield is the critical trait of interest and an ideal gauge of overall biological and economic performance. Grain yield in wheat is highly quantitative and can make selection efficiency difficult (Reynolds et al., 2012). It is well established that NDVI and other vegetation indices like NDRE and NWI, through high throughput multispectral imaging, correlate with cereal crop grain yields (Geipel et al., 2014; Gracia-Romero et al., 2017; Lozada et al., 2020). It has also been shown that SRI data can be utilized to improve tools like genomic selection for grain yield. Thus, grain yield is an appealing trait for a breeder to focus on when implementing HTP approaches (Reynolds et al., 2020; Montesinos López et al., 2022; Herr et al., 2023).

SRI heritability and correlation to grain yield are leading indicators of SRI and platform utility. A strong repeatable relationship to grain yield can determine data quality and efficiency of selection in large-scale field-based breeding applications. While the sensors evaluated in this study collected the same SRIs, each has a different manufacturer, sensor type, and processing pipeline. Sensor differences create the potential for variances in data value and quality. These variances are compared and discussed. Finally, the practical aspects of the platforms are compared for their potential cost relative to the benefit they could provide (i.e., the improved resolution of ground collected data is not worth the extra logistics required in data collection).

With the growing interest in utilizing high throughput phenotyping technology in plant breeding, this study aimed to compare the SRI data collected from breeding trials between ground systems and UAS and determine if the use of ground-based handheld systems provides an increased resolution and data quality that justify the negative aspects of the platform, like collection time, data noise, and secondary applications.

2 Materials and methods

2.1 Study population

The Washington State University (WSU) winter wheat breeding program has collected multispectral data with three different phenotyping systems as indicated in Figure 1: A handheld multispectral radiometer, the Cropscan MSR16R (CROPSCAN Inc., Rochester, MN, USA), a handheld full-range hyperspectral spectro-radiometer, the Spectra Vista Corporation (SVC) HR-1024i (Spectra Vista Corporation, Poughkeepsie, NY, USA), and a UAS-based system, a Sentera Quad Multispectral Sensor (Sentera Inc., Minneapolis, MN, USA) mounted on a DJI Inspire 1 rotor copter platform.

Data for all platforms was collected at anthesis due to its established relationship with grain yield (Duan et al., 2017; Lozada et al., 2020). Each population evaluated was sampled on the same day by both the UAS and compared handheld system. Handheld data was collected within a six-hour window of solar noon, which was the typical time required to collect data given the number of plots in the trials. UAS data was collected within a four-hour window of solar noon. In these trials, UAS data was often collected at or near solar noon to try and be in the middle of the handheld data collection timeframe, and flights often took 20 minutes. The UAS mounted with the Sentera camera flew a programmed route at an altitude of 45 m, with an 85% longitudinal and lateral overlap of georeferenced images. All data was collected on days with clear skies to limit variability in solar radiation. All trials were grown in Pullman, WA, as shown in Figure 1, and include:

- A genetically diverse Quality Association Mapping (QAM) panel;
- Unreplicated single plot yield trials of soft white and hard red winter wheat;
- Replicated alpha-lattice preliminary yield trials of soft white and hard red winter wheat;
- Replicated alpha-lattice advanced yield trials of soft white and hard red winter wheat.

Table 1 outlines the study populations' characteristics, including year, trial design, total number of unique entries, number of total plot observations, and HTP data type collected. Plots were planted using a double-disc 8-row small plot planter at a seed density of 250 seed per square meter. Total plot size was 1.5 meter wide by 3.5 meter long. Ground was prepared by grower cooperators using minimum-tillage techniques and practices customary of the region. Grain yield data were collected at all locations with a Zurn 150 harvester (Zurn Harvesting GmbH & Co. KG, Waldenburg, Germany). Weather data for each year can be found at https://weather.wsu.edu/ for the Pullman, WA location. Single environment adjusted means were calculated for all



FIGURE 1
The imaging systems used in this study were **(A)** a DJI Inspire 1 UAS with a Sentera Quad Multispectral Sensor, **(B)** Cropscan MSR16R, and **(C)** SVC HR-1024i. **(D)** highlights the study location.

TABLE 1 Study populations for HTP platform comparison.

Trial	Year	Design	Replication	Number of Entries	Total Observations	UAS	Cropscan	SVC
Single Plot	2018	Augmented Design with repeating checks	1	1438	1503	X	X	
Preliminary	2018	Alpha-Lattice	3	168	504	X	X	
QAM	2018	Augmented Design with repeating checks	1	480	528	X	X	
QAM	2019	Augmented Design with repeating checks	1	480	528	X	X	
Single Plot	2020	Augmented Design with repeating checks	1	178	195	X		X
Advanced	2020	Alpha-Lattice	3	48	144	X		X
Single Plot	2021	Augmented Design with repeating checks	1	213	227	X		X
Preliminary	2021	Alpha-Lattice	3	54	162	X		X
Advanced	2021	Alpha-Lattice	3	46	138	X		X

QAM, Quality Association Mapping Panel; X, Indicates HTP Data Type Collected.

observations of grain yield. Grain yield was the focus of observation given its importance as the final end-value selection parameter in many plant breeding programs.

2.2 UAS phenotypic data

The Sentera Quad Multispectral Sensor covered target bands of interest for winter wheat evaluation. The camera has four sensors that cover eight broad spectral bands between 450 nm and 970 nm. Collected UAS images are stitched and prepared for data extraction in Pix4Dmapper (Pix4D Inc., Denver, CO, USA), creating a single orthomosaic image for each sensor per location. Orthomosaic images were transferred to Quantum Geographic Information System (QGIS) for plot identification and then further processed with a custom R code for calibration, index calculation, and single plot mean data extraction. In 2018 and 2019, a single reflectance panel (85% reflectance) was used for radiometric calibration on RBG and red edge bands. Quantum efficiency coefficients were used to calculate NIR using:

$$NIR = (2.921 \times Blue) - (0.754 \times Red).$$

The NIR band was then normalized with a coefficient of 3.07 during the calculation of SRIs (Ortiz et al., 2021). In 2020 and 2021, a set of calibration panels (five panels ranging from 2% – 85% reflectance, MosaicMill Oy, Vantaa, Finland) was implemented. Iqbal et al. (2018) developed a simple radiometric calibration methodology using a set of calibration panels with a known variation of reflectance at each broadband wavelength of interest. The band layers are adjusted based on the relationship:

$$SR = DN \times m \pm b,$$

where digital numbers (DN) are the raw observed pixel values for collected orthomosaic images and Surface Reflectance (SR) is the true reflectance value. Slope (m) and intercept (b) are variables explaining the relationship between observed and true values of the reflectance panels. Once slope and intercept are calculated based on the reflectance panels' regression, the corresponding bands can be adjusted.

2.3 Ground phenotypic data

Like the Sentera sensor, the CROPSCAN MSR16R covers target bands of interest. The CROPSCAN radiometer has 16 broad spectral bands that range from 430 nm to 970 nm. Before data collection the sensor is calibrated using manufacture provided calibration panel. The sensor was attached to a pole and placed 1m directly above the wheat canopy in the middle of the plot. One plot is collected at a time, and a mean value for each spectral band is logged for that plot. An irradiance light sensor accounts for light variation and reduces noise in reflectance values. The CROPSCAN MSR system software is used to retrieve collected band values. Plot reflectance values were normalized across all observations by dividing each plot reflectance value by the standard deviation of reflectance values within a trial.

The SVC HR-1024i is a hyperspectral sensor collecting thousands of narrow band values between 338 nm and 2515 nm for each plot. Before sampling, the sensor was calibrated using the manufacture provided calibration panel. The SVC was held by hand at a height of 0.75 m above the plot at an approximate 20-degree angle. Collected SVC data reflectance curves were observed for each plot. Observations with abnormal reflectance curves below 1000 nm were removed from the evaluation. Band reflectance values were normalized as done with CROPSCAN data. Broadband values that reciprocate the collected UAS bands were then calculated by averaging all SVC narrowband values within the 50 nm desired broadband window.

2.4 Spectral reflectance indices calculation and data analysis

Comparing these systems is based on three overlying factors: grain yield correlation with water indices, grain yield correlation with vegetation indices, and overall utility for a large-scale breeding program. The processed spectral reflectance data collected from each plot for both the handheld and UAS platforms were used to calculate the vegetation indices NDVI, NDRE, Transformed Chlorophyll Absorption Reflectance Index (TCARI), Modified Triangular Vegetation Index (MTVI), and the water index, NWI. NDVI and NWI are the most commonly used in each of their corresponding categories and are ideal measures of plant stress and canopy water stress, respectively, in winter wheat (Prasad et al., 2007; Lozada et al., 2020). NDRE, TCARI and MTVI were chosen because of their past success in our breeding program in correlating to yield and accounting for environmental variability. The spectral reflectance bands used and formulas for these indices are shown in Table 2.

Broad-sense heritability (H2) was calculated for SRIs across all sampled locations for grain yield. Genotype, replication, block, environment, and genotype by environment variation were used as random effects in the calculation of H2 with the formula:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GE}^2}{x} + \frac{\sigma_\epsilon^2}{xr}}$$

where σ_G^2 is genetic variance, σ_{GE}^2 is variation due to genotype by environmental effect, σ_ϵ^2 represents variation due to error, x signifies the number of environments, and r represents the number of replications (Bernardo, 2002). Variance components used in heritability calculations were estimated using the "lme4" package in R. Heritability, in conjunction with correlation to grain yield, will indicate an index's success in indirect selection.

Phenotypic correlations among traits were calculated within the two datasets as Pearson correlations using "cor' function in R. Scatterplots and regressions for each unique year and platform combination was generated using "ggplot2" in R.

Principal component analysis (PCA) was conducted for the 2018-2019 and 2020-2021 populations using NDVI, NDRE, TCARI, MTVI and NWI for each platform as well as grain yield. PCA was conducted using the "FactoMineR" package in R.

3 Results

3.1 SRI correlation across platforms and with grain yield

Within each population, the correlation of indices was evaluated between collection methods as well as between indices and grain yield. In 2018-2019 correlations with grain yield were close to zero or slightly negative with handheld NWI, ranging from -0.23 to 0.22. Handheld NDVI was negatively associated with UAS NDVI with a correlation of -0.48, whereas handheld NWI and NDRE had a small positive association with UAS with a correlation of 0.24 and 0.2, respectively. In the 2020-2021 population, NDVI and NDRE had a moderate to high positive correlation between collection strategies and grain yield, as seen in Figure 2 and reinforced by PCA in Figure 3. NWI had a low correlation between UAS and handheld, with a coefficient of 0.07. When correlated to grain yield, UAS and handheld NWI had a negative association of -0.34 and -0.67, respectively. The negative correlation between water-based indices and grain yield was supported in PCA. This negative correlation is expected in NWI with both vegetation indices and grain yield. A lower NWI value indicates higher water content in crop canopy.

3.2 Linear relationship of SRIs to grain yield

A clearer relationship can be seen when linear regressions are conducted with SRIs and grain yield, as seen in Figure 3. Across all years and environmental conditions, UAS NDVI has an expected linear distribution relative to grain yield. Both handheld systems used in this study produced high NDVI values across observations while producing more non-normal distributions relative to grain yield. UAS NDVI generally has a stronger linear relationship to grain yield over the handheld counterpart. Only NDVI is shown because of its relevance to wheat. All other SRIs evaluated have similar trends.

3.3 PCA of SRIs across platforms and grain yield

The first principal component (Dim1) captured between 35.6% and 55.8% of the phenotypic variation. The second principal

TABLE 2 Spectral reflectance index equations.

Spectral Reflectance Index	Abbreviation	Equation	Reference
Normalized Difference Vegetation Index	NDVI	(R800 – R680)/(R800 + R680)	(Rouse et al., 1974)
Normalized Difference Red Edge	NDRE	(R800 – R700)/(R800 + R700)	(Gitelson and Merzlyak, 1996)
Transformed Chlorophyll Absorption Reflectance Index	TCARI	3 × [(R700 – 680) – 0.2 × (R700 – R550)(R700/R680)]	(Haboudane et al., 2002)
Normalized Water Index	NWI	(R970 – R800)/(R970 + R800)	(Gao, 1996)
Modified Triangular Vegetation Index	MTVI	((R700 – R550))/ √ ((2⋆R800 + 1) 2 – (6⋆R800 – 5⋆ √ R680) – 0.5))	(Haboudane et al., 2004)

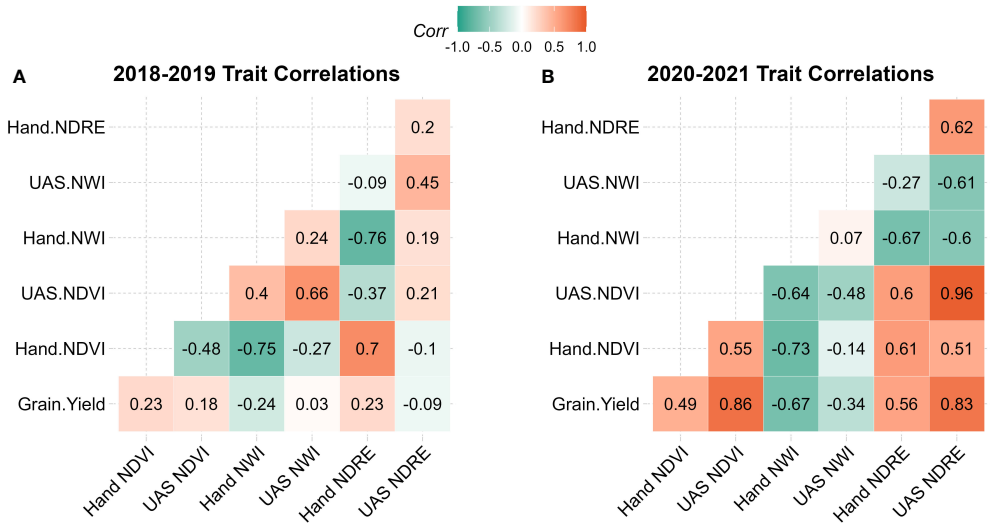


FIGURE 2
Correlation table of grain yield, handheld collected SRIs, and UAS collected SRIs in **(A)** 2018-2019 population and **(B)** 2020-2021 population. NDVI, Normalized Difference Vegetation Index; NDRE, Normalized Difference Red Edge; NWI, Normalized Water Index.

component (Dim2) was only able to explain between 30.9% and 22.4% of trait variation. PCA biplots of individuals in both populations group years along Dim1, indicating strong between-year environmental variability. QAM diversity panels group closely within their given year, while breeding trials (single plot, preliminary, and advanced) tend to spread across Dim2. As seen in Figure 4, the 2020 observations are tightly grouped due to ideal growing conditions reducing genetic expression in trait variability. In 2018-2019 handheld generated indices contributed most in Dim1 while UAS generated indices contributed more in Dim2. This differs from 2020-2021 where most indices were contributing to Dim1 whereas only handheld MTVI and UAS NWI were major contributors to Dim2.

3.4 Heritability of grain yield and SRIs in evaluated trials

Broad-sense heritability for spectral indices of all years evaluated was moderate to high, with a range of 0.50 to 0.80. Grain yield heritability was also calculated at 0.65 in 2018-2019 and 0.76 in 2020-2021, as seen in Table 3. Across both populations, UAS collected indices had a higher heritability than handheld collected indices. This difference was greater with NDVI and NDRE than with NWI.

4 Discussion

In this study, we have outlined the differences in the correlation and heritability performance of SRIs collected from handheld and UAS systems relative to grain yield. This study evaluated HTP data of a breeding population from a single location, over four highly differing years. This is typical of most breeding programs where lines are evaluated initially and selected based on performance at one location. Despite these factors, there are clear differences in the capability of the tested HTP systems for application in a breeding pipeline to improve grain yield selection potential as secondary traits. Both phenotypic correlation and heritability of SRIs were assessed to evaluate the utility in improving selection for grain yield. This section, along with the discussion of analytical results, will break down the less tangible aspects of the HTP systems used in this study and their relative potential utility in breeding applications.

4.1 SRI correlation and precision across platforms

The correlations of SRIs in the 2018-2019 dataset were generally lower than that of the 2020-2021 dataset. It is important to note that NWI is a water index that negatively associates with canopy water content. A higher NWI value indicates lower canopy water content, meaning that a strong negative correlation to grain yield is ideal (Bandyopadhyay et al., 2014). This relatively low correlation in 2018-2019 is potentially due to inadequate data quality caused by more primitive data standardization and poor sensor quality, where only one calibration panel was used. The poor data quality in 2018-2019 is also exemplified in the low correlations of corresponding SRIs between handheld and UAS. Similar findings were shown by Deng et al. (2018) and Díaz-Delgado et al. (2019), both highlighting inconsistencies in sensor performance and correlation, especially when sensor quality or calibration methods are inadequate. Finally, despite moderate heritability, the 2018-2019 SRI data correlates poorly with grain yield. This suggests that the collected data was not capturing the chlorophyll or water content targeted by SRIs, possibly because of the reduced calibration panel set. Ensuring that data collected is of the highest quality is always essential. As additional research was published suggesting the move from a

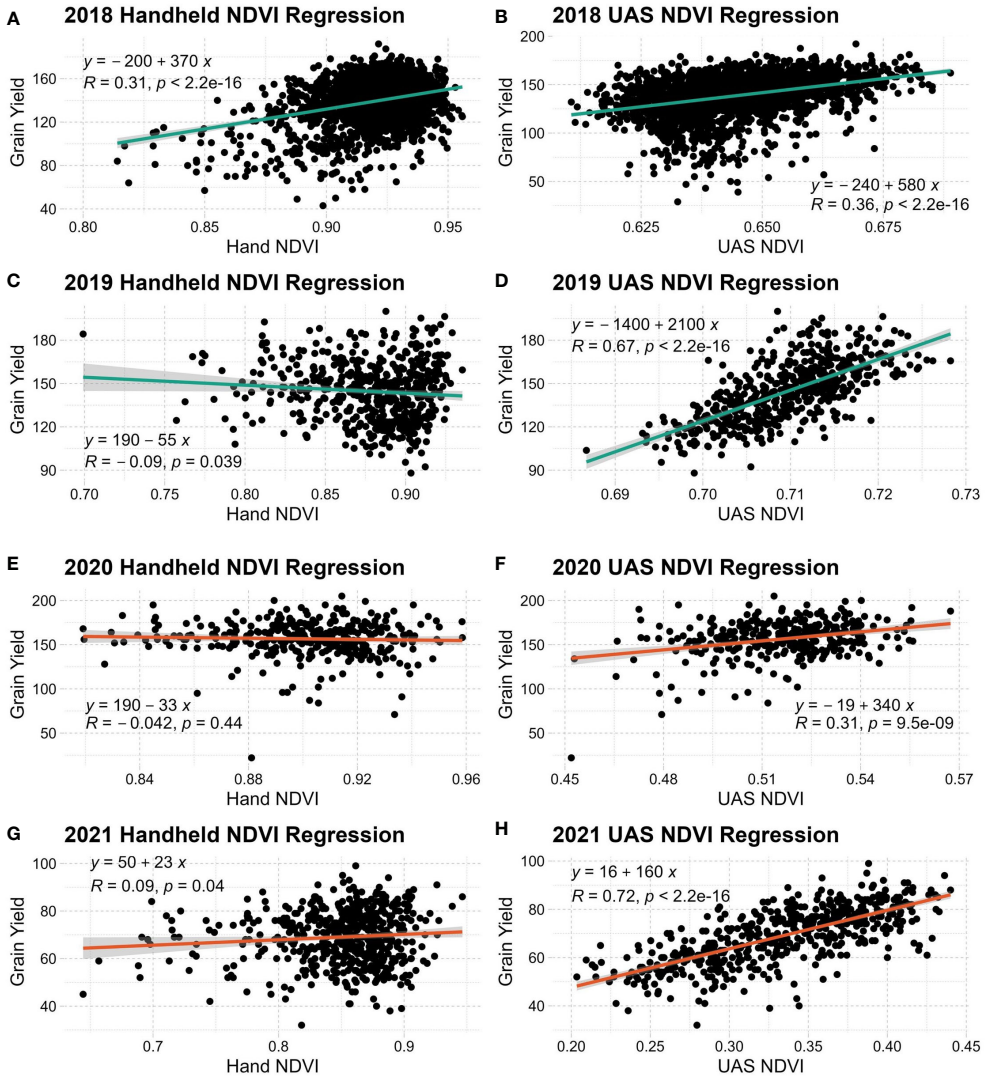


FIGURE 3
Linear regression of grain yield and handheld NDVI **(A, C, E, G)** or UAS NDVI **(B, D, F, H)** in each year evaluated. NDVI, Normalized Difference Vegetation Index.

single calibration panel to multiple panels, subsequent data was improved and yielded higher correlations with grain yield (Duan et al., 2017; Guo et al., 2019).

Unlike the 2018-2019 dataset, in 2020 and 2021, correlations are improved to moderate or high across SRI and platform. UAS data correlations are most likely improved over the 2018-2019 population due to an improved image calibration strategy using a set of five calibration panels. The 2020-2021 data also displays expected patterns across the correlation table between grain yield, handheld and UAS data indicating a more successful capture of target physiological characteristics relative to the 2018-2019 dataset. There are generally stronger correlations among grain yield UAS data in 2020-2021 relative to 2018-2019. These differences between datasets collected by the two ground based systems could possibly be because of variation in climatic conditions of the years. More likely, the improved correlations in 2020 and 2021 were because of improved data quality with the enhanced calibration strategies that were implemented.

The handheld radiometer systems do not have as quick of a collection speed as UAS, allowing for the introduction of error, similar results were found by Tattaris et al. (2016). This issue will be discussed in later sections. This study also validated Tattaris et al. (2016) in higher correlations of UAS derived vegetation indices to yield relative to ground based proximal sensors. In 2020 and 2021, handheld platforms did outperform UAS with NWI correlations. The outlier in correlation is most likely due to NWI's susceptibility to environmental variability and general sensor quality. These results are corroborated in Gutierrez et al. (2010) and Bandyopadhyay et al. (2014), highlighting difficulties in working with NWI. The SVC sensor used in 2020 and 2021 is a hyperspectral sensor capable of greater precision in reflectance evaluations. Reflectance bands used in calculating NWI are within the median

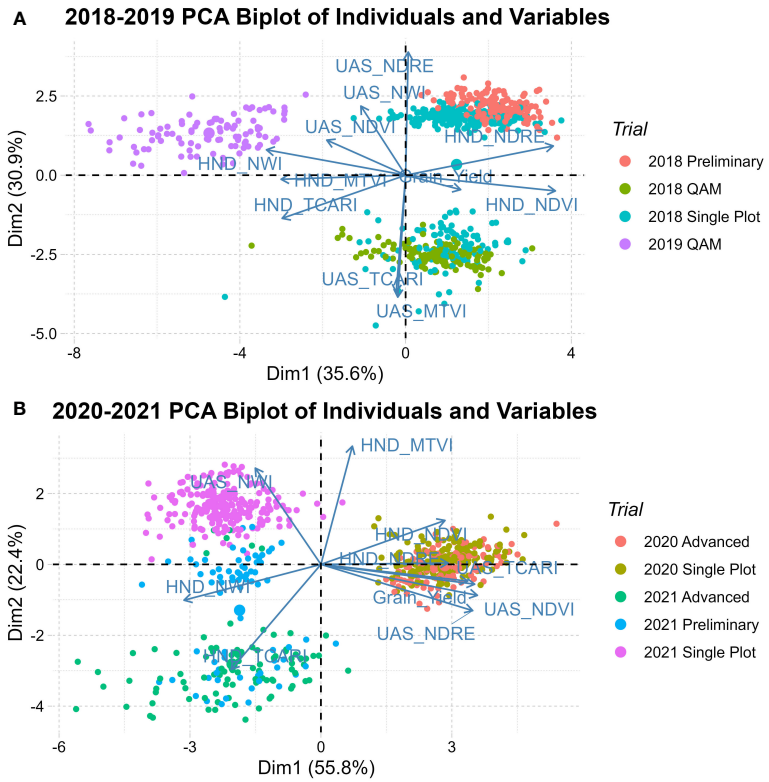


FIGURE 4
Principal component biplot of individuals and vector of variables in **(A)** 2018-2019 population and **(B)** 2020-2021 population showing the genetic relationships of QAM diversity panel, early generation single plot, preliminary, and advanced trials.

reflectance range of the SVC sensor, whereas the UAS camera works with secondary modified sensors.

4.2 SRI heritability and reliability in selection

Across all evaluated SRIs in both populations, UAS data produced a higher broad-sense heritability than handheld systems. This difference in heritability between systems is most likely due to the increased variability of SRI data introduced during a lengthened data collection window. Handheld systems have the disadvantage of collection efficiency, taking approximately 10 seconds per plot, whereas a UAS system can average under 2 seconds per plot. While the UAS, Cropscan and SVC all have methods for sensor calibration, slight changes in solar position and intensity likely impacted reflectance readings. It has been well established that minimization of error during the spectral reflectance data collection process is critical to the final data quality (Guo et al., 2016; Ortiz et al., 2021). Because the UAS system captures several plots at a time and the same plot multiple times, all within a 20-30 min window, it is likely that UAS reflectance data has a reduced potential for error relative to the handheld radiometers used. This difference in data quality is also observed in NDVI's relationships to grain yield across years, shown in Figure 3.

The moderate SRI heritability observed in this study is expected due to a portion of the study population being unreplicated trials. We also expect heritability in the 2018-2019 population to be lower than the 2020-2021 population due to the increased genetic diversity of the population from the inclusion of the QAM diversity panel (Bowman et al., 2015). SRI heritability was generally lower than grain yield, limiting the potential application for indirect selection. However, the moderate correlation and heritability of SRIs suggest the potential for improved genetic gain when utilized as secondary traits in selection. The utilization of SRI data for utilization in breeding for grain yield is most promising when incorporated in genomic selection strategies as a covariate or in multivariate models as shown my Lozada et al. (2020) and Montesinos López et al. (2022) respectively.

TABLE 3 Broad-sense heritability (H^2) of grain yield, UAS indices, and handheld indices.

Population	Grain Yield	Handheld NDVI	UAS NDVI	Handheld NWI	UAS NWI	Handheld NDRE	UAS NDRE
2018-2019	0.65	0.52	0.80	0.50	0.58	0.50	0.60
2020-2021	0.76	0.55	0.68	0.62	0.70	0.58	0.67

4.3 Platform utility in a breeding program

For most plant breeding programs grain yield is the primary trait of interest. The highly quantitative nature of the trait can make selection and prediction efficiency difficult (Reynolds et al., 2012). There is evidence that SRI data can complement and improve tools like genomic selection and machine learning prediction for use in the breeding strategy of grain yield (Montesinos López et al., 2022; Herr et al., 2023). It is important to validate that the methods used in secondary trait data collection are high quality, heritable, and correlate well to the primary trait of interest.

The overarching goal of this study was to determine if the use of ground-based handheld systems provides an increased resolution and data quality that justify the negative aspects of the platform, like collection time, data noise, and secondary applications. As mentioned above, handheld systems have the disadvantage of collection speed; this difference is amplified when capturing large breeding trails. A UAS can collect all data of a 1000 plot breeding trial in approximately 30 min, whereas the handheld system will take roughly 3 hours. In smaller research programs and applications, this difference would have minimal impact on the ability to collect desired datasets. However, in large breeding programs with multiple trial locations, collecting reflectance data across all locations at more than one or two critical time points can be difficult. Solar and weather limitations create narrow windows for image capture, and UAS imaging allows for flexibility in data collection timing. Under ideal environmental conditions, UAS allows for quick data capture across several locations in a single day. The variability seen in heritability and correlation between handheld and UAS is partly due to the differential in capture time. The increased time it takes a handheld radiometer system to collect data on an entire breeding trial, 2-3 hours, creates the potential for changes in solar radiation caused by solar angle or cloud cover. This will produce within field errors in collected reflectance data, creating challenges in distinguishing genetic, phenotypic, and environmental variability (Tattaris et al., 2016).

The HTP systems used in this study highlight the reality of working with technology in long term breeding research applications. When first evaluating the potential of HTP, UAS sensors were not common, thus the Cropscan system was utilized as a platform that was easy to implement in a field-based breeding program. In 2018, as more UAS and sensors became available, they were used in tandem with the hand-held Cropscan. When the Cropscan broke in 2020, alternative solutions were pursued for a ground based radiometer, leading to the use of the SVC system. Similarly, with the UAS calibration, when starting in 2018 the manufacturer recommendations of a single white panel were used. As new research came out it became evident for the need to implement higher quality, multi-panel radiometric calibration in 2020 and 2021. With technology constantly changing and improving, it is important to recognize the potential improvements these can make. It also important to note that new methodology or equipment can impact the quality and reliability of SRI data as shown in this study. As other breeding programs begin using HTP, it is valuable to evaluate different UAS calibration strategies and handheld platforms within the same population and year, and across a diverse set of environments, to clearly identify each technology's reliability.

Each of the three phenotyping systems used in this study has a different method of calibration, collection, and processing that influence the quality of data collected. The SVC and Cropscan radiometer systems initially use a white reference panel to calibrate the sensor. These radiometers do not collect actual images but a range of mean reflectance bands within the sensor field of view. The Cropscan system requires custom software for post-processing to populate reflectance values, and the SVC requires normalization and conversion of narrow hyperspectral band values to multispectral broadbands. A major disadvantage of these radiometers is their inability to screen for reflectance noise within the sensor field of view. The UAS used in this study collects images which are later stitched into an orthomosaic containing the desired reflectance values. These orthomosaics can be used to create soil masks, removing soil and other non-plant reflectance in calculating mean plot reflectance for later use in SRI calculation.

Calibration is another strategy for minimizing reflectance noise and standardizing collected data. All platforms in 2018-2019 and the handheld system in 2020-2021 used a simple single-panel radiometric calibration technique that utilizes the know reflectance of a white panel to adjust sensor readings based on the observed panel reflectance. This method is effective but is more limited in accurately adjusting each reflectance band (Iqbal et al., 2018). Radiometric calibration with a range of calibration panels, a method used for 2020-2021 UAS data collection, improves the former strategy by utilizing three to five reflectance panels with a known range of solar absorption. The range of panels can be used to produce a regression of expected reflectance against observed for each reflectance band of interest. This technique allows for more precise adjustment in individual band readings, producing more reliable reflectance values (Wang and Myint, 2015; Iqbal et al., 2018). The removal of soil noise, robust radiometric calibration, and short flight times minimize the error in the data collected, ensuring reflectance data quality across time and locations.

It is important to maximize limited resources in large-scale applied plant breeding research. Any implementation of HTP can be costly and time consuming. Because of this, when looking at implementing HTP into a breeding program, it is important to consider the versatility and range of the platform selected. The handheld radiometers used in this study can collect high resolution reflectance data across a broad spectral range with the potential for producing SRIs with moderate heritability and correlation to yield, yet are limited in their ability to account for soil noise or inconsistent solar radiation. The data capture speed limits the quality of data collection across locations and time. The lack of orthomosaic image capture in these systems also limits access to secondary traits of interest like plant height and canopy coverage estimates.

With the continual improvement in technology and software, the barrier to entry for UAS phenotyping continues to drop. The speed and efficiency of UAS minimizes labor and cost while providing quality data for further use in breeding strategies. However, it is important to consider best practices that will minimize unwanted environmental variability in collected UAS data. One way that this can be done is by utilizing multiple radiometric calibration panels as outlined in 2020 and 2021 UAS data collection (Iqbal et al., 2018). Another practice that can minimize unwanted variability is in timing of UAS flights. Most

sensors utilized on a UAS are passive sensors, therefore, it is important to adjust for shadowing and variability of solar radiation. For best results it is recommended to fly within a 4-5 hour window of solar noon on days without clouds (Ortiz et al., 2021). In plant breeding programs looking to incorporate high throughput phenotyping, the UAS is an efficient and versatile option that when used properly can produce high quality data.

Overall, it is important to know that not all HTP systems for data collection are created equal. Knowing what HTP traits are most important to the program, frequency and scale of data collection, and resources allocation will help determine which platform will be most beneficial in HTP data collection. When implemented properly, UAS are the more promising system for SRI collection in large-scale breeding programs.

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://doi.org/10.7273/000004802.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any identifiable images or data included in this article.

Author contributions

The project conceptualization and development of methodology was done by AH and AC.The formal investigation, analysis, visualization, original draft preparation, and data curation was conducted by AH. The resources, review and editing, funding acquisition, and supervision was done by AC. All authors contributed to the article and approved the submitted version.

Funding

Research was partially supported by the Agriculture and Food Research Initiative Competitive Grant awards 2022-67013-36426 and 2022-68013-36439 (WheatCAP) from the National Institute of Food and Agriculture, Hatch project 1014919, and the O.A. Vogel Research Endowment at Washington State University.

Acknowledgments

The authors would like to thank Dr. Sinduja Sankaran for use of the SVC equipment, and for many years of collaboration in helping establish an HTP system within the breeding program.

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

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

Babar, M., Reynolds, M., Van Ginkel, M., Klatt, A., Raun, W., and Stone, M. (2006). Spectral reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll, and canopy temperature in wheat. *Crop Sci.* 46, 1046–1057. doi: 10.2135/cropsci2005.0211

Bal, S. K., Wakchaure, G., Potekar, S., Choudhury, B. U., Choudhary, R. L., and Sahoo, R. N. (2021). Spectral signature-based water stress characterization and prediction of wheat yield under varied irrigation and plant bio-regulator management practices. *J. Indian Soc. Remote Sens.* 49, 1427–1438. doi: 10.1007/s12524-021-01325-6

Bandyopadhyay, K., Pradhan, S., Sahoo, R., Singh, R., Gupta, V., Joshi, D., et al. (2014). Characterization of water stress and prediction of yield of wheat using spectral indices under varied water and nitrogen management practices. *Agric. Water Manage.* 146, 115–123. doi: 10.1016/j.agwat.2014.07.017

Bernardo, R. (2002). *Breeding for quantitative traits in plants* (MN: Stemma press Woodbury).

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

Brown, M. E, Antle, J. M., Backlund, P., Carr, E. G., Easterling, W. E., Walsh, M. K., et al. (2015). Climate change, global food security and the US food system. Washington DC: USDA Technical Document. doi: 10.7930/J0862DC7

Carver, B. F. (2009). *Wheat: science and trade* (Ames, IA: John Wiley & Sons).

Deng, L., Mao, Z., Li, X., Hu, Z., Duan, F., and Yan, Y. (2018). UAV-based multispectral remote sensing for precision agriculture: A comparison between different cameras. *ISPRS J. photogrammetry Remote Sens.* 146, 124–136. doi: 10.1016/j.isprsjprs.2018.09.008

Díaz-Delgado, R., Ónodi, G., Kröel-Dulay, G., and Kertész, M. (2019). Enhancement of ecological field experimental research by means of UAV multispectral sensing. *Drones* 3, 7. doi: 10.3390/drones3010007

Duan, T., Chapman, S. C., Guo, Y., and Zheng, B. (2017). Dynamic monitoring of NDVI in wheat agronomy and breeding trials using an unmanned aerial vehicle. *Field Crops Res.* 210, 71–80. doi: 10.1016/j.fcr.2017.05.025

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

Geipel, J., Link, J., and Claupein, W. (2014). Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system. *Remote Sens.* 6, 10335–10355. doi: 10.3390/rs61110335

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

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. doi: 10.3389/fpls.2017.02004

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

Guo, J., Wang, Q., Tong, Y., Fei, D., and Liu, J. (2016). Effect of solar radiation intensity and observation angle on canopy reflectance hyperspectra for winter wheat. *Trans. Chin. Soc. Agric. Eng.* 32, 157–163. doi: 10.11975/j.issn.1002–6819.2016.10.022

Guo, Y., Senthilnath, J., Wu, W., Zhang, X., Zeng, Z., and Huang, H. (2019). Radiometric calibration for multispectral camera of different imaging conditions mounted on a UAV platform. *Sustainability* 11 (4), 978. doi: 10.3390/su11040978

Gutierrez, M., Reynolds, M. P., and Klatt, A. R. (2010). Association of water spectral indices with plant and soil water relations in contrasting wheat genotypes. *J. Exp. Bot.* 61, 3291–3303. doi: 10.1093/jxb/erq156

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

Herr, A. W., Adak, A., Carroll, M. E., Elango, D., Kar, S., Li, C., et al. (2023). Unoccupied aerial systems imagery for phenotyping in cotton, maize, soybean, and wheat breeding. *Crop Science.* 63, 1722–1749. doi: 10.1002/csc2.21028

Iqbal, F., Lucieer, A., and Barry, K. (2018). Simplified radiometric calibration for UAS-mounted multispectral sensor. *Eur. J. Remote Sens.* 51, 301–313. doi: 10.1080/22797254.2018.1432293

Lozada, D. N., Godoy, J. V., Ward, B. P., and Carter, A. H. (2020). Genomic prediction and indirect selection for grain yield in US Pacific Northwest winter wheat using spectral reflectance indices from high-throughput phenotyping. *Int. J. Mol. Sci.* 21, 165. doi: 10.3390/ijms21010165

Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., and Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Sci.* 282, 60–72. doi: 10.1016/j.plantsci.2019.01.007

Montesinos López, O. A., Montesinos López, A., and Crossa, J. (2022). *Multivariate statistical machine learning methods for genomic prediction* (Cham: Springer International Publishing). doi: 10.1007/978-3-030-89010-0

Myneni, R. B., Hall, F. G., Sellers, P. J., and Marshak, A. L. (1995). The interpretation of spectral vegetation indexes. *IEEE Trans. Geosci. Remote Sens.* 33, 481–486. doi: 10.1109/TGRS.1995.8746029

Ortiz, M. V., Sangjan, W., Selvaraj, M. G., McGee, R. J., and Sankaran, S. (2021). Effect of the solar zenith angles at different latitudes on estimated crop vegetation indices. *Drones* 5, 80. doi: 10.3390/drones5030080

Pauli, D., Chapman, S. C., Bart, R., Topp, C. N., Lawrence-Dill, C. J., Poland, J., et al. (2016). The quest for understanding phenotypic variation via integrated approaches in the field environment. *Plant Physiol.* 172, 622–634. doi: 10.1104/pp.16.00592

Prasad, B., Carver, B. F., Stone, M. L., Babar, M. A., Raun, W. R., and Klatt, A. R. (2007). Potential use of spectral reflectance indices as a selection tool for grain yield in winter wheat under great plains conditions. *Crop Sci.* 47, 1426–1440. doi: 10.2135/cropsci2006.07.0492

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PloS One* 8, e66428. doi: 10.1371/journal.pone.0066428

Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., et al. (2020). Breeder friendly phenotyping. *Plant Sci.* 295, 110396. doi: 10.1016/j.plantsci.2019.110396

Reynolds, M., Foulkes, J., Furbank, R., Griffiths, S., King, J., Murchie, E., et al. (2012). Achieving yield gains in wheat. *Plant Cell Environ.* 35, 1799–1823. doi: 10.1111/j.1365-3040.2012.02588.x

Rouse, J. W., Haas, R. H, Schell, J. A., Deering, D. W., and Harlan, J. C. (1974). Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. (Greenbelt, MD, USA: NASA/GSFC Final Report. NASA) pp. 1–137.

Sankaran, S., Khot, L. R., and Carter, A. H. (2015a). Field-based crop phenotyping: Multispectral aerial imaging for evaluation of winter wheat emergence and spring stand. *Comput. Electron. Agric.* 118, 372–379. doi: 10.1016/j.compag.2015.09.001

Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., et al. (2015b). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *Eur. J. Agron.* 70, 112–123. doi: 10.1016/j.eja.2015.07.004

Sapkota, B. R., Adams, C. B., Kelly, B., Rajan, N., and Ale, S. (2023). Plant population density in cotton: Addressing knowledge gaps in stand uniformity and lint quality under dryland and irrigated conditions. *Field Crops Res.* 290, 108762. doi: 10.1016/j.fcr.2022.108762

Sarkar, S., Oakes, J., Cazenave, A.-B., Burow, M. D., Bennett, R. S., Chamberlin, K. D., et al. (2022). Evaluation of the US peanut germplasm mini-core collection in the Virginia-Carolina region using traditional and new high-throughput methods. *Agronomy* 12, 1945. doi: 10.3390/agronomy12081945

Song, P., Wang, J., Guo, X., Yang, W., and Zhao, C. (2021). High-throughput phenotyping: Breaking through the bottleneck in future crop breeding. *Crop J.* 9, 633–645. doi: 10.1016/j.cj.2021.03.015

Tattaris, M., Reynolds, M. P., and Chapman, S. C. (2016). A direct comparison of remote sensing approaches for high-throughput phenotyping in plant breeding. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01131

Visitacion, G. J., Saludes, R. B., Luyun, R. A.Jr., Pinca, Y. M. M., and Eusebio, M. F. V. (2022). Statistical analysis of crop water stress in rainfed rice (Oryza sativa L.) using spectral and non-spectral indices. *Philippine J. Sci.* 15 (2), 587–603.

Wang, C., and Myint, S. W. (2015). A simplified empirical line method of radiometric calibration for small unmanned aircraft systems-based remote sensing. *IEEE J. selected topics Appl. Earth observations Remote Sens.* 8, 1876–1885. doi: 10.1109/JSTARS.2015.2422716

Wientjes, E., Philippi, J., Borst, J. W., and van Amerongen, H. (2017). Imaging the Photosystem I/Photosystem II chlorophyll ratio inside the leaf. *Biochim. Biophys. Acta (BBA) - Bioenergetics* 1858, 259–265. doi: 10.1016/j.bbabio.2017.01.008

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

Xiong, W., Reynolds, M., and Xu, Y. (2022). Climate change challenges plant breeding. *Curr. Opin. Plant Biol.* 70, 102308. doi: 10.1016/j.pbi.2022.102308

Xue, J., and Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *J. Sensors* 2017, 17. doi: 10.1155/2017/1353691

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

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 19 September 2023
DOI 10.3389/fpls.2023.1248152

Check for updates

OPEN ACCESS

EDITED BY
Bo Li,
Syngenta, United Kingdom

REVIEWED BY
Si Yang,
China Agricultural University, China
Wenhao Zhang,
University of the West of England, United Kingdom

*CORRESPONDENCE
Amrit Pokhrel
amritpokhrel@uga.edu

RECEIVED 26 June 2023
ACCEPTED 29 August 2023
PUBLISHED 19 September 2023

CITATION
Pokhrel A, Virk S, Snider JL, Vellidis G, Hand LC, Sintim HY, Parkash V, Chalise DP, Lee JM and Byers C (2023) Estimating yield-contributing physiological parameters of cotton using UAV-based imagery. *Front. Plant Sci.* 14:1248152. doi: 10.3389/fpls.2023.1248152

COPYRIGHT
© 2023 Pokhrel, Virk, Snider, Vellidis, Hand, Sintim, Parkash, Chalise, Lee and Byers. 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.

Estimating yield-contributing physiological parameters of cotton using UAV-based imagery

Amrit Pokhrel[1*], Simerjeet Virk[1], John L. Snider[1], George Vellidis[1], Lavesta C. Hand[1], Henry Y. Sintim[1], Ved Parkash[1], Devendra P. Chalise[1], Joshua M. Lee[1] and Coleman Byers[2]

[1]Department of Crop and Soil Sciences, University of Georgia, Tifton, GA, United States, [2]College of Engineering, University of Georgia, Athens, GA, United States

Lint yield in cotton is governed by light intercepted by the canopy (IPAR), radiation use efficiency (RUE), and harvest index (HI). However, the conventional methods of measuring these yield-governing physiological parameters are labor-intensive, time-consuming and requires destructive sampling. This study aimed to explore the use of low-cost and high-resolution UAV-based RGB and multispectral imagery 1) to estimate fraction of IPAR ($IPAR_f$), RUE, and biomass throughout the season, 2) to estimate lint yield using the cotton fiber index (CFI), and 3) to determine the potential use of biomass and lint yield models for estimating cotton HI. An experiment was conducted during the 2021 and 2022 growing seasons in Tifton, Georgia, USA in randomized complete block design with five different nitrogen treatments. Different nitrogen treatments were applied to generate substantial variability in canopy development and yield. UAV imagery was collected bi-weekly along with light interception and biomass measurements throughout the season, and 20 different vegetation indices (VIs) were computed from the imagery. Generalized linear regression was performed to develop models using VIs and growing degree days (GDDs). The $IPAR_f$ models had R^2 values ranging from 0.66 to 0.90, and models based on RVI and RECI explained the highest variation (93%) in $IPAR_f$ during cross-validation. Similarly, cotton above-ground biomass was best estimated by models from MSAVI and OSAVI. Estimation of RUE using actual biomass measurement and RVI-based $IPAR_f$ model was able to explain 84% of variation in RUE. CFI from UAV-based RGB imagery had strong relationship (R^2 = 0.69) with machine harvested lint yield. The estimated HI from CFI-based lint yield and MSAVI-based biomass models was able to explain 40 to 49% of variation in measured HI for the 2022 growing season. The models developed to estimate the yield-contributing physiological parameters in cotton showed low to strong performance, with $IPAR_f$ and above-ground biomass having greater prediction accuracy. Future studies on accurate estimation of lint yield is suggested for precise cotton HI prediction. This study is the first attempt of its kind and the results can be used to expand and improve research on predicting functional yield drivers of cotton.

KEYWORDS

fraction of intercepted photosynthetically active radiation, radiation use efficiency, harvest index, unmanned aerial vehicles, RGB imagery, multispectral imagery

1 Introduction

Cotton (*Gossypium* sp.) is one of the most important crops to the global textile and clothing industry (Sui et al., 2017; Lu, 2022). Cotton production supports the trillion dollar fashion industry, attracts revenues for countries, and provides employment opportunities to millions of people (Voora et al., 2020). The United Sates is a prime producer as well as exporter of this natural fiber, producing around 3 MT of cotton ($6 billion worth of value) and providing around 30% of global exports (USDA, 2022; USDA, 2023). Because of its significance, the majority of breeding and crop management efforts in the United States and Australia have been focused on lint yield improvements and yield stability (Conaty and Constable, 2020; Snider et al., 2021b). Physiologically, yield is a function of the cumulative amount of photosynthetically active radiation intercepted by the canopy during the growing season (IPAR), the efficiency with which a crop converts intercepted radiation into biomass (RUE), and the fraction of total biomass allocated to the economically important part of the crop, or harvest index (HI) (Monteith, 1972). Thus, yield improvements or declines resulting from breeding or management efforts can be attributed to alterations in any or all of the aforementioned traits. Nitrogen is one of the most important yield-governing factors that influences cotton growth and development. For example, nitrogen deficiency decreases leaf area expansion and canopy development (Wullschleger and Oosterhuis, 1990), which could reduce IPAR (Wajid et al., 2010). Reductions in single-leaf and whole-canopy photosynthetic rates commonly occur under nitrogen deficiency (Bondada and Oosterhuis, 2001; Snider et al., 2021a; Parkash et al., 2023), and previous reports have documented positive associations between N application rates and whole-canopy RUE (Bange and Milroy, 2000). Due to a reduction in the ability of the canopy to support fruit development under N deficiency or due to excessive vegetative growth and low fruit retention under N excess, yield can be significantly reduced (Gerik et al., 1994). As a result, it is important to understand the response of underlying physiological parameters (IPAR, RUE, and HI) to yield-altering N application rates. Traditionally in small-plot research, IPAR is calculated by measuring above and below canopy light intensity, RUE estimation requires above ground dry weight samples, and HI measurement needs hand harvested samples of cotton. However, these traditional methods of measuring the physiological parameters are both time and labor intensive, and require destructive sampling of plants (Ermanis et al., 2020).

Due to most physiological measurements being laborious and time-consuming, remote sensing has the potential of becoming a rapid and efficient non-destructive method for characterizing crop and vegetation bio-physical properties (Wiegand et al., 1991; Kumar et al., 2002). Over the years, remote sensing platforms have evolved from low-resolution orbital satellites in the 1970s to advanced unmanned aerial vehicles (UAVs) in recent years. In comparison to satellites, UAVs provide many improved features such as stability, reliability, autonomy, affordability, flights at lower altitudes, optimum data quality, high spatial and temporal resolution (Sankaran et al., 2015; Barbedo, 2019). While UAVs also have some limitations such as short battery life, technical knowledge, and limited airspace authorization (Jafarbiglu and Pourreza, 2022) they still provide an opportunity for rapid, reliable, and non-destructive measurement of crop biophysical characteristics. Furthermore, the availability of reliable and high spectral resolution sensors, such as Red Green Blue (RGB) (Bendig et al., 2015), multispectral (mainly red, blue, green, red-edge, and near-infrared spectral bands) (Deng et al., 2018), hyperspectral (a wide range of spectral bands) (Zhang et al., 2018), thermal (Gonzalez-Dugo et al., 2013), and depth (LiDAR) (Sun et al., 2018) have greatly expanded the capabilities of UAVs in agriculture. Low-cost sensors such as RGB and multispectral sensors on small, light-weight UAVs are commonly used by agricultural researchers because of their dependability, affordability, and ease of image processing and analysis (Jafarbiglu and Pourreza, 2022). RGB imagery has primarily been used for extracting soil or crop characteristics of interest (Huang et al., 2016; Li et al., 2019), and estimating crop heights (Bendig et al., 2015). By comparison, multispectral imagery incorporates red-edge (670-760 nm) and near-infrared (760-900 nm) wavebands that are inaccessible when using traditional RGB imagery (Ashapure et al., 2019). As a result, multispectral imagery has been widely used to generate vegetation indices (VIs) in many crops, which are used to estimate canopy cover (Xu et al., 2019), biomass (Yue et al., 2019; Wang et al., 2021), leaf area index (Boegh et al., 2002; Gutierrez et al., 2012), chlorophyll content (Raper and Varco, 2014), evapotranspiration (Hunsaker et al., 2003), nutrient status (Zhao et al., 2005; Ballester et al., 2017), and yield (Zhao et al., 2007b; Vatter et al., 2021).

Currently, limited studies have attempted to correlate VIs derived from UAV-based multispectral imagery to IPAR in cotton. However, VIs derived from satellite imagery, ground spectro-radiometers, and hyperspectral sensors have been shown as a suitable proxy for the fraction of incident PAR ($IPAR_f$) intercepted by the canopy. For example, multiple studies have reported the Normalized Difference Vegetation Index (NDVI) to be a good predictor of $IPAR_f$ for cotton (Prasad et al., 2021), garlic (Campoy et al., 2019), wheat (Tan et al., 2018; Pellegrini et al., 2020; Lv et al., 2021), corn (Tan et al., 2013), and soybean (Hatfield and Prueger, 2010). Furthermore, $IPAR_f$ is highly dependent on crop canopy structure and architectural traits like leaf area index (LAI), and several studies have shown that the Ratio Vegetation Index (RVI), Green Normalized Difference Vegetation Index (GNDVI), Enhanced Vegetation Index (EVI) and NDVI can be used to predict cotton LAI (Zhao et al., 2007a; Gutierrez et al., 2012; Chen, 2019) with a moderate to high accuracy (R^2 = 0.70 to 0.93).

Reports on the relationship between multispectral VIs and RUE are also limited in cotton. Most studies on RUE estimation for leaf and canopy have used Photochemical Reflectance Index (PRI) and Sun-Induced Fluorescence (SIF) estimates, obtained from narrow-band hyperspectral reflectance data and fine resolution spectrometers, respectively (Garbulsky et al., 2011; Zhang et al., 2016; Zhang et al., 2018; Merrick et al., 2020; Fu et al., 2022). PRI uses reflectance data obtained from 531 and 570 nm wavelengths that measures changes in xanthophyll cycle and pigment ratios (chlorophyll/carotenoid), (Hilker et al., 2008; Garbulsky et al., 2011; Zhang et al., 2016), while SIF estimates the fluorescence emission in

the far-red region (650-850 nm) from excited chlorophyll (Zhang et al., 2018; Merrick et al., 2020; Porcar-Castell et al., 2021), both of them ultimately relating to photosynthetic efficiency and RUE at leaf and canopy levels. For example, recent studies have used PRI or SIF or both to estimate RUE in crops like corn (Cheng et al., 2013) and wheat at the canopy level (Robles-zazueta et al., 2021), and cotton at the leaf level under water stress conditions (Zhang et al., 2018). However, obtaining SIF and hyperspectral VIs remains a challenge due to its high cost instruments, sensitivity to noise, and complex analytical procedures (Hilker et al., 2008; Porcar-Castell et al., 2021). Because chlorophyll content per unit leaf area can influence the quantum efficiency of primary photochemistry (Porcar-Castell et al., 2021), it is possible that multispectral VIs related to chlorophyll content are predictive of canopy-level RUE in cotton. Few studies have investigated the potential of multispectral VIs to estimate chlorophyll content, where Red-edge Chlorophyll Index (RECI) and Simplified Canopy Chlorophyll Content Index (SCCCI) are shown to be related to chlorophyll content in cotton and other crops (Raper and Varco, 2014; Ballester et al., 2017; Barbedo, 2019; Wang et al., 2021). Furthermore, previous research has shown that certain multispectral VIs, including NDVI, RVI, NIR, Normalized Difference Red-edge Index (NDRE), Modified Soil Adjusted Vegetation Index (MSAVI), Wide Dynamic Range Vegetation Index (WDRVI) can be used to estimate cotton biomass (Zhao et al., 2007a; Hatfield and Prueger, 2010; Gutierrez et al., 2012; Brandão et al., 2015; Chen and Wang, 2020). Because biomass is the product of IPAR ($IPAR_f \times PAR$) and RUE (Conaty and Constable, 2020), the estimation of $IPAR_f$ and biomass using multispectral VIs could potentially be used to derive RUE.

While the studies mentioned above have linked multispectral VIs to cotton LAI, biomass, and chlorophyll content, most of these relationships have been developed for a specific point in time or at a specific growth stage in these studies. As a result, the previously developed relationships are only predictive of these characteristics when measured at that specific time or growth stage. Reports have indicated that the relationship between VIs and crop growth characteristics such as LAI and biomass change with the phenological stage of the crop (Pinter et al., 2003; Gutierrez et al., 2012; Li Z. et al., 2022). A recent study from Li Z. et al. (2022) demonstrated significant improvements (R^2 = 0.83) in estimating the above-ground biomass of wheat throughout the season using multispectral VIs when used in conjunction with a well-established crop staging system such as growing degree days (GDD) or heat units. For cotton, crop growth and development are strongly tied to GDD accumulation (Ritchie et al., 2004; Sharma et al., 2021). Consequently, integrating GDD into VI-based functions can also be potentially used for predicting canopy-specific, yield-driving traits for cotton at any time during the season.

As crop biomass and yield can be predicted from aerial imagery (Gutierrez et al., 2012; Huang et al., 2016), it is likely that HI can be predicted from VIs used to derive these two traits. However, the authors are not aware of studies that have used multispectral imagery to capture in-field variation in HI for field-grown cotton. As discussed earlier, multiple VIs derived from multispectral imagery have shown a strong relationship with cotton biomass. Similarly, several studies have also suggested that multispectral imagery-based VIs such as NDVI, SCCI, NDRE and RVI as well as the individual reflectance bands including red and red-edge collected during the growing season have the potential to explain variation in lint yield (Yang et al., 2001; Zarco-Tejada et al., 2005; Zhao et al., 2007b; Gutierrez et al., 2012; Huang et al., 2013; Ballester et al., 2017). However, results from these studies indicate some discrepancies in the appropriate growth stage at which these VIs are effective, with few studies suggesting the relationship was better at the early flowering stage and others at the peak bloom stage. Huang et al. (2016) developed an alternate method to estimate cotton lint yield (R^2 = 0.83) after defoliation and before harvesting, based on the detection of white cotton pixels in an aerial RGB imagery. Feng et al. (2020) later named this ratio as Cotton Fiber Index (CFI) and showed a similar performance of CFI (R^2 = 0.90) in estimating cotton lint yield. CFI is the ratio of the number of white pixels (cotton bolls) to the total number of pixels in a given area and can be obtained from UAV-based RGB imagery following some image enhancement procedures.

Based on the available literature on the use of VIs to predict crop biophysical characteristics, VIs derived from UAV-based aerial imagery can potentially be utilized to predict yield-governing physiological parameters in field-grown cotton. However, studies predicting $IPAR_f$, RUE, and HI for field-grown cotton using UAV-based RGB and/or multispectral imagery have not been published. Thus, the main objectives of our study were to: 1) develop and validate models to estimate $IPAR_f$, RUE, and biomass of cotton throughout the season using VIs derived from UAV-based multispectral imagery and GDDs; 2) estimate cotton lint yield using cotton fiber index (CFI) extracted from UAV-based RGB imagery; and 3) investigate the potential of using biomass and lint yield estimates obtained from UAV-based RGB and multispectral imagery, respectively, to estimate cotton HI.

2 Materials and methods

2.1 Study site details and experimental design

The study was conducted at the Lang-Rigdon Farm (Figure 1) of the University of Georgia Tifton Campus in Tifton, Georgia, USA (31° 52' N, 83° 55' W) where the predominant soil type is classified as Tifton loamy sand (fine-loamy, kaolinite, plinthic kandiudults) (Soil Survey Staff, Natural Resources Conservation Service and United States Department of Agriculture, no date). During the 2021 and 2022 growing seasons, a field experiment was arranged in a randomized complete block design with five replications using a cotton cultivar DP 1646 B2XF and five nitrogen application rates of 0, 44, 89, 135, and 179 kg N ha^{-1}. DP 1646 B2XF was the most widely grown cotton cultivar in the southeastern US during the time of this experiment (USDA Agricultural Marketing Service, 2020), and the five nitrogen application rates were implemented to create variability in crop growth and yield. N was applied as granular urea (46-0-0): 25% at planting and 75% at the initiation of floral bud development (squaring stage). All other agronomic management practices, except N application, were conducted based on the

recommendations outlined in the University of Georgia Cotton Production Guide (Hand et al., 2022). The cotton crop was sown on June 1, 2021 and April 26, 2022, and machine harvested on October 25, 2021 and September 22, 2022. The weather data which includes daily maximum and minimum temperatures, and daily solar radiation (Supplementary Figure 1) for the study site from planting until harvest was obtained from the University of Georgia Weather Network (http://www.georgiaweather.net/). The 2022 growing season had higher average daily minimum and maximum temperatures by 0.15°C and 1.16°C, respectively than the 2021 growing season. The 2021 growing season had cloudier days with daily solar radiation below 10 MJ m^{-2} than the 2022 growing season. Daily maximum and minimum temperatures were used to calculate GDDs throughout the season for cotton during both years.

The accumulated GDDs of cotton over time is a predictor of growth and phenological stage during the season. The GDDs is calculated based on the average daily temperatures and base temperature of cotton using equation 1 (Ritchie et al., 2004).

$$\text{GDDs} = \sum_{i=1}^{n}\left[\frac{\text{Max. Temperature } ^\circ\text{C}_i + \text{Min. Temperature } ^\circ\text{C}_i}{2} - \text{Base Temperature } ^\circ\text{C}\right] \quad (1)$$

where $i = 1$ signifies first day of planting and n is each sampling date. For cotton, the base temperature is 15.6°C (Snider et al., 2021b; Hand et al., 2022) and the upper threshold of maximum temperature is 33.9°C (Hand et al., 2022). The growth and development of cotton plants is assumed to be zero below the base temperature and above the upper threshold temperature.

2.2 Data collection

Physiological measurements and UAV-based imagery were collected on the same dates, beginning 5 weeks after planting, throughout the 2021 and 2022 growing seasons in approximately two-week intervals, depending on weather conditions on target sampling dates. The actual dates of data collection along with the GDD for each sampling date are provided in Table 1.

2.2.1 Physiological parameters and yield measurements

An AccuPAR LP-80 ceptometer (METER Environment, Pullman, WA) was used for light interception measurements from the middle two rows of each plot between 1000 and 1400 h under cloudless conditions. An integrated linear sensor and an attached external sensor to a tripod stand were used to measure below-canopy photosynthetically active radiation (PAR_{below}) and above-canopy photosynthetically active radiation (PAR_{above}), respectively. The PAR_{below} is an average value from the linear sensor positioned perpendicular and parallel to the row. These values were used to calculate the fraction of intercepted photosynthetically active radiation ($IPAR_f$) on a given day using equation 2.

$$IPAR_f = (PAR_{above} - PAR_{below})/PAR_{above} \quad (2)$$

Furthermore, cumulative incident PAR from planting until a specific sampling date was calculated by summing incident daily PAR (assuming PAR is 45% of solar radiation) for all days included in the defined time frame. Finally, cumulative IPAR (MJ m^{-2}) at the sampling date was the product of $IPAR_f$ and cumulative incident PAR.

Above-ground dry biomass in 2021 was obtained by harvesting all plants from a one-meter length of the row from one of the middle two rows of each plot, while in 2022, the sample size was increased to a two-meter length of the row. The fresh samples were dried in a forced-air oven at 80°C for 72 hours and then weighed to obtain above-ground dry biomass in g m^{-2}.

Radiation use efficiency (RUE) (g MJ^{-1}) was determined by dividing the change in above-ground dry biomass (g m^{-2}) by the

TABLE 1 Information on sampling date and types of measurements taken during the 2021 and 2022 growing seasons.

Year	DAP	Date	GDDs	Measurements	UAV-based imagery
2021	44	July 14	465	LI, AGB	RGB and multispectral
	57	July 27	614	LI, AGB	RGB and multispectral
	70	August 9	763	LI, AGB	RGB and multispectral
	94	September 2	1045	LI, AGB	RGB and multispectral
	116	September 24	1252	LI, AGB	RGB and multispectral
	144	October 25	1458	Lint Yield	RGB
2022	43	June 7	372	LI, AGB	RGB and multispectral
	59	June 23	569	LI, AGB	RGB and multispectral
	75	July 9	768	LI, AGB	RGB and multispectral
	100	August 3	1060	LI, AGB	RGB and multispectral
	114	August 17	1220	LI, AGB	RGB and multispectral
	150	September 22	1588	Lint Yield	RGB

DAP, Days after planting; GDDs, Growing degree days; LI, Light Interception; AGB, Above-ground biomass; UAV, Unmanned Aerial Vehicle; RGB, Red Green Blue.

change in cumulative IPAR (MJ m^{-2}), where the first sampling date was taken as a reference point as shown in equation 3.

$$RUE = (Dry\ biomass_n - Dry\ biomass_1)/(Cumulative\ IPAR_n - Cumulative\ IPAR_1) \quad (3)$$

where, dry biomass $_1$ and cumulative IPAR $_1$ are the reference measurements of the first sampling date, and n represents each subsequent sampling date in the season.

Further, the harvest index (HI) for each plot was calculated as the ratio of lint yield (kg ha^{-1}) to the highest above-ground dry biomass obtained during the season (kg ha^{-1}). For lint yield (kg ha^{-1}), the middle two rows from each plot during both years were mechanically harvested using a two-row spindle cotton picker (John Deere 9930 (John Deere, Moline, IL)), and the harvested seed-cotton was ginned at the University of Georgia MicroGin (Li et al., 2011) to obtain a realistic measure of gin turnout and lint yield (kg ha^{-1}).

2.2.2 UAV data collection and image processing

Both RGB and multispectral imagery were acquired during 1000 to 1400 h using a flight altitude of 45 m. The RGB aerial imagery was acquired using an integrated RGB sensor on DJI Phantom™ Pro 4 V2.0 (Shenzhen, China) in 2021 and on DJI Mavic™ Air 2 (Shenzhen, China) in 2022. The DJI Mavic Air 2 was used on the September 22, 2022 sampling date (prior to harvesting) because of its higher resolution. Multispectral images were acquired during both years using a MicaSense RedEdge-MX™ (Seattle, WA) sensor mounted on a DJI Inspire™ 2 (Shenzhen, China) UAV platform. Detailed information on the sensors and the flight settings used for aerial imagery data collection is provided in Table 2. Prior to any data collection each year, ground control points (GCPs) were placed at four corners of the field and were georeferenced using a handheld Trimble GNSS Unit (Sunnyvale, CA) with a GPS/GNSS RTK correction (± 0.50 cm) enabled. All flights were created and implemented using the Pix4Dcapture® software (Pix4D, Lausanne, Switzerland).

After each UAV flight, all images were downloaded and processed in the Pix4Dmapper® software (Pix4D, Lausanne, Switzerland) using the GCPs and pre-flight calibration to create one orthorectified mosaic images per bands for each sampling date. The Pix4Dmapper® software uses modified structure-from-motion (SFM) approach to create the orthorectified mosaic images. For radiometric calibration of multispectral images, before each flight the MicaSense RedEdge-MX™ multispectral sensor was used to capture reference images of a MicaSense Calibrated Reflectance Panel provided by the manufacturer. These captured reference images for each multispectral band are used act as an input in Pix4Dmapper processing options to perform radiometric calibration and correction for each of the five bands.

2.2.3 Reflectance values and image feature extraction

Orthorectified mosaic multispectral images were used to extract raw reflectance values for each plot. The reflectance values for each band (red, green, blue, red-edge, and near-infrared) for all sampling dates were extracted in ArcMap® 10.7.1 (ESRI, Redlands, CA). Cotton canopy and soil were segmented in each image using a classification index: multiplication of the Normalized Difference Vegetation Index (NDVI) and Excessive Greenness Index (ExG) (Moghimi et al., 2018). A global threshold (pixel values greater than 0.02) was applied to the classified image to obtain a binary mask layer to separate the canopy from bare soil. The binary mask layer was applied to the orthorectified mosaic image for each reflectance band. A polygon specifying the region of interest (ROI) – the middle two rows of each plot – was created and used to extract the averaged reflectance values for all multispectral bands for each plot. The reflectance values for each band were further used to compute several VIs as shown in Table 3 using Microsoft Excel® (Redmond, WA).

RGB images taken immediately before harvest during each year were used to compute cotton fiber index (CFI) values for each plot (Huang et al., 2016; Feng et al., 2020). CFI estimates the total open cotton bolls in an ROI and is calculated using equation 4.

TABLE 2 Flight plan details and technical specifications for the RGB and multispectral sensors used for aerial imagery collection in the study sites for the 2021 and 2022 growing seasons.

Flight details and specifications	DJI's Phantom™ Pro 4 V2.0	DJI's Mavic™ Air 2	MicaSense RedEdge-MX™
Flight Altitude	45 m	45 m	45 m
Spatial Resolution	1.33 cm	1.62 cm	3.29 cm
Image Overlap	Side 70% Front 80%	Side 70% Front 80%	Side 80% Front 80%
Flight Speed	4.11 m s^{-1}	4.91 m s^{-1}	3.81 m s^{-1}
Sensor	1-inch CMOS; 20 Megapixel	1/2-inch CMOS; 48 Megapixel	RedEdge-MX sensor
Bits per pixel	16	16	12
Spectral Range	RGB	RGB	Blue (475 nm ± 20 nm) Green (560 nm ± 20 nm) Red (668 nm ± 10 nm) Red-edge (717 nm ± 10 nm) Near Infrared (840 nm ± 40 nm)

TABLE 3 List of 20 different vegetation indices (VIs) computed in this study from the raw RGB and Multispectral bands.

Vegetation Indices	Formulation	References
ExG (Excessive Greenness Index)	(2×G)−R−B	Woebbecke et al. (1995)
NDVI (Normalized Difference Vegetation Index)	(NIR−R)/(NIR+R)	Rouse (1974)
GNDVI (Green Normalized Difference Vegetation Index)	(NIR−G)/(NIR+G)	Gitelson et al. (2003a)
NDRE (Normalized Difference Red Edge Index)	(NIR−RE)/(NIR+RE)	Gitelson and Merzlyak (1997)
RVI (Ratio Vegetation Index)	NIR/R	Tucker (1979)
SCCCI (Simplified Canopy Chlorophyll Content Index)	NDRE/NDVI	Raper and Varco (2014)
SAVI (Soil Adjusted Vegetation Index)	(1 + 0.5)((NIR−R)/(NIR+R+0.5))	Huete (1988)
EVI (Enhanced Vegetation Index)	(2.5×NIR−R)/(NIR+6×R−7.5×B+1)	Huete et al. (2002)
EVI2 (Enhamced Vegetation Index2)	(2.5×NIR−R)/(NIR+2.5×R+1)	Jiang et al. (2008)
MSAVI (Modified Soil Adjusted Vegetation Index)	$[(2NIR+1)-\sqrt{((2NIR+1)^2-8(NIR-R))}]/2$	Qi et al. (1994)
VARI (Visible Atmospherically Resistant Index)	(G−R)/(G+R−B)	Gitelson et al. (2002)
WDRVI (Wide Dynamic Range Vegetation Index)	(0.2×NIR−R)/(0.2×NIR+R)	Gitelson (2004)
RECI (Rededge Chlorophyll Index)	(NIR/RE)-1	Gitelson et al. (2003b)
RE/R (Redegde to Red ratio)	RE/R	Ballester et al. (2013)
NIR/G (Near infrared to Green ratio)	NIR/G	Sripada et al. (2005)
RGBVI (Red Green Blue Vegetation Index)	$(G-B\times R)/(G^2+(B\times R))$	Bendig et al. (2015)
GRVI (Green Red Vegetation Index)	(G−R)/(G+R)	Falkowski et al. (2005)
TCARI (Transformed Chlorophyll Absorption Reflectance Index)	3[(RE−R)−0.2(RE−G)×(RE/R)]	Haboudane et al. (2002)
OSAVI (Optimized Soil Adjusted Vegetation Index)	(1 + 0.16)[(NIR−R)/(NIR+R+0.16)]	Rondeaux et al. (1996)
TCARI/OSAVI	TCARI/OSAVI	Haboudane et al. (2002)

R, G, B, RE, and NIR represents Red, Green, Blue, Red-edge, and Near-infrared bands, respectively, and were obtained from multispectral imagery collected at different sampling dates during the 2021 and 2022 season.

$$\text{CFI} = \text{Number of white pixels in the ROI} / \text{Total number of pixels in the ROI} \quad (4)$$

As Feng et al. (2020) and Huang et al. (2016) suggested, a simple global threshold to RGB image was unable to completely differentiate white cotton pixels from background pixels of defoliated vegetation and soil. Therefore, a series of image filtering and image enhancing techniques (Figure 2) were performed in MATLAB® 2022b 9.13.0 (The MathWorks Inc., Natick, MA) and ArcMap® 10.7.1 (ESRI, Redlands, CA) for accurate detection of white pixels. First, a bilateral filter (Figure 2: Step 2), using a non-linear filtering method was applied for smoothing the images to remove background leaves and soil surface noises, while preserving the shapes and high-intensity values of cotton pixels. Then, a 5 x 5 Laplacian filter (Figure 2: Step 3) was applied to increase the contrast and sharpness to enhance the edges of white cotton pixels. An additional smoothing filter, with arithmetic mean (Figure 2: Step 4), was further utilized to remove the extra noise introduced by the Laplacian filter. Finally, a threshold (pixel values greater than 150) was applied to separate white cotton pixels, and a zonal histogram tool was used to obtain white pixel counts from each ROI to calculate CFI.

2.2.4 Vegetation indices

Based on the previous literature, 20 different vegetation indices (VIs) that showed strong relationships with leaf area index, biomass, chlorophyll content, plant height, canopy cover, and lint yield of cotton were selected and computed using various combinations of different multispectral bands. These indices are presented in Table 3. Further, the relationship of these VIs along with the individual multispectral reflectance bands (green, blue, red, red-edge, and near-infrared) with functional yield-governing parameters in cotton were examined.

2.3 Model development and statistical analysis

2.3.1 Variable selection

As noted in section 2.2.1, the response (independent) variables included measured physiological parameters such as $IPAR_f$,

above-ground biomass, and HI, as well as RUE_n which is expressed as biomass produced (g) per PAR intercepted (MJ) by the canopy in reference to the first sampling date during the growing seasons, and vegetation index as the predictor (dependent variable). Initially, individual scatter plots were created for 20 different vegetation indices using pooled data from both growing seasons, with $IPAR_f$ and above-ground biomass on the Y-axis versus VIs derived from multispectral imagery throughout the growing season on the X axis (Supplementary Figures 2, 3). Similarly, for RUE, RUE_n was plotted on the Y-axis versus the average of VIs for the period during which RUE_n was determined on the X-axis (Supplementary Figure 4). These plots between $IPAR_f$, above-ground biomass, and VIs revealed a non-linear relationship with higher variation for $IPAR_f$ and above-ground biomass as VI values increased. The non-linear association between $IPAR_f$ and biomass indicated the necessity for an additional predictor in addition to the VIs. In a study to predict above-ground wheat biomass throughout the season, Li et al. (2022) suggested a strong linear relationship of VI's with biomass at each GDD, and the overall relationship throughout the season evolved with GDD. Similarly, changes were seen in our study where the slope and intercept of standard linear relationships between $IPAR_f$ or above-ground biomass and VIs changed at different GDD throughout the season (Figure 3). Therefore, along with VI, GDD was added as a dependent variable (predictor) of $IPAR_f$, and biomass throughout the season in the model as shown in equation (4). In contrast, the association between RUE_n and the average of VIs was found to be linear with constant variance and the average of VI during a specified time in the season was used as a single predictor of RUE

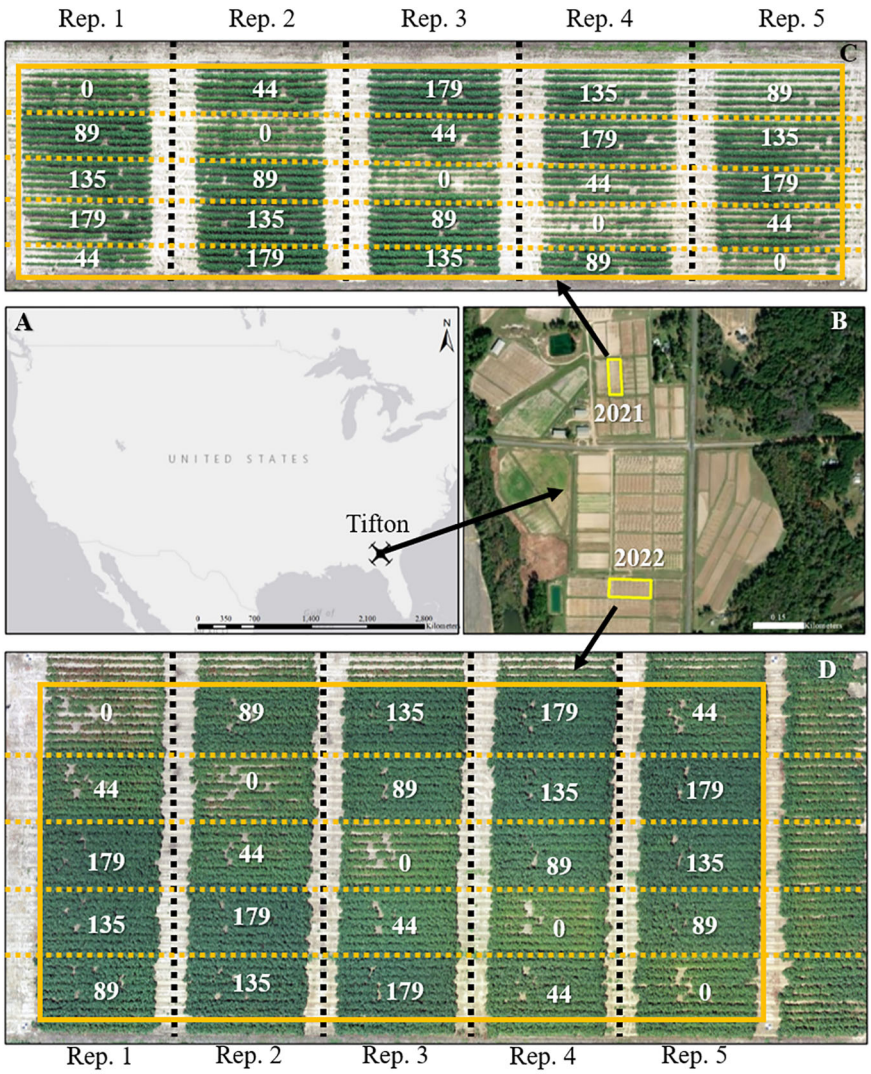


FIGURE 1
Geographical location of the study sites for the 2021 and 2022 growing seasons **(A, B)**. Aerial imagery of the field sites and layout of the experimental plots during the 2021 **(C)** and 2022 **(C)** growing seasons. A solid yellow box depicts the experimental area; black dashed lines separate replications, and yellow dashed lines separate each plot. Values in **(C, D)** are the nitrogen application rates in kg N ha^{-1}.

(equation 5).

$$\text{IPAR}_f \text{ or above-ground biomass} = f(\text{instantaneous VI and GDD}) \quad (4)$$

$$\text{RUE}_n = f(\text{average of VI within the RUE calculation period}) \quad (5)$$

2.3.2 Model selection

For model development, generalized linear regression (GLR) modeling was performed which assumes that the response variables can have a variety of distributions depending on their characteristics and the predictors. Generalized linear models (GLMs) from GLR consist of three components: the random component, the systematic component, and the link function to connect the random and systematic components (Nelder and Wedderburn, 1972; Dunn and Smyth, 2018). The random component represents the distribution of the response variable given the predictors. $IPAR_f$ is a fractional value that always lies between 0 and 1. As a result, the beta distribution with a logistic link function, which is effective for continuous data ranging from 0 to 1, was chosen as a random component for $IPAR_f$. Biomass throughout the season is a positive value; therefore, gamma distribution that includes an exponential relationship with a log link function was chosen as a random component for above-ground biomass. RUE seems to be linearly related to the average of VIs with constant variance; therefore, normal distribution with identity link function (equivalent to standard linear regression) was chosen as the random component. The systematic component of GLM represents the linear predictors which were identified in equations 4 and 5.

2.3.3 Model validation

The 2021 and 2022 growing seasons had different planting and harvesting times, resulting the cotton plants with seasonal variances in growth and development that can be captured in the UAV imagery. As a result of distinct growth patterns, all data collected throughout the season from both growing seasons for $IPAR_f$, RUE, and biomass were pooled together, and the pooled data were randomly stratified by the sampling date in JMP® Pro 16.0.0 (SAS, Cary, NC) where 60% of the data (training data) was used for model generation and the remaining 40% (validation data) for cross-validation. Outliers were identified using histogram, boxplot, and interquartile range methods and were removed from the data before generating training models. There were only two data points identified as outliers, which could have been introduced by noise during reflectance measurement. Cross-validation was performed for each model by plotting predicted versus measured fit for the validation dataset. The models generated from GLR using different VIs were ranked based on the generalized coefficient of determination (R^2), Akaike information criterion value (AICc), and Bayesian information criterion value (BIC) for training dataset, and the coefficient of determination (R^2cv) and root mean sum square error (RMSEcv) for cross-validation. Higher R^2

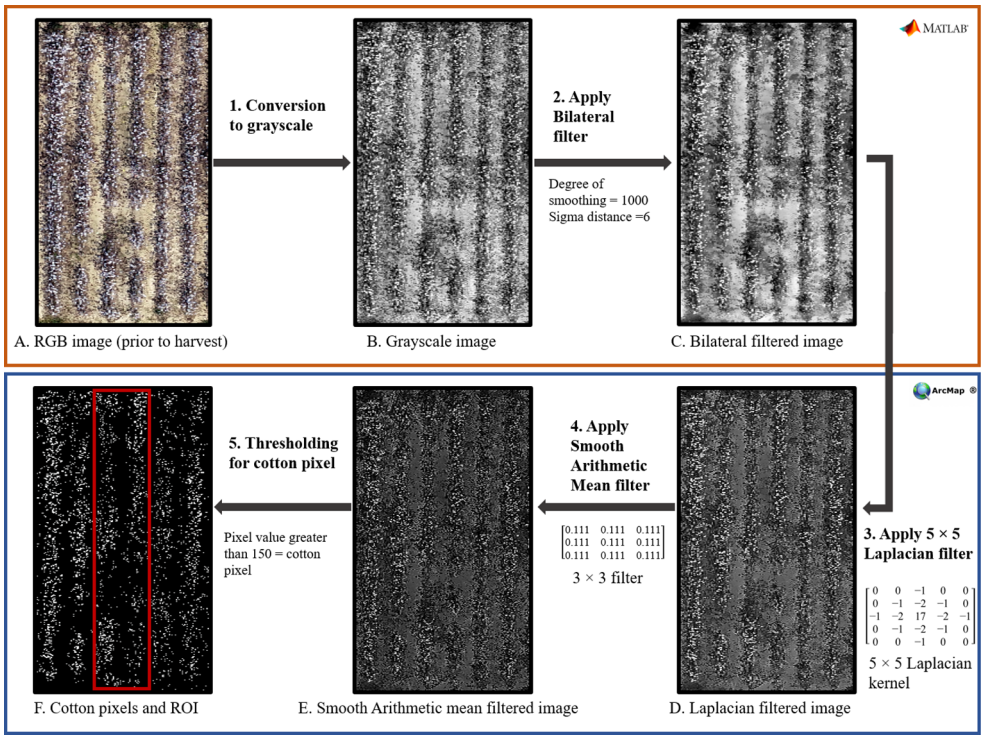


FIGURE 2
Flowchart to illustrate the different steps (1-5) followed for computing cotton fiber index (CFI) from aerial RGB images using various image filtering and enhancing tools in Matlab® and ArcMap®. The top orange box (Steps 1 and 2; Images **A–C**) shows the steps performed in Matlab®, and the bottom blue box (Steps 3, 4, and 5; Images D–F) shows the steps performed in ArcMap®.

and R^2cv values along with lower AICc, BIC, and RMSEcv values were viewed as better model performance qualities.

HI was estimated as the ratio of predicted lint yield to predicted above-ground biomass. For lint yield estimation, standard linear regression was performed between the machine-harvested lint yield and CFI. The cross-validation for lint yield modeling could not be performed due to the limited number of data points (N=50). Finally, the estimated HI was compared to the observed HI (Section 2.3.1) to obtain the coefficient of determination (R^2) and RMSE values. All modeling and statistical analyses were performed in JMP® Pro 16.0.0 (SAS, Cary, NC) and graphs were prepared using Sigma Plot 14.0 (Systat Software Inc., San Jose, CA).

3 Results

3.1 Fraction of intercepted photosynthetically active radiation

The $IPAR_f$ models generated using training data had generalized R^2 values ranging from 0.66 to 0.90 (Table 4). The model equations are provided in the Supplementary Table 1. Upon comparing the model performance for both training and cross-validation, RVI, RECI, NDRE, and SCCCI were the top four VIs with the highest R^2 and R^2cv values, and lowest AICc, BIC, and RMSEcv values. The predicted $IPAR_f$ from these four models was able to explain 89% to 93% of the variation in measured $IPAR_f$ with RMSEcv ranging from 0.080 to 0.097 during cross-validation (Figure 4). Compared to the VIs, models from near-infrared and red reflectance bands explained 85% and 87% of measured $IPAR_f$; however, the RMSEcv was higher for these raw bands (0.139, and 0.115, respectively).

3.2 Above-ground biomass

For above-ground biomass, models developed from the training dataset had generalized R^2 values ranging from 0.69 to 0.87 indicating a moderate to strong relationship with the VIs (Table 5). These model equations are provided in Supplementary Table 2. Based on the model performance, MSAVI, OSAVI, RVI, and SAVI were the four VI-based models with the highest R^2 and R^2cv values, and the lowest AICc, BIC, and RMSEcv values. The estimated above-ground biomass from these models was able to explain 83-84% of variation with RMSEcv ranging from 240.64 to 251.34 g m^{-2} during cross-validation (Figure 5). GNDVI and NIR/G had the highest R^2 values of 0.86 and 0.87, respectively, during training, but the R^2 values reduced to 0.75 and 0.76, respectively during cross-validation. Out of the five raw bands, the model with

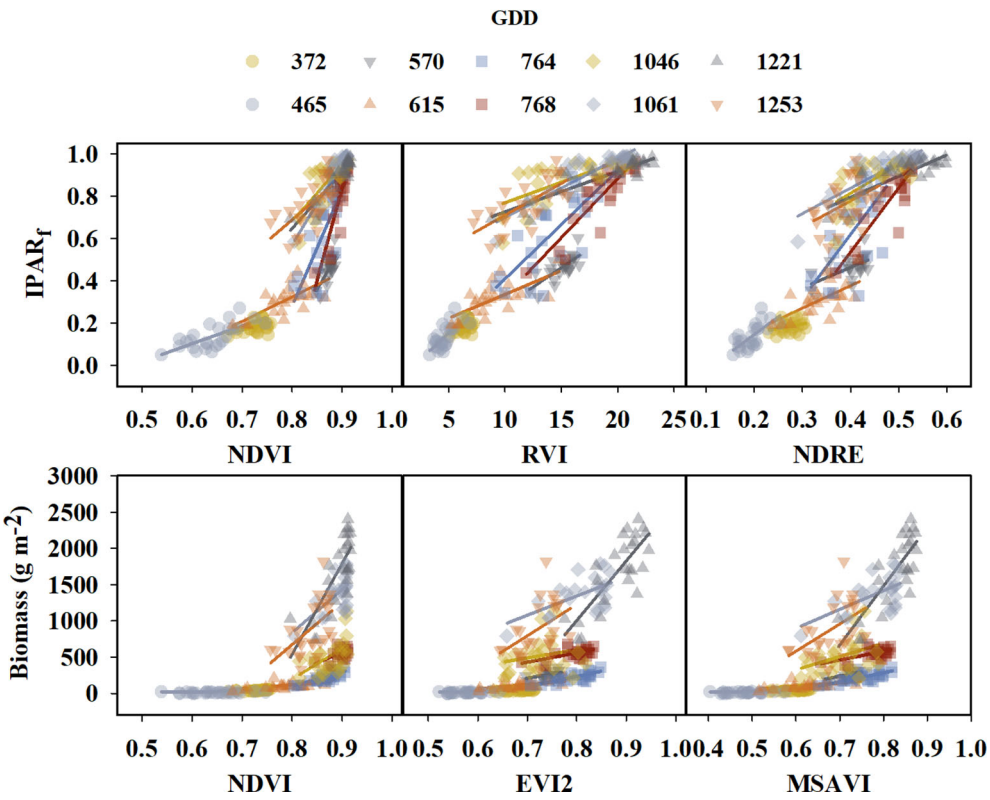


FIGURE 3
Graphs illustrating linear relationship between the fraction of light interception ($IPAR_f$) and above-ground biomass for few selected VIs at different growing degree day (GDD). Different colors and symbols represent different GDDs and are specified in the legend at the top of the graph. Solid lines represent linear regression functions.

the NIR band was able to explain 79% of the variation in the measured above-ground biomass during cross-validation.

3.3 Radiation use efficiency

The models created using training data to predict RUE explained 2% to 43% of the variation in RUE indicating a poor to low relationship of VIs to the RUE (Table 6). The model equations are provided in Supplementary Table 3. While some of the VIs such as WDRVI, OSAVI, MSAVI, and TCARI explained as much as 43% variation in the RUE for the training data, they did not perform well during validation (Rcv = 25-33%). RECI, NIR/G, NDRE, and SCCCI performed best during training and cross-validation with lower AICc, BIC, and RMSEcv values, and the highest R^2cv value. A plot of predicted versus measured RUE using validation data (Figure 6) showed that the above-mentioned models explained 37% to 40% of the variation in cotton RUE with around 0.4 g MJ^{-1} RMSEcv. Among the five raw reflectance bands, the RUE prediction model using the red-edge and green bands explained 37% and 34%, respectively, of variation in RUE during cross-validation.

Based on the model identified for $IPAR_f$ and biomass in sections 3.1 and 3.2, the models that performed the best during validation were used to derive RUE using equation 3. The estimated RUE from MSAVI-based biomass model and RVI-based $IPAR_f$ model only explained 18% of the variation in measured RUE with an RMSE of 0.462 (Figure 7A). However, replacing the MSAVI-based biomass model with the ground-measured biomass increased the prediction accuracy to 84% with the RMSE of 0.207 (Figure 7B).

TABLE 4 Model performance parameters for predicting fraction of Intercepted Photosynthetically Active Radiation ($IPAR_f$) using different vegetation indices (VIs) and raw reflectance bands for the studies conducted in the 2021 and 2022 growing seasons (model equations provided in Supplementary Table 1).

VIs	R^2	AICc	BIC	R^2cv	RMSEcv
RVI	0.90	-359.43	-347.69	0.93	0.080
RECI	0.89	-351.13	-339.39	0.92	0.084
NDRE	0.88	-332.31	-320.57	0.91	0.091
SCCCI	0.87	-318.60	-306.87	0.89	0.097
WDRVI	0.85	-297.40	-285.66	0.90	0.099
MSAVI	0.83	-278.31	-266.58	0.91	0.100
NIR/G	0.86	-307.54	-295.80	0.88	0.101
OSAVI	0.82	-275.32	-262.59	0.90	0.103
SAVI	0.81	-261.05	-249.31	0.90	0.105
VARI	0.82	-269.98	-258.25	0.90	0.106
NDVI	0.82	-272.21	-260.47	0.88	0.107
RE/R	0.82	-270.38	-258.65	0.89	0.107
EVI2	0.79	-246.48	-234.74	0.89	0.109
EVI	0.79	-247.96	-236.22	0.89	0.110
GRVI	0.80	-257.86	-246.12	0.89	0.111
REDEDGE	0.83	-271.05	-259.37	0.84	0.111
GNDVI	0.81	-265.86	-254.13	0.86	0.113
RED	0.81	-255.02	-243.34	0.87	0.115
RGBVI	0.79	-247.66	-235.93	0.87	0.115
TCARI	0.77	-245.89	-234.15	0.87	0.116
TCARI/OSAVI	0.78	-240.74	-229.01	0.86	0.119
GREEN	0.77	-227.54	-215.86	0.82	0.126
BLUE	0.73	-207.82	-196.14	0.82	0.134
NIR	0.70	-191.24	-179.56	0.85	0.139
ExG	0.66	-178.21	-166.48	0.77	0.148

R^2, AICc, and BIC represent the generalized coefficient of determination, Akaike Information Criterion and Bayesian Information Criterion values, respectively, for the training data (N = 149). R^2cv and RMSEcv represent the coefficient of determination and root mean sum square error, respectively, for cross-validation (N = 100).

3.4 Lint yield and harvest index

Cotton fiber Index (CFI) from UAV-based RGB imagery showed a linear relationship with machine-harvested lint yield and explained 69% of the variation (RMSE = 244.07 kg ha^{-1}) in the lint yield (Figure 8). However, the estimated harvest index (HI) utilizing the CFI-based lint yield model and the four models that performed the best in predicting above-ground biomass (Figure 9) wasn't significant (p-value > 0.05) in explaining measured HI. A plot of predicted versus measured HI for the two growing seasons (2021 and 2022) (Figure 9) shows that the estimated HI was able to explain 40% to 49% of the variation in cotton HI, for the 2022 growing season, while only 1% to 4% for the 2021 growing season.

4 Discussion

The yield-driving physiological parameters – intercepted photosynthetically active radiation (IPAR), radiation use efficiency (RUE), and harvest index (HI), are the key contributors to yield in cotton under different management conditions (Conaty and Constable, 2020). However, the current methods of ground measurement for these parameters are resource intensive and require destructive sampling of plants. Remote sensing using UAVs and low-cost sensors have proven to be an efficient approach in estimating crop biophysical traits such as biomass, chlorophyll content, and leaf area index (Gutierrez et al., 2012; Raper and Varco, 2014; Wang et al., 2021). Therefore, this study investigates the use of low-cost and high-resolution UAV-based RGB and multispectral imagery to predict the aforementioned yield-driving physiological parameters [IPAR, RUE, and HI] of cotton.

4.1 Estimation of $IPAR_f$, RUE, and biomass

The first objective of this study was to develop and validate models to estimate the fraction of IPAR ($IPAR_f$), RUE, and biomass of cotton throughout the season using reflectance data (which includes multispectral vegetation indices (VIs) and raw reflectance bands) and growing degree days (GDD). Past research has demonstrated that the relationship between crop biophysical characteristics and VIs can be altered by growth stage (Gutierrez et al., 2012; Li Z. et al., 2022). For example, during cotton growth and development, biomass distribution, specific leaf nitrogen and chlorophyll contents in the canopy change, resulting in varied spectral responses throughout the season (Snider et al., 2021b). Gutierrez et al. (2012) showed the need of separate models to explain above-ground biomass and LAI of cotton plants at different days after planting. In this study, the inclusion of the GDD term as a predictor helped in addressing the above limitation and allowed us to directly account for the changes in the spectral reflectance of cotton plants at different growth stages during the growing season.

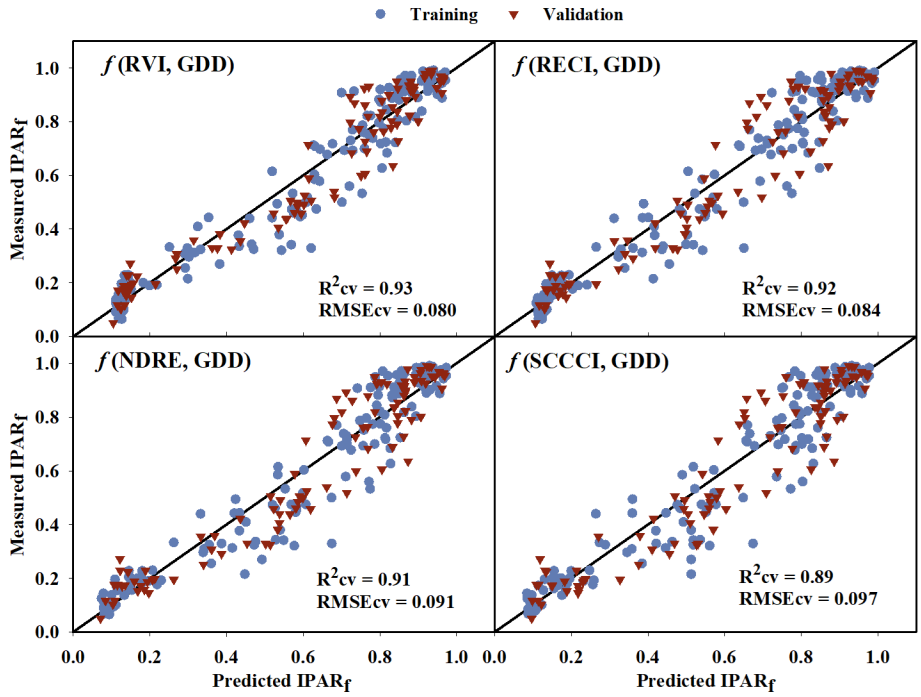


FIGURE 4
Predicted versus measured fraction of Intercepted Photosynthetically Active Radiation ($IPAR_f$) for the four models that performed the best in predicting $IPAR_f$. Blue circles represent the training data, and red triangles represent the validation data. R^2cv and RMSEcv represent the coefficient of determination and root mean sum square error, respectively, for cross-validation. The diagonal line is a reference line with a slope equal to 1. GDD, Growing Degree Days; RVI, Ratio Vegetation Index; RECI, Red-edge Chlorophyll Index; NDRE, Normalized Difference Red-edge Index; SCCCI, Simplified Canopy Chlorophyll Content Index.

The proposed model of VIs along with GDD can be used to estimate the $IPAR_f$ and aboveground biomass throughout the season.

During cross-validation, the developed models explained up to 93% of the variation in measured $IPAR_f$ of the cotton canopy (Table 4), where RVI, RECI, NDRE, and SCCCI had the highest coefficient of determination (Figure 4). Past studies have relied on either a physical radiative transfer model (Clevers et al., 1989) or hyperspectral VIs (Tan et al., 2013; Tan et al., 2018) to estimate the fraction of light absorption by crop canopies. However, there is limited information on the relationship between multispectral VIs and $IPAR_f$. The amount of light intercepted by a cotton canopy is highly influenced by canopy structure and architectural traits such as leaf area index (Brodrick et al., 2013; Bhattacharya, 2019). Multiple studies have reported the association between cotton leaf area index (LAI) and multispectral VIs (Hatfield et al., 1984; Zhao et al., 2007b; Gutierrez et al., 2012), and insights on the relationship between $IPAR_f$ and VIs can be drawn from these studies. RVI, the ratio of near-infrared to red reflectance, has been shown to be effective (R^2 = 0.69 to 0.93) in explaining cotton LAI (Zhao et al., 2007a; Gutierrez et al., 2012; Chen, 2019). In the visible and near-infrared wavelength region, the near-infrared is reflected by the canopy and is sensitive to green photosynthetically active vegetation, whereas the red band is strongly absorbed by chlorophyll, making RVI a good predictor of LAI (Tucker, 1979). In this study, models based on raw bands of near-infrared and red reflectance were also able to moderately estimate $IPAR_f$ throughout the growing season, further highlighting the importance of these two reflectance bands in explaining the variability in $IPAR_f$. NDVI, which is also a function of near-infrared and red reflectance bands, has been used to predict cotton LAI and canopy cover (Ritchie et al., 2010; Adams et al., 2021). Guillen-Climent et al. (2012) also reported that out of several VIs, NDVI was best correlated with

TABLE 5 Model performance parameters for predicting above-ground biomass using different vegetation indices (VIs) and raw reflectance bands for the studies conducted in the 2021 and 2022 growing seasons (model equations provided in Supplementary Table 2).

Vis	R^2	AICc	BIC	R^2cv	RMSEcv
MSAVI	0.87	1896.01	1907.75	0.84	240.64
OSAVI	0.87	1895.28	1907.01	0.84	243.22
RVI	0.86	1904.01	1915.75	0.84	246.96
SAVI	0.86	1909.57	1921.31	0.83	251.34
EVI	0.84	1924.82	1936.56	0.82	260.19
EVI2	0.84	1924.57	1936.30	0.82	260.26
WDRVI	0.86	1906.92	1918.66	0.81	263.85
NIR	0.82	1910.56	1922.24	0.79	267.15
NDVI	0.85	1917.75	1929.48	0.79	278.17
NDRE	0.85	1915.79	1927.53	0.79	280.39
VARI	0.80	1958.09	1969.83	0.78	284.61
RECI	0.84	1925.15	1936.89	0.78	286.98
GRVI	0.80	1962.05	1973.78	0.77	290.94
SCCCI	0.84	1928.49	1940.23	0.77	293.92
NIR/G	0.87	1896.81	1908.55	0.76	295.95
RE/R	0.82	1939.01	1950.75	0.76	300.85
RED	0.82	1903.43	1915.11	0.70	303.97
GNDVI	0.86	1910.01	1921.75	0.75	305.15
TCARI	0.82	1942.01	1953.75	0.70	330.58
TCARI/OSAVI	0.82	1944.09	1955.83	0.70	333.46
RGBVI	0.78	1971.23	1982.97	0.69	339.64
BLUE	0.78	1935.68	1947.36	0.62	344.66
REDEDGE	0.77	1943.90	1955.58	0.61	348.08
GREEN	0.78	1938.76	1950.44	0.56	362.22
ExG	0.69	2022.66	2034.39	0.56	405.25

R^2, AICc, and BIC represent the generalized coefficient of determination, Akaike Information Criterion and Bayesian Information Criterion values, respectively, for the training data (N = 147). R^2cv and RMSEcv represent the coefficient of determination and root mean sum square error, respectively, for cross-validation (N = 100).

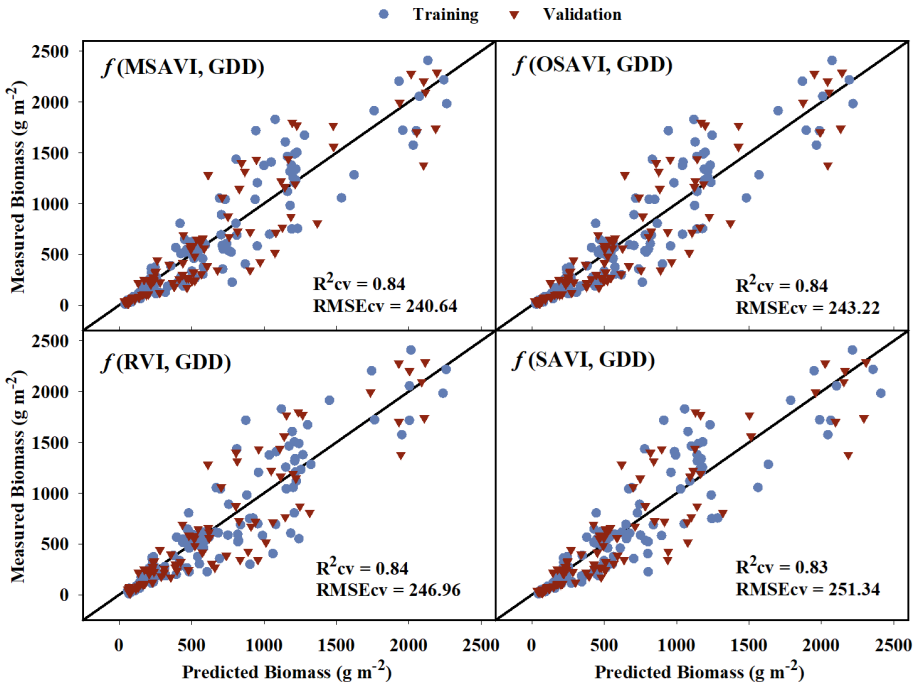


FIGURE 5
Predicted versus measured above-ground biomass for the four models that performed the best in predicting above-grond biomass models. Blue circles represent the training data, and red triangles represent the validation data. R^2cv and RMSEcv represent the coefficient of determination and root mean sum square error, respectively, for cross-validation. The diagonal line is a reference line with a slope equal to 1. GDD, Growing Degree Days; MSAVI, Modified Soil Adjusted Vegetation Index; OSAVI, Optimized Soil Adjusted Vegetation Index; RVI, Ratio Vegetation Index; SAVI, Soil Adjusted Vegetation Index.

$IPAR_f$ for crops with homogenous canopies such as wheat, maize, and soybean. However, in the current study, NDVI saturated after the third sampling date (800 GDD) with strong absorption of red wavelength light by the canopy (red reflectance moved closer to zero) and couldn't explain the variation in $IPAR_f$ during the latter part of the season. A similar trend of NDVI saturating at high LAI was reported by Chen (2019) and Gutierrez et al. (2012). The other best performing VIs, RECI, NDRE, and SCCCI, are also reported to explain LAI and canopy chlorophyll content. RECI has been found to be closely related to maize LAI (R^2 = 0.95), even at higher LAI values (0 to > 6), without reaching saturation. Similarly, RECI performed the best without saturation even at higher GDD in this study. This is because RECI is based on near-infrared and red-edge reflectance, and the red-edge wavelength of light is not strongly absorbed by chlorophyll compared to the red wavelength (Gitelson et al., 2003b). In a similar manner, NDRE, and SCCCI are also based on near-infrared and red-edge reflectance and have been found to be associated with leaf and canopy chlorophyll content (Gitelson et al., 2003a; Raper and Varco, 2014; Ballester et al., 2017), which explains their effectiveness in explaining $IPAR_f$ in this study.

The developed models using reflectance data and GDD were able to explain 55% to 84% of the variation in above-ground biomass for the validation dataset (Table 5). MSAVI, OSAVI, RVI, and SAVI were the best performing VIs with the highest coefficient of determination (Figure 5). The near-infrared band explained the most variation (79%) in above-ground biomass out of the five raw bands. For crops and other vegetation, the red band is highly absorbed by chlorophyll, whereas the near-infrared is reflected from the outer leaf surfaces and spongy mesophyll cells. (Jackson, 1986). The VIs computed based on these two bands, such as NDVI and RVI, have been linked to above-ground biomass previously (Gitelson et al., 2003a; Zhao et al., 2005; Zhao et al., 2007b; Chao et al., 2019). RVI is one of the best VIs to explain above-ground biomass in this study, which is similar to previous findings. NDVI is also widely used VI that has been extensively related to above-ground biomass. However, it tends to saturate early with increasing biomass, and is influenced by soil reflectance, particularly under low vegetation cover (Huete, 1988; Qi et al., 1994; Gutierrez et al., 2012). VIs such as SAVI, MSAVI, and OSAVI are all modifications of NDVI which are more sensitive to changes in vegetation cover and are less influenced by soil reflectance during the lag phase of vegetative growth (Rouse, 1974; Huete, 1988; Qi et al., 1994; Rondeaux et al., 1996), which explain their strong relationship with biomass throughout the season in the current study. Previous studies have also reported RVI (Gutierrez et al., 2012), MSAVI (Brandão et al., 2015), and OSAVI (Zhao et al., 2007b) to be a good predictor of cotton above-ground biomass at early and peak growth stages. However, recent studies have shifted their attention towards use of hyperspectral (Junhua et al., no date) and LIDAR (Sun et al., 2018; Furbank et al., 2019) sensors but these are currently expensive and require extensive data analysis for crop aboveground biomass estimation. In our study, we noticed that with increasing GDD, the estimated above-ground biomass values moved further away from measured values. This source of error

and uncertainty in the biomass model could be attributed to the indeterminate growth habit of cotton because the above-ground biomass during the later season is a combination of reproductive and vegetative biomass, and the reflectance data cannot fully capture the reproductive structures in the lower canopy.

The VIs developed from multispectral imagery only explained up to 40% of the variation in cotton RUE (Table 6). This could be due to the fact that RUE is affected by various environmental factors and the interaction of different physiological processes, making it difficult to accurately assess using multispectral remote sensing (Furbank et al., 2019). However, the relationship between multispectral VIs and RUE can help in the identification of valuable predictors for future modeling. RECI, NIR/G, NDRE, and SCCCI were the top four VIs in predicting RUE (Figure 6). Similarly, the red-edge and green raw bands explained 37% and 34% of the variation in RUE for the validation data set. All the aforementioned VIs and raw reflectance bands have shown to be sensitive to canopy chlorophyll content for crops including corn and cotton (Tumbo et al., 2002; Gitelson et al., 2003a; Bausch et al., 2008; Ritchie et al., 2010; Raper and Varco, 2014; Ballester et al., 2017). The specific absorption coefficient of chlorophyll for the green and red-edge bands is lower than the red and blue bands, making the VIs based on green and red-edge bands sensitive to changes in chlorophyll content in plant tissues (Gitelson and Merzlyak, 1997; Gitelson et al., 2003b). Because the quantum efficiency of primary photochemistry is influenced by canopy chlorophyll content per unit leaf area (Porcar-Castell et al., 2021), the aforementioned VIs may relate to canopy-level RUE. There are limited studies estimating canopy RUE from multispectral VIs. However, Robles-zazueta et al. (2021) showed potential of

TABLE 6 Model performance parameters for predicting Radiation Use Efficiency (RUE) using different vegetation indices (VIs) and raw reflectance bands for the studies conducted in the 2021 and 2022 growing seasons (model equations provided in Supplementary Table 3).

VIs	R^2	AICc	BIC	R^2cv	RMSEcv
RECI	0.38	123.82	131.84	0.40	0.395
NIR/G	0.41	117.46	125.48	0.38	0.401
NDRE	0.39	121.16	129.17	0.37	0.403
SCCCI	0.35	128.72	136.74	0.37	0.405
REDEDGE	0.30	137.61	145.63	0.37	0.402
GNDVI	0.43	113.26	121.28	0.36	0.406
RVI	0.42	115.29	123.31	0.35	0.409
GREEN	0.36	127.07	135.08	0.34	0.412
WDRVI	0.43	114.37	122.39	0.33	0.417
OSAVI	0.43	114.40	122.42	0.32	0.420
EVI	0.42	115.74	123.76	0.32	0.420
NDVI	0.41	117.03	125.04	0.32	0.420
RED	0.40	120.18	128.2	0.32	0.420
SAVI	0.43	114.00	122.02	0.31	0.421
MSAVI	0.43	113.64	121.66	0.31	0.423
EVI2	0.42	114.91	122.93	0.31	0.424
NIR	0.37	125.27	133.29	0.26	0.438
TCARI/OSAVI	0.42	115.41	123.16	0.25	0.439
VARI	0.37	125.89	133.91	0.25	0.441
TCARI	0.43	113.24	121.26	0.24	0.444
RE/R	0.41	118.24	126.26	0.23	0.447
GRVI	0.35	129.39	137.41	0.22	0.452
BLUE	0.30	137.59	145.61	0.19	0.458
RGBVI	0.34	131.01	139.03	0.18	0.460
ExG	0.02	175.65	183.67	0.06	0.493

R^2, AICc, and BIC represent the coefficient of determination, Akaike Information Criterion and Bayesian Information Criterion values, respectively, for the training data (N = 115). R^2cv and RMSEcv represent the coefficient of determination and root mean sum square error, respectively, for cross-validation (N = 82).

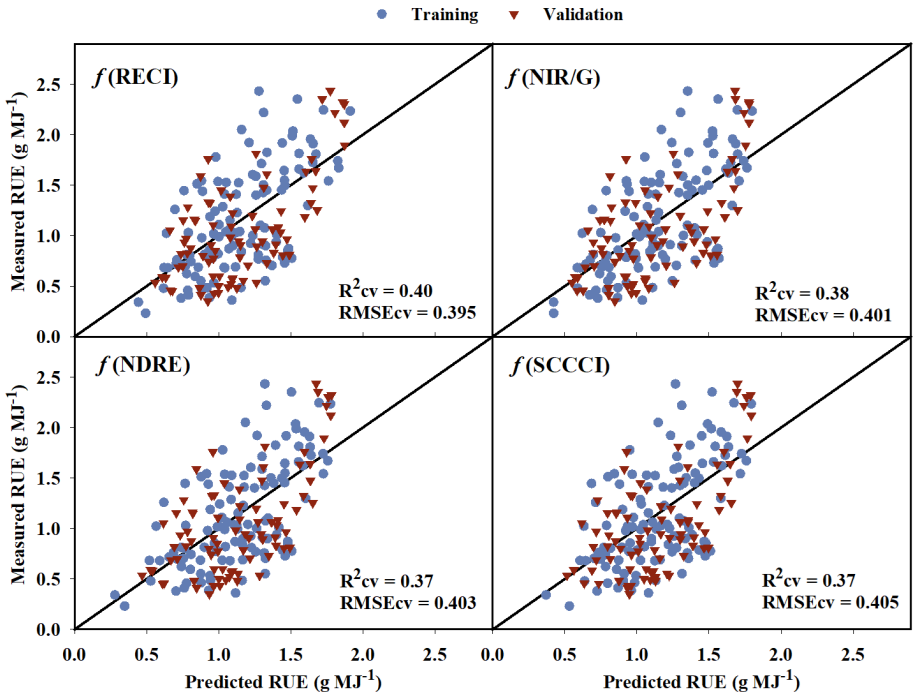


FIGURE 6
Predicted versus measured radiation use efficiency (RUE) for the the four models that performed the best in predicting RUE. Blue circles represent the training data, and red triangles represent the validation data. R^2cv and RMSEcv represent the coefficient of determination and root mean sum square error, respectively, for cross-validation. The diagonal line is a reference line with a slope equal to 1. RECI, Red-edge Chlorophyll Index; NIR/G, Near infrared to Green Ratio; NDRE, Normalized Difference Red-edge Index; SCCCI, Simplified Canopy Chlorophyll Content Index.

hyperspectral VIs and partial least square regression (PSLR) modelling to predict RUE in wheat, where EVI and PRI predicted RUE with 70% accuracy.

Additionally, this study showed the potential of using a mechanistic model (RUE = Biomass/IPAR) from estimated IPAR and biomass to derive RUE in cotton. The estimated RUE from the measured biomass and RVI-based $IPAR_f$ model was able to explain 84% of measured RUE (Figure 7). A similar approach to obtain RUE from measured biomass and remotely sensed light interception has been used in corn (Tewes and Schellberg, 2018) and soybean (Phillips et al., 2020) as well. Further, in this study, actual biomass measurements were replaced with estimates from the MSAVI-based biomass model, which resulted in a reduced prediction accuracy of 18%. The reason for the reduced performance could be due to the error and inaccuracies associated with biomass prediction (Robles-zazueta et al., 2021). As previously discussed, cotton is an indeterminate crop, and its above-ground biomass during the season consists of both vegetative and reproductive structures. Therefore, the spectral VIs may not represent the biomass contributed by the bolls in lower canopy.

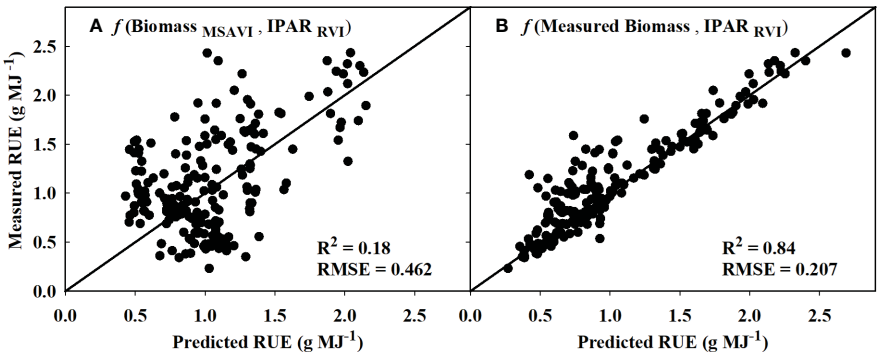


FIGURE 7
Predicted versus measured plot for the radiation use efficiency (RUE) values obtained using **(A)** MSAVI-based biomass model/RVI-based IPAR model and **(B)** measured biomass/RVI-based IPAR model. R^2 and RMSE represent the coefficient of determination and root mean sum square error. The diagonal line is a reference line with a slope equal to 1. IPAR, Intercepted Photosynthetically Active Radiation; MSAVI, Modified Soil Adjusted Vegetation Index; RVI, Ratio Vegetation Index.

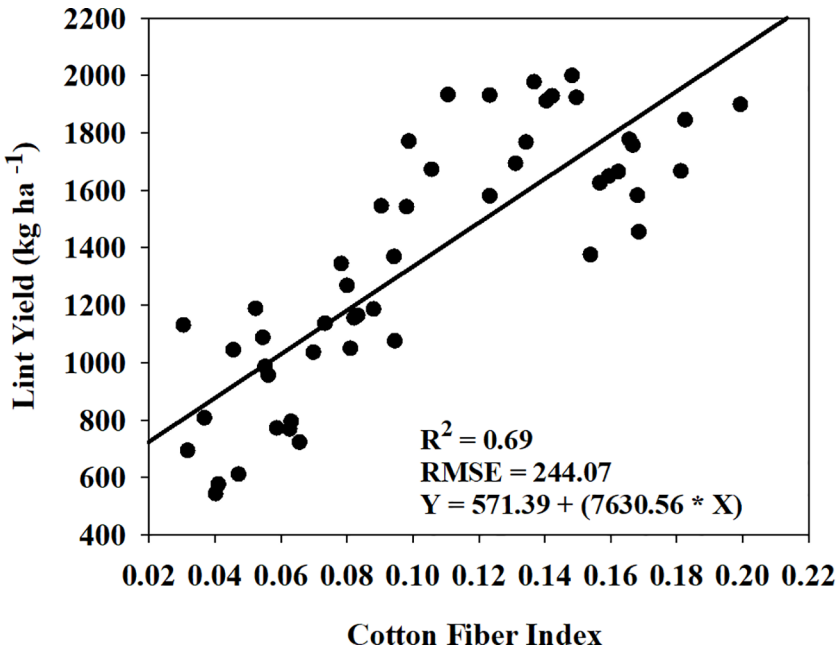


FIGURE 8
Relationship and prediction equation between machine-harvested lint yield and cotton fiber index (CFI) for both the 2021 and 2022 growing seasons. R^2 and RMSE represent the coefficient of determination and root mean square error, respectively, for the relationship. In the given equation, Y represents lint yield and X represents CFI.

Moreover, due to cotton's indeterminate growth pattern, it is challenging to uniformly sample above-ground biomass throughout the season.

4.2 Estimation of lint yield and HI

The second objective of this study was to estimate cotton lint yield using cotton fiber index (CFI). CFI explained 69% of the variation in cotton lint yield (Figure 8), which is somewhat lower than the result reported by Huang et al. (2016) (R^2 = 0.83) and Feng et al. (2020) (R^2 = 0.90). However, the performance of CFI in explaining lint yield variation was better than previous studies that used in-season vegetation indices to predict yield (R^2 = 0.47 to 0.60) (Gutierrez et al., 2012; Ballester et al., 2017). There are a few limitations of using CFI from UAV-based RGB imagery which could introduce some error and possibly the reason for low accuracy observed in our study. CFI does not consider the cotton bolls present in the lower and middle of the canopy (Feng et al., 2020; Siegfried et al., 2023), and there is a potential for misclassification of soil or background pixels as white cotton bolls. Also, the UAV flight height can have an influence on the image resolution for cotton pixel detection. The lower flight height can increase the detection accuracy but can limit the amount of area covered as well as increase time for image processing. To increase the yield-predicting performance of CFI, future research could focus on identifying low cost remote sensing predictors, such as UAV-based plant height or white flower detection during the season, which can account for the cotton bolls in the middle and below the canopy. Utilizing machine learning and convolutional neural networks to accurately delineate cotton pixels from the surrounding soil pixels could also improve yield prediction (Li F. et al., 2022; Shi et al., 2022).

Finally, the last objective of this study was to investigate the utility of the developed models for lint yield and above-ground biomass in predicting cotton harvest index (HI). No significant relationship was observed between the measured and predicted HI when the data from both years was combined. It was also found that the HI models performed differently for each growing season (Figure 9), with the estimated HI explaining 40-49% of the variation in measured HI for the 2022 growing season. The lack of a significant relationship over the two years of the study could be attributed to the combined errors in the lint yield and biomass prediction. As discussed earlier, for an indeterminate crop such as cotton, nadir UAV imagery cannot fully account for the reproductive structure or cotton bolls in the lower and middle canopy that contribute to both above-ground biomass and lint yield (Shi et al., 2022; Siegfried et al., 2023). Furthermore, previous studies have reported varying responses of cotton HI (indifferent to inverse relationship) to different nitrogen levels (Kimball and Mauney, 1993; Gerik et al., 1994; Li et al., 2017), which could be another reason for the varied responses observed in the different seasons.

4.3 Model applications, limitations, and implication for future efforts

The models developed to estimate the yield-contributing physiological parameters in cotton showed low to strong

performance. The models developed in this study can estimate the yield determining physiological parameters throughout the growing season and are relatively simple with only two required input predictors- VI from multispectral imagery and GDD (growth stage parameter). This makes use of these parameters a viable method to measure the light interception and above-ground biomass rapidly and accurately during the cotton growing season and also expands the potential of these models to be used by researchers and industry in cotton management and breeding. Previous work showed that low nitrogen stress in cotton primarily reduces the light interception of the cotton canopy, resulting in significant yield loss (Pokhrel et al., 2023). The models, therefore, can be applied to measure light interception and identify yield-limiting low nitrogen circumstances in-season. The models have ability to aid in agronomic decision making in order to prevent significant yield loss. Similarly, excessive irrigation can cause excessive above-ground biomass resulting in a lower number of cotton bolls and boll mass, ultimately causing significant yield loss (Ermanis et al., 2020). The models from this study can be potentially applied to regulate the irrigation requirement of a cotton plant during the growing season based on their light interception and biomass gain to avoid yield loss due to excess growth. For cotton breeding, the models can be utilized for high-throughput phenotyping of a large number of cotton genotypes for yield determining physiological parameters without the need for labor-intensive and time-consuming manual measurements. Finally, results attained in this study can be used to expand and improve future research on predicting functional yield drivers of cotton from aerial imagery.

One of the limitations of this study is that the modeling and validation were limited to data collected within 372 GDD to 1253 GDD, with no extrapolation, and it would be beneficial to test these models outside this range in future studies. Furthermore, the models were developed based on a two-year study of a single cotton cultivar's response to five different nitrogen application rates at an experimental site. Inclusion of training data from different cotton cultivars across multiple production environments can enhance model performance. It is suggested that future research should focus on increasing the model's robustness for its transferability across a wide range of genotypes and environmental conditions. Independent validation from different locations would also help in determining the model's prediction ability. In the future, more robust models can be created by exposing the crop to a broader range of growth and yield-altering factors. Additionally, due to the limited number of data points, validation for the lint yield modeling was not performed. Validation for the use of CFI in lint yield prediction can be done in the future. There is also the possibility of misclassification while separating cotton white pixels and background soils for CFI calculation. Future research can work on assessing quality of those filtering techniques to avoid the misclassifications. Moreover, top-view UAV imagery from 45 m height may not fully capture the cotton bolls present in the middle section of cotton plants. Future research can investigate the

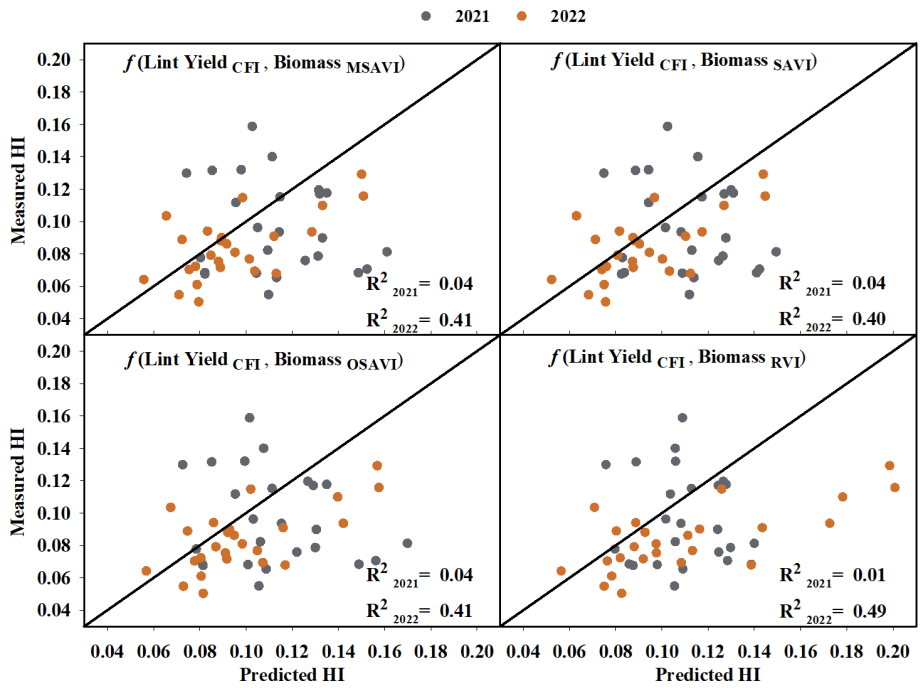


FIGURE 9
Predicted versus measured harvest index (HI) using CFI-based lint yield and the highest seasonal biomass obtained from the four models that performed the best in predicting above-ground biomass. Grey circles represent the 2021 growing season, and orange circles represent the 2022 growing season. R^2_{2021} and R^2_{2022} represent the coefficient of determination for predicted versus measured regression for the 2021 and 2022 growing season, respectively. The diagonal line is a reference line with a slope equal to 1. CFI, Cotton Fiber Index; MSAVI, Modified Soil Adjusted Vegetation Index; OSAVI, Optimized Soil Adjusted Vegetation Index; RVI, Ratio Vegetation Index; SAVI, Soil Adjusted Vegetation Index.

influence of different flight parameters such as flight height and sensor resolution on accurate estimation of cotton lint yield and above-ground biomass prediction. This would potentially aid in increasing the estimation accuracy of the cotton harvest index. Overall, this study is the first attempt of its kind to use low-cost UAV RGB and multispectral imagery to predict yield-determining physiological parameters of cotton throughout the season and the results obtained here shows a strong potential to utilize and expand the use of remotely-sensed imagery for estimating yield-driving functional traits of cotton.

5 Conclusions

The objectives of this study were to predict $IPAR_f$, RUE, and biomass of cotton during the growing season using VIs derived from UAV-based multispectral imagery and GDD, and to estimate cotton HI from biomass models and CFI-based lint yield estimates. Estimated $IPAR_f$ using models based on RVI, RECI, NDRE and SCCCI showed strong relationships with actual $IPAR_f$ values during cross-validation. RUE was best explained by VIs that used red-edge, green, and near-infrared bands such as RECI, NIR/G, NDRE, and SCCCI, which are linked to the chlorophyll content per unit leaf area in past studies. Models based on MSAVI, OSAVI, RVI, and SAVI explained most of the variation in above-ground biomass during cross-validation. CFI had a moderate relationship with the machine-harvested lint yields. Estimated HI based on CFI-based lint yield estimates and biomass models showed no significant relationships with actual HI values and only weak relationships with actual values during the 2022 growing season. Further research towards accurate estimation of lint yield and biomass is recommended to predict cotton harvest index. Thus, we can conclude that UAV-based RGB and multispectral imagery can be utilized to predict some yield-determining physiological parameters in cotton.

Data availability statement

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

Author contributions

Conceptualization: JS, SV, and AP. Data curation: SV and AP. Formal analysis: AP and SV. Funding acquisition: JS and HS. Investigation: AP, SV, VP, DC, JL, and CB. Methodology: SV, JS and AP. Project administration: SV, JS, and AP. Resources: SV, JS and HS. Writing-Original draft: AP, SV, and JS. Writing-review and editing: SV, JS, AP, GV, LH, VP, DC, JL, and CB. All authors contributed to the article and approved the submitted version.

Funding

The authors would like to thank Georgia Cotton Commission (Project 22-447GA) and Cotton Incorporated (grant numbers AWD00013386 and AWD00014749) for their financial support for the research.

Acknowledgments

The authors would like to thank the University of Georgia for the material support on this research. We would also like to thank Lola Sexton, William Vance, Gurpreet Virk, Navneet Kaur, Bailey Lawson, and Ravi Meena for their support in the lab and the field.

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.2023.1248152/full#supplementary-material

References

Adams, C. B., Ritchie, G. L., and Rajan, N. (2021). Cotton phenotyping and physiology monitoring with a proximal remote sensing system. *Crop Sci.* 61 (2), 1317–1327. doi: 10.1002/csc2.20434

Ashapure, A., Jung, J., Chang, A., Oh, S., Maeda, M., and Landivar, J. (2019). A comparative study of RGB and multispectral sensor-based cotton canopy cover modelling using multi-temporal UAS data. *Remote Sens.* 11 (23). doi: 10.3390/rs11232757

Ballester, C., Jiménez-Bello, M. A., Castel, J. R., and Intrigliolo, D. S. (2013). Usefulness of thermography for plant water stress detection in citrus and persimmon trees. *Agric. For. Meteorol.* 168, 120–129. doi: 10.1016/j.agrformet.2012.08.005

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 (11), 1149–1167. doi: 10.3390/rs9111149

Bange, M. P., and Milroy, S. P. (2000). "Assessing effects of canopy nitrogen and light distribution on radiation use efficiency of cotton," in *New frontiers in cotton research. Proceedings of the Second World Cotton Conference, Athens Greece* (Thessaloniki, Greece: P. Petridis Publishers. Athens, Greece), 498–501.

Barbedo, J. G. A. (2019). A review on the use of unmanned aerial vehicles and imaging sensors for monitoring and assessing plant stresses. *Drones* 3 (2), 1–27. doi: 10.3390/drones3020040

Bausch, W. C., Halvorson, A. D., and Cipra, J. (2008). Quickbird satellite and ground-based multispectral data correlations with agronomic parameters of irrigated maize grown in small plots. *Biosyst. Eng.* 101 (3), 306–315. doi: 10.1016/j.biosystemseng.2008.09.011

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 Observ. Geoinform.* 39, 79–87. doi: 10.1016/j.jag.2015.02.012

Bhattacharya, A. (2019). "Radiation-Use efficiency under different climatic conditions," in *Changing Climate and Resource Use Efficiency in Plants* (Academic Press in UK and USA: Elsevier), 51–109. doi: 10.1016/b978-0-12-816209-5.00002-7

Boegh, E., Soegaard, H., Broge, N., Schelde, K., Thomsen, A., Hasager, C. B., et al. (2002). Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. *Remote Sens. Environ.* 81 (2–3), 179–193. doi: 10.1016/S0034-4257(01)00342-X

Bondada, B. R., and Oosterhuis, D. M. (2001). Canopy photosynthesis, specific leaf weight, and yield components of cotton under varying nitrogen supply. *J. Plant Nutr.* 24 (3), 469–477. doi: 10.1081/PLN-100104973

Brandão, Z. N., Sofiatti, V., Bezerra, J. R. C., Ferreira, G. B., and Medeiros, J. C. (2015). Spectral reflectance for growth and yield assessment of irrigated cotton. *Aust. J. Crop Sci.* 9 (1), 75–84.

Brodrick, R., Bange, M. P., Milroy, S. P., and Hammer, G. L. (2013). Physiological determinants of high yielding ultra-narrow row cotton: Canopy development and radiation use efficiency. *Field Crops Res.* 148, 86–94. doi: 10.1016/j.fcr.2012.05.008

Campoy, J., Campos, I., Plaza, C., Calera, M., Jiménez, N., Bodas, V., et al. (2019). Water use efficiency and light use efficiency in garlic using a remote sensing- based approach. *Agric. Water Manage.* 219, 40–48. doi: 10.1016/j.agwat.2019.03.032

Chao, Z., Liu, N., Zhang, P., Ying, T., and Song, K. (2019). Estimation methods developing with remote sensing information for energy crop biomass: A comparative review. *Biomass Bioenergy* 122 (February), 414–425. doi: 10.1016/j.biombioe.2019.02.002

Chen, P. (2019). "'Cotton leaf area index estimation using unmanned aerial vehicle multi-spectral images'," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. 6251–6254.

Chen, P., and Wang, F. (2020). New textural indicators for assessing above-ground cotton biomass extracted from optical imagery obtained via unmanned aerial vehicle. *Remote Sens.* 12 (24). doi: 10.3390/rs12244170

Cheng, Y., Middleton, E. M., Zhang, Q., Huemmrich, K. F., Campbell, P. K. E., Corp, L. A., et al. (2013). Integrating solar induced fluorescence and the photochemical reflectance index for estimating gross primary production in a cornfield. *Remote Sens.* 5, 6857–6879. doi: 10.3390/rs5126857

Clevers, J. G. P. W., Leeuwen, H. J. C. V., and Verhoef, W. (1989). *Estimating APAR by means of vegetation indices: A sesitivity analysis*. The Netherlands.

Conaty, W. C., and Constable, G. A. (2020). 'Factors responsible for yield improvement in new Gossypium hirsutum L. cotton cultivars. *Field Crops Res.* 250 (May 2019), 107780. doi: 10.1016/j.fcr.2020.107780

Deng, L., Mao, Z., Li, X., Hu, Z., Duan, F., and Yan, Y. (2018). UAV-based multispectral remote sensing for precision agriculture: A comparison between different cameras. *ISPRS J. Photogramm. Remote Sens.* 146, 124–136. doi: 10.1016/j.isprsjprs.2018.09.008

Dunn, P. K., and Smyth, G. K. (2018). *Generalized Linear Models with Examples in R*. Eds. R. DeVeaux, S. E. Fienberg and I. Olkin (New York, NY: Springer US). doi: 10.1007/978-1-4419-0118-7

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 (2), 186–196. doi: 10.1111/jac.12453

Falkowski, M. J., Gessler, P. E., Morgan, P., Hudak, A. T., and Smith, A. M. S. (2005). Characterizing and mapping forest fire fuels using ASTER imagery and gradient modeling. *For. Ecol. Manage.* 217 (2–3), 129–146. doi: 10.1016/j.foreco.2005.06.013

Feng, A., Zhou, J., Vories, E. D., Sudduth, K. A., and Zhang, M. (2020). Yield estimation in cotton using UAV-based multi-sensor imagery. *Biosyst. Eng.* 193, 101–114. doi: 10.1016/j.biosystemseng.2020.02.014

Fu, P., Montes, C. M., Siebers, M. H., Gomez-casanovas, N., McGrath, J. M., Ainsworth, E. A., et al. (2022). Advances in field-based high-throughput photosynthetic phenotyping. *J. Exp. Bot.* 73 (10), 3157–3172. doi: 10.1093/jxb/erac077

Furbank, R. T., Jimenez-Berni, J. A., George-Jaeggli, B., Potgieter, A. B., and Deery, D. M. (2019). Field crop phenomics: enabling breeding for radiation use efficiency and biomass in cereal crops. *New Phytol.* 223 (4), 1714–1727. doi: 10.1111/nph.15817

Garbulsky, M. F., Peñuelas, J., Gamon, J., Inoue, Y., and Filella, I. (2011). The photochemical re fl ectance index (PRI) and the remote sensing of leaf, canopy and ecosystem radiation use efficiencies A review and meta-analysis. *Remote Sens. Environ.* 115, 281–297. doi: 10.1016/j.rse.2010.08.023

Gerik, T. J., Jackson, B. S., Stockle, C. O., and Rosenthal, W. D. (1994). Plant nitrogen status and boll load of cotton. *Agron. J.* 86 (3), 514–518. doi: 10.2134/agronj1994.00021962008600030011x

Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *J. Plant Physiol.* 161 (2), 165–173. doi: 10.1078/0176-1617-01176

Gitelson, A. A., Kaufman, Y. J., Stark, R., and Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* 80 (1), 76–87. doi: 10.1016/S0034-4257(01)00289-9

Gitelson, A. A., Gritz, Y., and Merzlyak, M. N. (2003a). 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., Gritz, Y., and Merzlyak, M. N. (2003b). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* 160 (3), 271–282. doi: 10.1078/0176-1617-00887

Gitelson, A. A., and Merzlyak, M. N. (1997). Remote estimation of chlorophyll content in higher plant leaves. *Int. J. Remote Sens.* 18 (12), 2691–2697. doi: 10.1080/014311697217558

Gonzalez-Dugo, V., Zarco-Tejada, P., Nicolás, E., Nortes, P. A., Alarcón, J. J., Intrigliolo, D. S., et al. (2013). Using high resolution UAV thermal imagery to assess the variability in the water status of five fruit tree species within a commercial orchard. *Precis. Agric.* 14 (6), 660–678. doi: 10.1007/s11119-013-9322-9

Guillen-Climent, M. L., Zarco-tejada, P. J., Berni, J. A. J., North, P. R. J., and Villalobos, F. J. (2012). Mapping radiation interception in row-structured orchards using 3D simulation and high-resolution airborne imagery acquired from a UAV. *Precis. Agric.* 13, 473–500. doi: 10.1007/s11119-012-9263-8

Gutierrez, M., Norton, R., Thorp, K. R., and Wang, G. (2012). Association of spectral reflectance indices with plant growth and lint yield in upland cotton. *Crop Sci.* 52 (2), 849–857. doi: 10.2135/cropsci2011.04.0222

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 (2–3), 416–426. doi: 10.1016/S0034-4257(02)00018-4

Hand, C., Culpepper, S., Harris, G., Kemerait, B., Liu, Y., Perry, C., et al. (2022) '*2022 GEORGIA Cotton Production Guide*'. Available at: https://secure.caes.uga.edu/extension/publications/files/pdf/AP124-2_1.PDF.

Hatfield, J. L., Asrar, G., and Kanemasu, E. T. (1984). Intercepted photosynthetically active radiation estimated by spectral reflectance. *Remote Sens. Environ.* 14 (1–3), 65–75. doi: 10.1016/0034-4257(84)90008-7

Hatfield, J. L., and Prueger, J. H. (2010). Value of using different vegetative indices to quantify agricultural crop characteristics at different growth stages under varying management practices. *Remote Sens.* 2 (2), 562–578. doi: 10.3390/rs2020562

Hilker, T., Coops, N. C., Wulder, M. A., Black, T. A., and Guy, R. D. (2008). The use of remote sensing in light use efficiency based models of gross primary production: A review of current status and future requirements. *Sci. Total Environ.* 404, 411–423. doi: 10.1016/j.scitotenv.2007.11.007

Huang, Y. B., Sui, R. X., Thomson, S. J., and Fisher, D. K. (2013). Estimation of cotton yield with varied irrigation and nitrogen treatments using aerial multispectral imagery. *Int. J. Agric. Biol. Eng.* 6 (2), 37–41. doi: 10.3965/j.ijabe.20130602.00?

Huang, Y., Brand, H. J., Sui, R., Thomson, S. J., Furukawa, T., and Ebelhar, M. W. (2016). Cotton yield estimation using very high-resolution digital images acquired with a low-cost small unmanned aerial vehicle. *Trans. ASABE* 59 (6), 1563–1574. doi: 10.13031/trans.59.11831

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 25 (3), 295–309. doi: 10.1016/0034-4257(88)90106-X

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Enviroment* 83, 195–213. doi: 10.1016/S0034-4257(02)00096-2

Hunsaker, D. J., Pinter, P. J., Barnes, E. M., and Kimball, B. A. (2003). Estimating cotton evapotranspiration crop coefficients with a multispectral vegetation index. *Irrigation Sci.* 22 (2), 95–104. doi: 10.1007/s00271-003-0074-6

Jackson, R. D. (1986). Remote sensing of biotic and abiotic plant stress. *Annu. Rev. Phytopathol.* 24, 265–287. doi: 10.1146/annurev.py.24.090186.001405

Jafarbiglu, H., and Pourreza, A. (2022). A comprehensive review of remote sensing platforms, sensors, and applications in nut crops. *Comput. Electron. Agric.* 197 (July 2021), 106844. doi: 10.1016/j.compag.2022.106844

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 (10), 3833–3845. doi: 10.1016/j.rse.2008.06.006

Junhua, B., Shaokun, L., Keru, W., Xueyan, S., and Bing, C. (n.d.). *Estimation models of cotton aboveground fresh biomass based on field hyperspectral remote sensing*, 311–316. Available at: http://europepmc.org/abstract/CBA/636815.

Kimball, B. A., and Mauney, J. R. (1993). Response of cotton to varying CO2, irrigation, and nitrogen: yield and growth. *Agron. J.* 85 (3), 706–712. doi: 10.2134/agronj1993.00021962008500030035x

Kumar, L., Schmidt, K., Dury, S., and Skidmore, A. (2002). "Imaging spectrometry and vegetation science," in *Imaging Spectrometry. Remote Sensing and Digital Image Processing*. Ed. F.D., J. S. M. D. Meer (Springer), 111–155.

Li, C., Knowlton, A., Brown, S., and Ritchie, G. (2011). A comparative study of a microgin with a lab gin stand and commercial gins in southeast United States. *Appl. Eng. Agric.* 27 (2), 167–176. doi: 10.13031/2013.36488

Li, P., Dong, H., Zheng, C., Sun, M., Liu, A., Wang, G., et al. (2017). Optimizing nitrogen application rate and plant density for improving cotton yield and nitrogen use efficiency in the North China Plain. *PloS One* 12 (10), 1–15. doi: 10.1371/journal.pone.0185550

Li, B., Xu, X., Han, J., Zhang, L., Bian, C., Jin, L., et al. (2019). The estimation of crop emergence in potatoes by UAV RGB imagery. *Plant Methods* 15 (1), 1–13. doi: 10.1186/s13007-019-0399-7

Li, F., Bai, J., Zhang, M., and Zhang, R. (2022). Yield estimation of high-density cotton fields using low-altitude UAV imaging and deep learning. *Plant Methods* 18 (1), 1–11. doi: 10.1186/s13007-022-00881-3

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

Lu, S. (2022) *Market Size of the Global Textile and Apparel Industry: 2016 to 2021/2022, FASH455 Global Apparel & Textile Trade and Sourcing*. Available at: https://shenglufashion.com/2018/12/18/market-size-of-the-global-textile-and-apparel-industry-2016-to-2021-2022/ (Accessed 11 September 2022).

Lv, Z., Meng, R., Man, J., Zeng, L., Wang, M., and Xu, B. (2021). Modeling of winter wheat fAPAR by integrating Unmanned Aircraft Vehicle-based optical, structural and thermal measurement. *Int. J. Appl. Earth Observ. Geoinform.* 102, 102407. doi: 10.1016/j.jag.2021.102407

Merrick, T., Jorge, M. L. S. P., Silva, T. S. F., Pau, S., Broadbent, E. N., Bennartz, R., et al. (2020). Characterization of chlorophyll fluorescence, absorbed photosynthetically active radiation, and reflectance-based vegetation index spectroradiometer measurements. *Int. J. Remote Sens.* 41 (17), 6755–6782. doi: 10.1080/01431161.2020.1750731

Moghimi, A., Yang, C., Miller, M. E., Kianian, S. F., and Marchetto, P. M. (2018). A novel approach to assess salt stress tolerance in wheat using hyperspectral imaging. *Front. Plant Sci.* 9, 1–17. doi: 10.3389/fpls.2018.01182

Monteith, J. L. (1972). Solar radiation and productivity in tropical ecosystems. *J. Appl. Ecol.* 9 (3), 747–766. doi: 10.2307/2401901

Nelder, J. A., and Wedderburn, R. W. M. (1972). Generalized linear models. *J. R. Stat. Soc.* 135 (3), 370–384. doi: 10.1016/B978-0-08-044894-7.01331-2

Parkash, V., Snider, J. L., Sintim, H. Y., Hand, L. C., Virk, G., and Pokhrel, A. (2023). Differential sensitivities of photosynthetic processes and carbon loss mechanisms govern N-induced variation in net carbon assimilation rate for field-grown cotton. *J. Exp. Bot.* 74 (8), 2638–2652. doi: 10.1093/jxb/erad038

Pellegrini, P., Sadras, V. O., Oesterheld, M., Bella, C. M., Piñeiro, G., et al. (2020). Simple regression models to estimate light interception in wheat crops with Sentinel-2 and a handheld sensor. *Crop Sci.* 60, 1607–1616. doi: 10.1002/csc2.20129

Phillips, X. A., Kandel, Y. R., Licht, M. A., and Mueller, D. S. (2020). Estimating soybean radiation use efficiency using a UAV in Iowa. *Agronomy* 10 (2002), 1–13. doi: 10.3390/agronomy10122002

Pinter, P. J., Hatfield, J. L., Schepers, J. S., Barnes, E. M., Moran, M. S., Daughtry, C. S.T., et al. (2003). Remote sensing for crop management. *Photogramm. Eng. Remote Sens.* 69 (6), 647–664. doi: 10.14358/PERS.69.6.647

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. doi: 10.1016/j.fcr.2023.108976

Porcar-Castell, A., Malenovský, Z., Magney, T., Van Wittenberghe, S., Fernández-Marín, B., Maignan, F., et al. (2021). Chlorophyll a fluorescence illuminates a path connecting plant molecular biology to Earth-system science. *Nat. Plants* 7 (8), 998–1009. doi: 10.1038/s41477-021-00980-4

Prasad, N. R., Patel, N. R., Danodia, A., and Manjunath, K. R. (2021). Comparative performance of semi – empirical based remote sensing and crop simulation model for cotton yield prediction. *Model. Earth Syst. Environ.*. doi: 10.1007/s40808-021-01180-x

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., and Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sens. Environ.* 48 (2), 119–126. doi: 10.1016/0034-4257(94)90134-1

Raper, T. B., and Varco, J. J. (2014). Canopy-scale wavelength and vegetative index sensitivities to cotton growth parameters and nitrogen status. *Precis. Agric.* 16 (1), 62–76. doi: 10.1007/s11119-014-9383-4

Ritchie, G. L., Bednarz, C. W., Jost, P. H., and Brown, S. M. (2004). *Cotton growth and development* (Georgia, United States: University of Georgia Cooperative Extension Service Tifton, GA). doi: 10.32473/edis-ag235-2005

Ritchie, G. L., Sullivan, D. G., Vencill, W. K., Bednarz, C. W., and Hook, J. E. (2010). Sensitivities of norMalized difference vegetation index and a green/red Ratio index to cotton ground cover fraction. *Crop Sci.* 50 (3), 1000–1010. doi: 10.2135/cropsci2009.04.0203

Robles-zazueta, C. A., Molero, G., Pinto, F., and Foulkes, M. J. (2021). Field-based remote sensing models predict radiation use efficiency in wheat. *J. Exp. Bot.* 72 (10), 3756–3773. doi: 10.1093/jxb/erab115

Rondeaux, G., Steven, M., and Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* 55 (2), 95–107. doi: 10.1016/0034-4257(95)00186-7

Rouse, J. W. (1974). *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation* (NTRS - NASA Technical Reports Server).

Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., et al. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *Eur. J. Agron.* 70, 112–123. doi: 10.1016/j.eja.2015.07.004

Sharma, A., Deepa, R., Sankar, S., Pryor, M., Stewart, B., Johnson, E., et al. (2021). Use of growing degree indicator for developing adaptive responses: A case study of cotton in Florida. *Ecol. Indic.* 124, 107383. doi: 10.1016/j.ecolind.2021.107383

Shi, G., Du, X., Du, M., Li, Q., Tian, X., Ren, Y., et al. (2022). Cotton yield estimation using the remotely sensed cotton boll index from UAV images. *Drones* 6 (9). doi: 10.3390/drones6090254

Siegfried, J., Adams, C. B., Rajan, N., Hague, S., Schnell, R., and Hardin, R. (2023). Combining a cotton "Boll Area Index" with in-season unmanned aerial multispectral and thermal imagery for yield estimation. *Field Crops Res.* 291. doi: 10.1016/j.fcr.2022.108765

Snider, J., Bange, M., and Heitholt, J. (2021b). "Cotton," in *Crop Physiology Case Histories for Major Crops*. Eds. V. O. Sadras and D. F. Calderini (UK and USA: Academic Press), 714–746. doi: 10.1016/B978-0-12-819194-1.00022-0

Snider, J., Harris, G., Roberts, P., Meeks, C., Chastain, D., Bange, M., et al. (2021a). Cotton physiological and agronomic response to nitrogen application rate. *Field Crops Res.* 270. doi: 10.1016/j.fcr.2021.108194

Soil Survey Staff, Natural Resources Conservation Service and United States Department of Agriculture *Web soil survey*. Available at: https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx (Accessed 20 January 2023). (no date).

Sripada, R. P., Heiniger, R. W., White, J. G., and Weisz, R. (2005). Aerial color infrared photography for determining late-season nitrogen requirements in corn. *Agron. J.* 97 (5), 1443–1451. doi: 10.2134/agronj2004.0314

Sui, R., Byler, R. K., and Delhom, C. D. (2017). Effect of nitrogen application rates on yield and quality in irrigated and rainfed cotton. *J. Cotton Sci.* 21 (2), 113–121. doi: 10.56454/XZQP5457

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, 1–17. doi: 10.3389/fpls.2018.00016

Tan, C., Samanta, A., Jin, X., Tong, L., Ma, C., Knyazikhin, Y., et al. (2013). Using hyperspectral vegetation indices to estimate the fraction of photosynthetically active radiation absorbed by corn canopies. *Int. J. Remote Sens.* 34 (24), 8789–8802. doi: 10.1080/01431161.2013.853143

Tan, C., Wang, D., Zhou, J., Du, Y., Luo, M., and Zhang, Y. (2018). Remotely assessing fraction of Photosynthetically Active Radiation (fPAR) for wheat canopies based on hyperspectral vegetation indexes. *Front. Plant Sci.* 9, 1–9. doi: 10.3389/fpls.2018.00776

Tewes, A., and Schellberg, J. (2018). Towards remote estimation of radiation use efficiency in maize using UAV-based low-cost camera imagery. *Agronomy* 8 (1), 1–15. doi: 10.3390/agronomy8020016

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8 (2), 127–150. doi: 10.1016/0034-4257(79)90013-0

Tumbo, S. D., Wagner, D. G., and Heinemann, P. H. (2002). 'Hyperspectral characteristics of corn plants under different chlorophyll levels'. *Trans. ASAEE* 45 (3), 815–823.

USDA (2022). *Cotton: World Markets and Trade*. USDA.

USDA (2023). *Cotton and Wool Outlook: February 2023* (USDA).

USDA Agricultural Marketing Service (2020) *Varieties Planted 2020 Crop*. Available at: https://www.ams.usda.gov/mnreports/cnavar.pdf.

Vatter, T., Gracia-Romero, A., Kefauver, S. C., Nieto-Taladriz, M. T., Aparicio, N., and Araus, J. L. (2021). Preharvest phenotypic prediction of grain quality and yield of durum wheat using multispectral imaging. *Plant J.* 109 (6), 1507–1518. doi: 10.1111/tpj.15648

Voora, V., Larrea, C., and Bermudez, S. (2020). *Global Market Report: Cotton, Sustainable Commodities Marketplace Series 2019*. Canada.

Wajid, A., Ahmad, A., Khaliq, T., Alam, S., Hussain, A., Hussain, K., et al. (2010). Quantification of growth, yield and radiation use efficiency of promising cotton cultivars at varying nitrogen levels. *Pakistan J. Bot.* 42 (3), 1703–1711.

Wang, T., Liu, Y., Wang, M., Fan, Q., Tian, H., Qiao, X., et al. (2021). Applications of UAS in crop biomass monitoring: A Review. *Front. Plant Sci.* 12, 1–16. doi: 10.3389/fpls.2021.616689

Wiegand, C. L., Richardson, A. J., Escobar, D. E., and Gerbermann, A. H. (1991). Vegetation indices in crop assessments. *Remote Sens. Environ.* 35 (2–3), 105–119. doi: 10.1016/0034-4257(91)90004-P

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 (1), 259–269. doi: 10.13031/2013.27838

Wullschleger, S. D., and Oosterhuis, D. M. (1990). Canopy development and photosynthesis of cotton as influenced by nitrogen nutrition. *J. Plant Nutr.* 13 (9), 1141–1154. doi: 10.1080/01904169009364140

Xu, R., Li, C., and Paterson, A. H. (2019). Multispectral imaging and unmanned aerial systems for cotton plant phenotyping. *PloS One* 14 (2), e0205083. doi: 10.1371/journal.pone.0205083

Yang, C., Bradford, J. M., and Wiegand, C. L. (2001). Airborne multispectral imagery for mapping variable growing conditions and yields of cotton, grain sorghum, and corn. *Trans. ASABE* 44 (6), 1983–1994.

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

Zarco-Tejada, P. J., Ustin, S. L., Whiting, M. L., Zarco-Tejada, P. J., Ustin, S. L., and Whiting, M. L. (2005). Temporal and spatial relationships between within-field yield variability in cotton and high-spatial hyperspectral remote sensing imagery. *Agron. J.* 97 (3), 641–653. doi: 10.2134/agronj2003.0257

Zhang, C., Filella, I., Garbulsky, M. F., and Penuelas, J. (2016). Affecting factors and recent improvements of the Photochemical Reflectance Index (PRI) for remotely sensing foliar, canopy and ecosystemic radiation-use efficiences. *Remote Sens.* 8. doi: 10.3390/rs8090677

Zhang, Y. J., Hou, M. Y., Xue, H. Y., Liu, L. T., Sun, H. C., Li, C. D., et al. (2018). Photochemical reflectance index and solar-induced fluorescence for assessing cotton photosynthesis under water-deficit stress. *Biol. Plantarum* 62 (4), 817–825. doi: 10.1007/s10535-018-0821-4

Zhao, D., Huang, L., Li, J., and Qi, J. (2007a). A comparative analysis of broadband and narrowband derived vegetation indices in predicting LAI and CCD of a cotton canopy. *ISPRS J. Photogramm. Remote Sens.* 62 (1), 25–33. doi: 10.1016/j.isprsjprs.2007.01.003

Zhao, D., Reddy, K. R., Kakani, V. G., Read, J. J., and Koti, S. (2007b). Canopy reflectance in cotton for growth assessment and lint yield prediction. *Eur. J. Agron.* 26 (3), 335–344. doi: 10.1016/j.eja.2006.12.001

Zhao, D. H., Li, J. L., and Qi, J. G. (2005). Identification of red and NIR spectral regions and vegetative indices for discrimination of cotton nitrogen stress and growth stage. *Comput. Electron. Agric.* 48 (2), 155–169. doi: 10.1016/j.compag.2005.03.003

10.3389/fpls.2023.1248152

Glossary

AICc	Akaike Information Criterion Value
BIC	Bayesian Information Criterion
CFI	Cotton Fiber Index
EVI	Enhanced Vegetation Index
EVI2	Enhanced Vegetation Index 2
ExG	Excessive Greenness Index
GCPs	Ground Control Points
GDDs	Growing Degree Days
GNDVI	Green Normalized Difference Vegetation Index
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GRVI	Green Red Vegetation Index
HI	Harvest Index
IPAR	Intercepted Photosynthetically Active Radiation
$IPAR_f$	Fraction of Intercepted Photosynthetically Active Radiation
LAI	Leaf Area Index
MSAVI	Modified Soil Adjusted Vegetation Index
N	Nitrogen
NDRE	Normalized Difference Red Edge Index
NDVI	Normalized Difference Vegetation Index
NIR	Near infrared
NIR/G	Near infrared to Green ratio
OSAVI	Optimized Soil Adjusted Vegetation Index
PAR	Photosynthetically Active Radiation
PRI	Photochemical Reflectance Index
R^2	Coefficient of Determination
RE/R	Redegde to Red ratio
RECI	Rededge Chlorophyll Index
RGB	Red Green Blue Composite
RGBVI	Red Green Blue Vegetation Index
RMSE	Root Mean Sun Square Error
ROI	Region of interest
RTK	Real Time Kinematics
RUE	Radiation Use Efficiency
RVI	Ratio Vegetation Index
SAVI	Soil Adjusted Vegetation Index
SCCCI	Simplified Canopy Chlorophyll Content Index
SIF	Sun-induced Fluorescence
TCARI	Transformed Chlorophyll Absorption Reflectance Index
UAV	Unmanned Aerial Vehicle
VARI	Visible Atmospherically Resistant Index
VIs	Vegetation Indices
WDRVI	Wide Dynamic Range Vegetation Index

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 29 September 2023
DOI 10.3389/fpls.2023.1237988

Check for updates

OPEN ACCESS

EDITED BY
Zhenjiang Zhou,
Zhejiang University, China

REVIEWED BY
Zhaoyu Zhai,
Nanjing Agricultural University, China
Haikuan Feng,
Beijing Research Center for Information Technology in Agriculture, China

*CORRESPONDENCE
Shuo Shi
shishuo@whu.edu.cn

RECEIVED 10 June 2023
ACCEPTED 12 September 2023
PUBLISHED 29 September 2023

CITATION
Shi Z, Shi S, Gong W, Xu L, Wang B, Sun J, Chen B and Xu Q (2023) LAI estimation based on physical model combining airborne LiDAR waveform and Sentinel-2 imagery.
Front. Plant Sci. 14:1237988.
doi: 10.3389/fpls.2023.1237988

COPYRIGHT
© 2023 Shi, Shi, Gong, Xu, Wang, Sun, Chen 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.

LAI estimation based on physical model combining airborne LiDAR waveform and Sentinel-2 imagery

Zixi Shi[1], Shuo Shi[1,2,3]*, Wei Gong[1,3,4], Lu Xu[1], Binhui Wang[5], Jia Sun[6], Bowen Chen[1,7] and Qian Xu[1]

[1]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei, China, [2]State Key Laboratory of Geo-Information Engineering, Xi'an, Shaanxi, China, [3]Perception and Effectiveness Assessment for Carbon-Neutrality Efforts, Engineering Research Center of Ministry of Education, Wuhan, Hubei, China, [4]Wuhan Institute of Quantum Technology, Wuhan, Hubei, China, [5]School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan, Hubei, China, [6]School of Geography and Information Engineering, China University of Geosciences, Wuhan, Hubei, China, [7]Chinese Antarctic Centre of Surveying and Mapping, Wuhan University, Wuhan, Hubei, China

Leaf area index (LAI) is an important biophysical parameter of vegetation and serves as a significant indicator for assessing forest ecosystems. Multi-source remote sensing data enables large-scale and dynamic surface observations, providing effective data for quantifying various indices in forest and evaluating ecosystem changes. However, employing single-source remote sensing spectral or LiDAR waveform data poses limitations for LAI inversion, making the integration of multi-source remote sensing data a trend. Currently, the fusion of active and passive remote sensing data for LAI inversion primarily relies on empirical models, which are mainly constructed based on field measurements and do not provide a good explanation of the fusion mechanism. In this study, we aimed to estimate LAI based on physical model using both spectral imagery and LiDAR waveform, exploring whether data fusion improved the accuracy of LAI inversion. Specifically, based on the physical model geometric-optical and radiative transfer (GORT), a fusion strategy was designed for LAI inversion. To ensure inversion accuracy, we enhanced the data processing by introducing a constraint-based EM waveform decomposition method. Considering the spatial heterogeneity of canopy/ground reflectivity ratio in regional forests, calculation strategy was proposed to improve this parameter in inversion model. The results showed that the constraint-based EM waveform decomposition method improved the decomposition accuracy with an average 12% reduction in RMSE, yielding more accurate waveform energy parameters. The proposed calculation strategy for the canopy/ground reflectivity ratio, considering dynamic variation of parameter, effectively enhanced previous research that relied on a fixed value, thereby improving the inversion accuracy that increasing on the correlation by 5% to 10% and on R^2 by 62.5% to 132.1%. Based on the inversion strategy we proposed, data fusion could effectively be used for LAI

inversion. The inversion accuracy achieved using both spectral and LiDAR data (correlation=0.81, R^2 = 0.65, RMSE=1.01) surpassed that of using spectral data or LiDAR alone. This study provides a new inversion strategy for large-scale and high-precision LAI inversion, supporting the field of LAI research.

KEYWORDS

Leaf Area Index (LAI), remote sensing, full-waveform LiDAR, physical model, forest canopy, GORT model, data fusion

1 Introduction

Leaf area index (LAI) is one of the prime determinants of photosynthesis, which makes it an important quantity controlling physical and biological processes of plant canopies and assessing forest growth potential (Chen and Black, 1992; Barclay and Goodman, 2000). As a fundamental attribute of global vegetation, LAI has been listed as an essential climate variable by the global climate change research community (Fang et al., 2019). The ability to accurately and rapidly acquire LAI is an indispensable component of process-based ecological research facilitating the understanding of gas-vegetation exchange phenomenon at an array of spatial scales from the leaf to the landscape (Zheng and Moskal, 2009).

Remote sensing data provides large-scale, systematic land surface observations consistently over the globe (Pan et al., 2008). With remote sensing technology, LAI can be mainly derived from a variety of sensors including passive optical sensors, active light detection and ranging (LiDAR) instrument, and microwave sensors (Fang et al., 2019). In passive optical sensors, multispectral and hyperspectral sensors provide spectral measurements across the electromagnetic spectrum, which are sensitive to subtle variations in reflected energy and, therefore, have a giant potential for detecting differences in vegetation (Mananze et al., 2018). However, LAI retrieval using multispectral and hyperspectral data has potential problems, such as the low signal-to-noise ratios (SNRs) of some remote sensing data, the "curse of dimensionality" and problems of saturation (Liu et al., 2016). In addition, optical remote sensing is only capable of capturing information from the horizontal canopy, resulting in a lack of information pertaining to vertical canopy (Xu et al., 2022). In active LiDAR instrument, full-waveform LiDAR systems can digitize the entire reflected energy, resulting in complete waveforms from the top of the canopy to the ground which reflect vertical profiles (Lefsky et al., 1999; Mallet and Bretar, 2009). It has been used to estimate LAI based on canopy structure and radiation transfer principles, especially by means of the correlation with the gap fraction (Wang and Fang, 2020). The primary advantage of LAI estimation using full-waveform LiDAR lies in its ability to capture detailed structural information beneath the canopy through the complete energy waveform, thereby mitigating estimation errors stemming from leaf aggregation. However, the high-density data obtained from airborne LiDAR is limited to the measurement range, whereas spaceborne full-waveform LiDAR data is characterized by its substantial volume. Multi-source remote sensing data have their own advantages and disadvantages in LAI estimation. On this basis, researches on using multi-source remote sensing data fusion to estimate LAI is becoming a research hotspot for which it can give full play to the advantages of different remote sensing data (Clevers and vanLeeuwen, 1996; Yang et al., 2011; Ma et al., 2014; Qu et al., 2015).

Existing methods for fusing spectral and LiDAR data are mostly based on empirical models. The empirical model directly relates inputs to outputs by pure statistical means, the advantage of which lies in its simplicity (Weiss et al., 2020). Thomas et al. (2011) estimated LAI with LiDAR and multispectral data by constructing fused LiDAR-optical indices. Ma et al. (2014) combined LiDAR data with the MODIS and MISR products to retrieve canopy height and LAI by multivariate linear regression model and geometric-optical mutual-shadowing (GOMS) model. Luo et al. (2019) estimated maize LAI using the combined hyperspectral imagery and LiDAR pseudo-waveforms by random forest (RF) regression algorithm. Zhang et al. (2022) estimated the LAI of a short-crop using UAS-based SfM and LiDAR point clouds, as well as the spectral information from multispectral imagery. Zhang et al. (2023) used six typical machine learning algorithms to construct prediction models of LAI, among which the XGBoost model showed the best performance. It also showed that the fusion of data could significantly improve the predictive ability of the models. However, the empirical model is strongly grounded with a large amount of statistical data and can only be applied within a relatively localized area because their performance is highly dependent on vegetation types, canopy structures, sensors, and temporal change (Xu et al., 2020). Empirical models, even superior deep learning models, can also suffer from statistical problems such as overfitting (Verrelst et al., 2015; Neinavaz et al., 2016).

In contrast to empirical methods, the physical model can be better generalized and its physical principles is helpful for the analysis of fusion mechanism (Myneni et al., 1997; Verrelst et al., 2019; Kennedy et al., 2020). Based on LiDAR waveform, commonly used models include gap fraction models (homogeneous canopy or clumping-aware canopy) and three-dimensional (3-D) radiative transfer models. Within multiple gap fraction models, gap fraction can be performed by directly computing its aggregate value (Luo et al., 2013; Fieber et al., 2014; Tseng et al., 2016; Jiang et al., 2022) or by considering the vertical accumulation of gaps

within the canopy layers (Ni et al., 1999; Yang et al., 2019). Gap fraction models, with minimal input parameters, streamline LAI computation through forward modeling (Zhao et al., 2011). The geometric-optical and radiative transfer (GORT) model, one of the gap fraction models, capitalizes on the capability of lidar waveform data to characterize the underlying canopy structure (Ni-Meister et al., 2001). Accurate vertical profiles of LAI can be derived by GORT (Tang et al., 2012). This methodology has been employed in the derivation of products for the spaceborne lidar GEDI and exhibits a strong applicability of large-scale LAI estimation (Tang et al., 2014; Wang et al., 2023). Regarding the 3-D radiative transfer models such as Discrete Anisotropic Radiative Transfer (DART), they are proficient in simulating lidar waveforms effectively (Gastellu-Etchegorry et al., 2016; Gastellu-Etchegorry et al., 2017). However, due to its multitude of model parameters and complex scenarios, time-consuming aspects arise during the inversion process. Taking the above reasons into consideration, we consider conducting a data fusion inversion study based on the GORT model. It is still a challenge to combine spectral imagery and LiDAR waveform for LAI retrieval based on physical models. In LAI estimation, spectral features including sensitive bands' reflectance and spectral indices, play a pivotal role in accuracy (Potithep et al., 2013; Liang et al., 2015). Waveform information such as height, echo energy ratio and leaf coverage are key parameters of LAI estimation using full-waveform LiDAR (Pope and Treitz, 2013; Ma et al., 2015). Improving the physical model to use the above parameters so as to make use of the respective advantages of multi-source data is the key for LAI estimation using both spectral and LiDAR data. In addition, the accuracy of the physical model is susceptible to the initial assignment value of model parameters and the quality of the input data (Houborg et al., 2007). Adjusting the input parameters of the model based on the study area and source data is also an important measure to ensure the accuracy of LAI estimation.

In response to the above problems, the main objectives of this paper are: (1) Developing an estimation strategy based on GORT model to achieve data fusion. (2) Extracting accurate parameters for LAI estimation from both spectral and LiDAR waveform data. (3) Assessing the performance of the joint data fusion for LAI estimation. Specifically, for the first time, we attempt to fuse spectral and waveform data within the GORT model, achieving a joint LAI estimation. We enhance the waveform decomposition method to extract more accurate waveform parameters. Furthermore, on the basis of the existing retrieval, we improve the model input parameter canopy/ground reflectivity ratio (ρ_v/ρ_g) to make it more suitable for the large-scale forest with the heterogeneity of the spectrum. The significance of this study lies in providing novel insights into the fusion of active and passive remote sensing data, thereby contributing to the enhancement of accurate large-scale LAI estimation.

2 Study area and data

The study area is located in Harvard Forest, a 4000 acres forest in Petersham, Mass., which is now among the most studied forests in the world. As a critical node in USA national ecological network (LTER, NEON and ForestGEO), Harvard Forest department gathers and produces various datasets from its ecological scientific researches. At the same time, many remote sensing ecosystem projects choose Harvard Forest as study area, which provide multi-source remote sensing data. For these reasons, we chose the Harvard Forest to carry out remote sensing research of multi-source data.

2.1 LVIS airborne LiDAR data

NASA's Land, Vegetation, and Ice Sensor (LVIS), is an airborne, wide-swath, full waveform imaging laser altimeter system, which emits 1064nm wavelength laser pulses to collect data on surface topography and 3-d structure with medium 25m footprint.

During summer 2021, LVIS operated as a NASA Facility to calibrate and validate the space-based LiDAR sensor GEDI (Global Ecosystem Dynamics Investigation) by conducting overflow ground tracks over the Eastern United States and French Guiana. The LVIS Classic instrument was flown on Gulfstream V at a flight altitude of 41,000', covering Harvard Forest completely on August 6, 2021. The data products of LVIS include Level 1B Geolocated LVIS Waveforms (HDF format) and Level 2 Geolocated Surface Elevation and Height Product (ASCII Text format), from which ecosystem structure parameters can be derived (Blair et al., 1999; Blair and Hofton., 2020).

In this research, we used data products of the LVIS flight on August 6, 2021, obtained from 'https://nsidc.org/data/LVISC1B/versions/1'. From these, we extracted multiple parameters for each pulse into a comprehensive dataset (.csv). The parameters included: laser shot (shotnumber), longitude (lon), latitude (lat), elevation of the highest detected signal (zt), elevation of the lowest detected mode within the waveform (zg), return waveform (rxwave), signal mean noise level (sigmean).

2.2 Sentinel-2 multispectral images

The Copernicus Sentinel-2 mission comprises a constellation of two polar-orbiting satellites. It offers free multi-spectral images with high spatial resolution (four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution). The orbital swath width is 290 km with high revisit time (5 days with 2 satellites under cloud-free conditions which results in 2-3 days at mid-latitudes), which support to accurately monitor land surface changes especially vegetation changes and are beneficial to biophysical indicators estimation (Drusch et al., 2012). The Sentinel-2 data has good temporal and spatial resolution with high-quality. Numerous studies have shown that it provides accurate retrieval of LAI. (Korhonen et al., 2017; Hu et al., 2020; Zhou et al., 2020; Sun et al., 2022).

Sentinel Applications Platform (SNAP), released by European Space Agency (ESA), has been accelerating Earth observation innovation since 2014. It can help to process and analyze

Sentinel-2 imagery. In this experiment, the Sentinel-2 image was processed by the Sentinel-2 Toolbox (S2TBX) in SNAP. We acquired an L2A image of Sentinel-2 covering the Harvard Forest on July 31, 2021 (https://doi.org/10.5270/S2_-znk9xsj). The base map of Figure 1 is the Sentinel-2 RGB image of the study area processed by SNAP. Based on SNAP, we obtained canopy cover map from Sentinel-2 L2A image. Subsequently, using the latitude and longitude information of the laser pulse from the LVIS, we extracted parameters at each pulse point, including spectral reflectance and canopy cover (CC) values.

2.3 Ground based LAI in Harvard Forest

The Harvard Forest Data Archive contains various datasets from scientific research at the Harvard Forest, in which HF150 dataset collects Leaf Area Index at Harvard Forest HEM and LPH Towers since 1998 (https://harvardforest.fas.harvard.edu/harvard-forest-data-archive). Leaf area index is measured with the LAI-2000 canopy analyzer with one LAI sensor made at multiple plots within each forest type - usually 12 plots within the old-growth hemlock forest, and 36 plots on Little Prospect Hill (Orwig and Hadley, 2022). The time, distance from tower, compass direction from tower from geographic north, LAI value of each plot is given in the dataset, which help to correlate exact coordinates and values of plots. The spatial distribution of ground plots is shown in Figure 1.

3 Methods

To better utilize the advantages of spectral imagery and LiDAR waveform, and considering the scale difference between satellite and airborne data, we first designed a LAI estimation strategy based on a physical model GORT, using airborne LiDAR waveform as the main body and satellite spectral data as support. In order to improve the LAI inversion accuracy, we optimized the model input data and parameters: optimizing waveform decomposition algorithm for more precise waveform energy data, improving the method for obtaining the canopy/ground reflectivity ratio as model parameter to obtain values more consistent with the actual research area. Figure 2 shows an overview of the methods used in this paper.

3.1 A fusion strategy proposal based on GORT model deconstruction

In order to perform joint estimation of LAI using both spaceborne multispectral images and airborne LiDAR waveforms, we decomposed the GORT model and developed a data fusion strategy.

The geometric-optical and radiative transfer (GORT) model is for the bidirectional reflectance of a vegetation cover combines principles of geometric optics and radiative transfer (Li et al., 1995). Ni-Meister et al. (2001) developed a method based on the modified

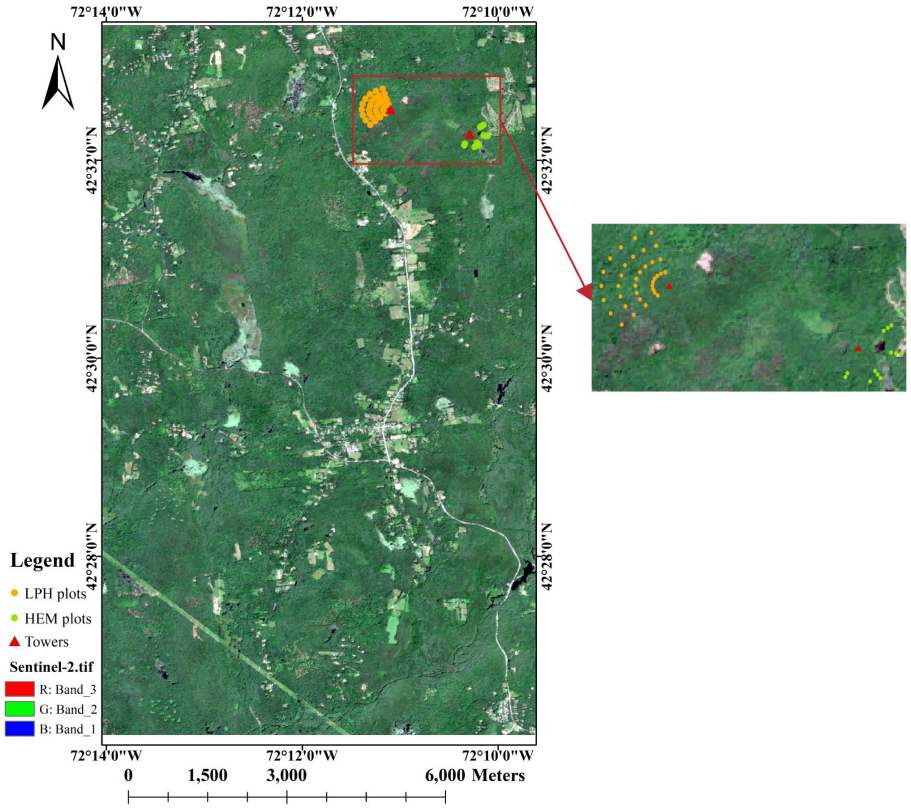


FIGURE 1
The spatial distribution of ground plots in Harvard Forest (The base map is Sentinel-2 RGB image.).

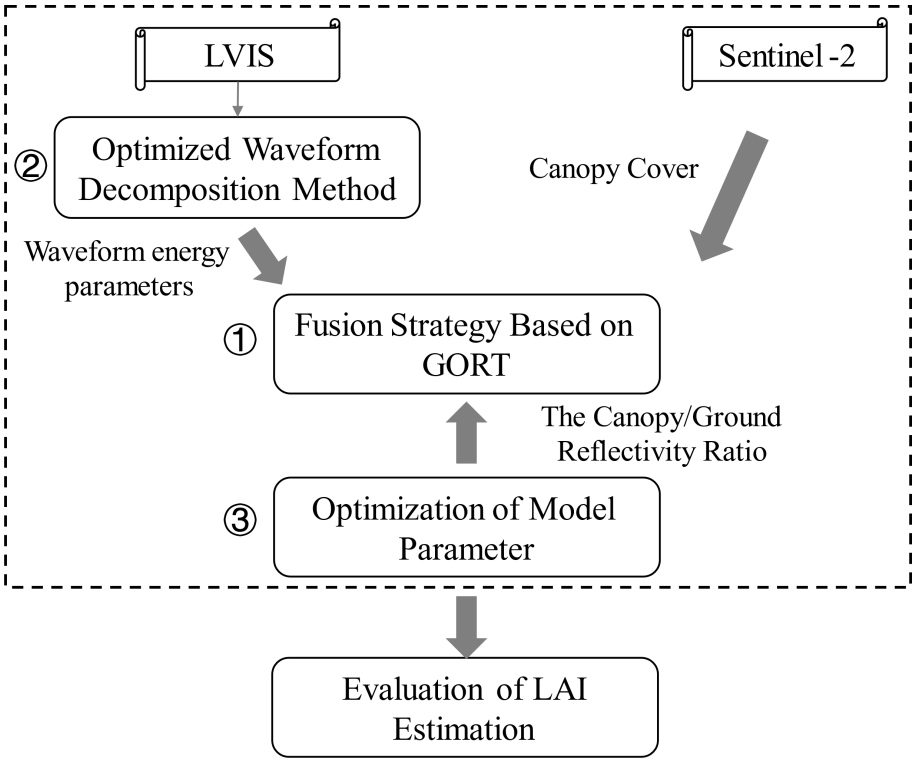


FIGURE 2
The overview of the methods used in this paper.

GORT model to derive gap probability and canopy cover from LiDAR waveforms. In the modified GORT model, the cumulative canopy height profile (CHP) was calculated by using a logarithmic transformation of (1-canopy cover). The canopy cover and gap probability could be calculated as follows (Ni-Meister et al., 2001):

$$fcoVer(h) = 1 - P_{gap}(h) = \frac{R_V(h)}{R_V(0)} \frac{1}{1+\frac{\rho_v}{\rho_g}\frac{R_g}{R_V(0)}} \quad (1)$$

In formula (1), $P_{gap}(h)$ and $fcoVer(h)$ represented the gap probability and canopy cover percentage above a particular height h within canopy respectively. The terms $R_V(h)$, $R_V(0)$ and R_g were the integrated laser energy returns from the canopy top to height h, from canopy top to canopy bottom, and from the ground return individually. The canopy and ground reflectance were ρ_v and ρ_g respectively.

Canopy cover (CC) is typically defined as the extent of ground area covered by the foliage of trees or other vegetation, as projected from a vertical viewpoint onto a horizontal plane (Fiala et al., 2006). From the perspective of remote sensing approaches, two types of canopy cover estimates are commonly derived: metrics describing the 2D horizontal extent of canopy, which is often expressed for a given cover type as a percentage of pixels (Silvan-Cardenas and Wang, 2010); or as 3D LiDAR metrics that represent the transmission of light through the canopy (Morsdorf et al., 2006; Korhonen et al., 2011; Moran et al., 2020). Researches have demonstrated a strong correlation between canopy cover extracted from spectral data and LiDAR-derived inversion values, with the latter frequently exhibiting higher accuracy (Smith et al., 2009; Ma et al., 2017). Figure 3 shows a schematic diagram of the principles of acquiring forest canopy information using active and passive remote sensing sensors.

Based on the above researches, we assumed that the canopy cover obtained by spectral imagery and LiDAR waveform are equal in this study:

$$fcoVer(z) = \frac{R_V(z)}{R_V(0)} \frac{1}{1+\frac{\rho_v}{\rho_g}\frac{R_g}{R_V(0)}} = FVC \quad (2)$$

where $fcoVer(z)$ represented canopy cover percentage above height z. We set the accumulated canopy cover percentage above height z as equal to the estimated canopy cover (FVC) in the pixel at the position of the LiDAR pulse using spectral imagery. z was set to the height corresponding to 80% of the canopy energy in the return. Therefore, the estimate of ρ_v/ρ_g could be expressed as:

$$\frac{\rho_v}{\rho_g} = \frac{R_V(z)}{FVC \star R_g} - \frac{R_V(0)}{R_g} \quad (3)$$

Based on the above formulas, the canopy/ground reflectivity ratio value of each pulse could be calculated according to LiDAR waveform and spectral imagery.

Tang et al. (2012) deduced the formula of LAI derivation based on the GORT model. The effectiveness of the method has been proved by experiments. Total LAI can be calculated as (Tang et al., 2012):

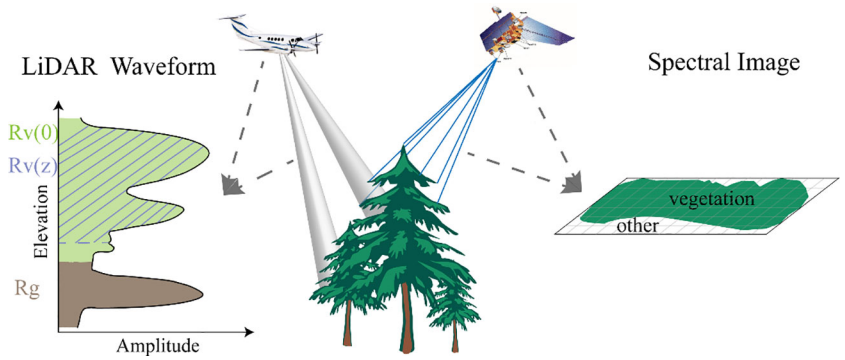


FIGURE 3
The schematic diagram illustrating the principle of acquiring forest canopy information through active and passive remote sensing methods.

$$LAI_{total} = \frac{C}{G} \times \ln\left(1 + \frac{R_V(0)}{\frac{\rho_v}{\rho_g} \times R_g}\right) \quad (4)$$

where C represented the clumping index which adjusted the linear relationship between effective LAI and true LAI (Chen, 1996). We chose the clumping index value of 1.58 for the in Harvard Forest (Tang et al., 2012). G was the projection coefficient and was set to be 0.5 assuming a random foliage distribution within the canopy (Ni-Meister et al., 2001). $R_V(0)$ and R_g were the integrated laser energy returns from canopy top to canopy bottom, and from the ground return individually, which were obtained by waveform processing. ρ_v/ρ_g was calculated as formula (3) using both spectral data and LiDAR waveforms. Based on formula 3 and 4, LAI was estimated by combining spaceborne and airborne data. The strength of this strategy is that we used the physical model to realize the data fusion, and improve the parameter ρ_v/ρ_g by using the spaceborne spectral reflectance.

3.2 Optimization of waveform decomposition method for accurate waveform energy parameters extraction

Waveform energy parameters are the main input data in LAI inversion based on the GORT model, including canopy backward energy, ground backward energy and waveform energy integral returned at different altitudes, which are useful for segmentation, classification and inversion purposes, in both forested and urban areas (Mallet and Bretar, 2009). Selecting an appropriate processing method to "purify" the original waveform data is vital to extract structure parameters of forest, so as to accurately invert physical parameters of vegetation. In this paper, waveform processing procedure was designed for the processing of the Level 1B Geolocated LVIS Waveforms and Level 2 Geolocated Surface Elevation and Height Product. Specifically, to extract energy parameters more accurately from waveforms, the waveform decomposition algorithm was improved. The processing flow chart is shown in Figure 4.

Firstly, the background noise of the echo was removed based on the average noise parameter "sigmean" of each echo calculated in flight provided by LVIS Level 1B product. Values less than the noise threshold was eliminated.

Secondly, the Gaussian filtering algorithm was used to remove other types of noise and smoothed the waveform. The methods of waveform denoising mainly include Gaussian filtering, mean filtering, Fourier low-pass filtering, etc. (Zhang et al., 2020). Gaussian filtering has small time-frequency window area and a simple design, which makes it widely used in the field of signal processing. By measuring the echo denoising effect and adjusting the parameters, a Gaussian filter with better denoising effect was finally selected for echo denoising.

Finally, waveform decomposition method was applied to decompose the echo and extracted effective waveform parameters. Since the backscattered echo signal can be considered as the superposition of multiple Gaussian signals, the Gaussian decomposition method was used to fit the original signal to the superposition of multiple Gaussian function curves (Zhou et al., 2022). The backscattered echo can be expressed as:

$$W(t) = \varepsilon + \sum_{m=1}^{N_p} A_m exp\left[-\frac{(t - t_m)^2}{2\sigma_m^2}\right] \quad (5)$$

In formula (5), $W(t)$ is the amplitude of the waveform at time t; ε is the bias of the Gaussian waveform; N_p is the number of Gaussian components; A_m, t_m, σ_m are the amplitude, peak position and waveform width of the waveform of the mth Gaussian component respectively.

There are two main steps in the waveform decomposition: 1) estimation of the initial parameters; 2) optimization of the parameters and fitting the waveform (Zhou et al., 2021). After extracting accurate initial values of parameters, the commonly used waveform fitting methods include LM (Levenberg-Marquardt optimization algorithm) method (Wagner et al., 2006) and EM (Expectation-Maximization algorithm) method (Persson et al., 2005), the accuracy of which has no significant difference (Zhou et al., 2022). In the global optimization algorithm, waveform components that are too close in distance are prone to being merged during optimization, leading to significant fitting errors. In this paper, in order to address recognition errors caused by some echo components being close to the target, constraints were placed

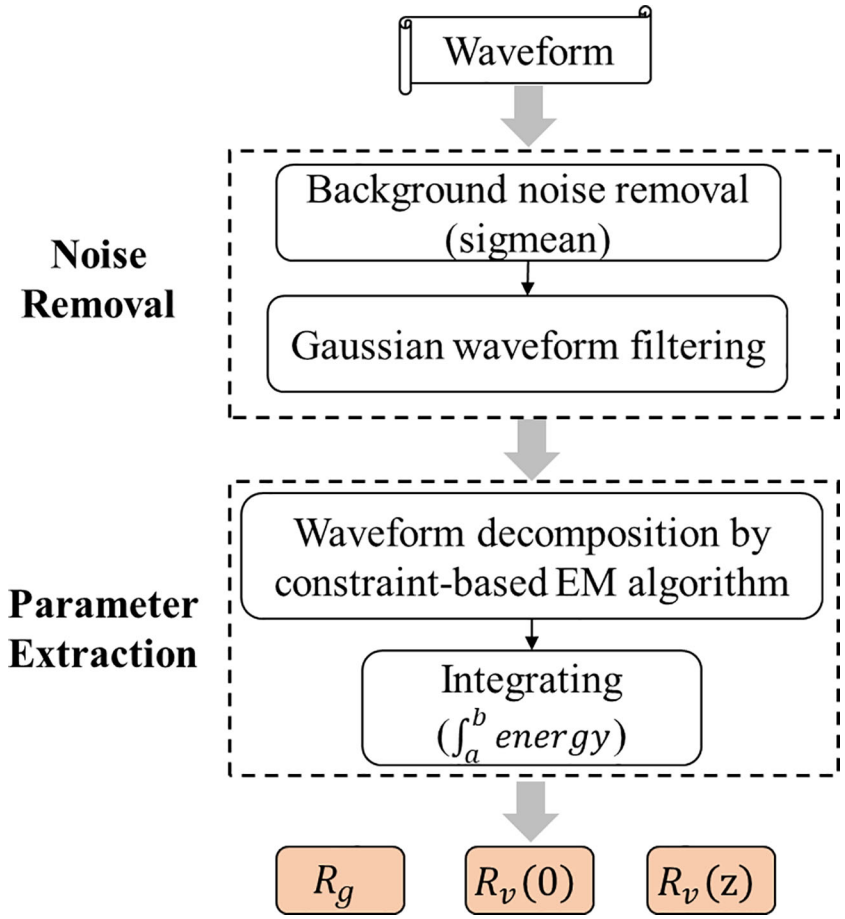


FIGURE 4
The flow chart of waveform processing ($R_v(z)$, $R_v(0)$ and R_g are the integrated laser energy returns from the canopy top to height z, from canopy top to canopy bottom, and from the ground return individually.).

on the values of each peak to improve the accuracy of the fitting based on EM algorithm. The specific approach of the constraint-based EM was designed as follows:

E-step: Compute the posterior probabilities for each component given the data points using Bayes' rule. The posterior probability for the jth component of the Gaussian mixture model for the ith data point was given by:

$$\gamma_{ij} = \frac{w_j p_j(x_i)}{\sum_{k=1}^{K} w_k p_k(x_k)} \tag{6}$$

where γ_{ij} represented the probability that the ith data point belonged to the jth component, $p_j(x_i)$ was the probability density function of a Gaussian distribution at x_i of the j th component.

M-step: Update the parameters of the Gaussian mixture model using the posterior probabilities computed in the E-step. The update equations for the means (μ_j), and standard deviations (σ_j) of the jth component were given by:

$$P_{[a,b]}(x) = \begin{cases} a & if \quad x < a \\ x & if \quad a \le x \le b \\ b & if \quad x > b \end{cases} \tag{7}$$

$$a = \mu_{j_0} - \sigma_{j_0}, \quad b = \mu_{j_0} + \sigma_{j_0} \tag{8}$$

$$\mu_j = P_{[a,b]}\left(\frac{\sum_{i=1}^{N} \gamma_{ij} x_i}{\sum_{i=1}^{N} \gamma_{ij}}\right) \tag{9}$$

$$\sigma_j^2 = \frac{\sum_{i=1}^{N} \gamma_{ij} \left(x_i - \mu_j\right)^2}{\sum_{i=1}^{N} \gamma_{ij}} \tag{10}$$

where $P_{[a,b]}(x)$ was the projection operator that maps x onto the interval $[a, b]$, μ_{j_0}, σ_{j_0} were the initial values of the mean and standard deviation for the jth component.

Repeat E-step, M-step until convergence was achieved. The constrained EM algorithm for LiDAR waveform decomposition imposed constraints on the parameter range within the optimization problem, leading to enhanced stability and accuracy of the algorithm.

After waveform processing, we identified the last waveform component as the ground component, and the rest as the canopy components. R_g was the area enclosed by the amplitude of the last waveform component and the coordinate axis within its start-stop

range. $R_V(0)$ was the integral value of the other waveform components within their start-stop range. $R_V(z)$ was the integral value of the waveform from the initial position of canopy component to the height of z.

In this section, we implemented the processing of LiDAR waveforms, especially by adding constraints on the peak positions in the waveform decomposition algorithm to obtain better waveform decomposition results and calculate more accurate waveform energy parameters.

3.3 Optimization of model parameter ρ_v/ρ_g for large scale forest by gridding study area

When using the strategy proposed in Section 3.1 to calculate the LAI of each pulse position, both spectral and LiDAR data are used to calculate the model parameter ρ_v/ρ_g at this point, that is, ρ_v/ρ_g varies with the input data in each position. Although the parameter setting strategy is more accurate than taking a fixed value for the entire study area, it causes large data uncertainty and increases the computation of inversion. For example, abnormal waveform of a pulse will result in abnormal calculated value of ρ_v/ρ_g, thus leading to abnormal LAI inversion results at this point.

To solve this problem, we proposed a method to optimize the model parameter ρ_v/ρ_g, that was, gridded the study area and calculated the mean value of ρ_v/ρ_g in each grid. Then the LAI estimation model was constructed by each grid. In a large area, the canopy/ground reflectivity ratio varies with forest environmental conditions. Gridding the study area not only accounts for the spatial heterogeneity within Harvard Forest but also reduces the computational complexity of LAI modeling, thereby mitigating uncertainty arising from anomalous input data. Theoretically, the optimization method is helpful to improve the inversion accuracy.

Specifically, we divided the study area into 2331 rectangular grids (63*37), the coordinate size of each grid was 0.002° * 0.002° (about 36118m^2). In order to reduce the uncertainty, the statistical method of histogram distribution was used to eliminate extreme abnormal ratio values that without the range of the mean plus or minus three standard deviations in each grid. Then an average reflectivity ratio was assigned for all the pulses in grid to reduce the uncertainty caused by pulse quality. The reflectivity ratio of each grid was the trimmed mean of all pulses in the grid.

4 Result and discussion

4.1 Waveform energy parameters based on optimized waveform decomposition method

For the purpose of obtaining accurate canopy and ground energy parameters, we processed a total of 187,302 LiDAR waveforms in the study area through denoising, smoothing, and waveform decomposition based on least squares optimization. During this process, the waveform energy parameters *Rg* and *RV* obtained using different least squares optimization methods were compared.

The correspondence between the original waveforms and Gaussian model estimates of *Rg* and *RV* using different optimization algorithms is shown in Table 1. Overall the correspondence is high for both *RV* and *Rg*, the R^2 of which are all above 0.99. Further comparative analysis between the canopy and ground components reveal that the overall canopy R^2 is lower than that of the ground. The RMSE for the canopy is significantly higher than that of the ground, with a maximum difference of 35.656. It can be attributed to two main factors. Firstly, the canopy component often has one or more echo components in LiDAR waveforms, whereas the ground component typically only has one. Multiple echo components result in larger errors in waveform decomposition, leading to poorer correspondence of *RV*. Secondly, due to the significant differences in canopy coverage, the waveform energy fluctuation of canopy is also much larger, resulting in a much higher RMSE compared to the ground.

Comparing the results obtained from two different optimization methods, it is found that correspondence is higher for both *Rg* and *RV* when the bound of peaks are constrained, compared to the result without parameter constraints (higher R^2 and lower RMSE). To investigate the reasons for the result, we compare the waveform fitting results of two different optimization methods. The results of waveform decomposition using unconstrained optimization algorithm and optimization algorithm with peak boundary constraints are shown in the Table 2. We conducted an average statistical analysis of the fitting results for all waveforms in the study area and found that the constrained optimization algorithm yielded a fitting waveform with $\overline{R^2}$ of 0.979, $\overline{MAE}$ of 6.907, $\overline{MSE}$ of 103.016, and $\overline{RMSE}$ of 8.596. The fact that the average R^2 is almost close to 1

TABLE 1 Correlation results of waveform and Gaussian model integrals using different optimization algorithms for the ground (Rg) and canopy (Rv) components of the waveforms.

Waveform energy parameters	Optimization algorithms	R^2	RMSE
Rv	Constraint-based EM	0.991	62.980
	EM	0.990	65.926
Rg	Constraint-based EM	0.998	27.324
	EM	0.997	33.766

"Constraint-based EM" means the Expectation-Maximization algorithm with peaks boundary constraints, "EM" means the original Expectation-Maximization algorithm.

TABLE 2 Performance of waveform decomposition using different optimization algorithms.

Optimization algorithms	$\overline{R^2}$	$\overline{MAE}$	$\overline{MSE}$	$\overline{RMSE}$
Constraint-based EM	0.979	6.907	103.016	8.596
EM	0.978	6.971	104.740	8.681

"Constraint-based EM" means the Expectation-Maximization algorithm with peaks boundary constraints, "EM" means the original Expectation-Maximization algorithm.

and the small values of the average MAE and RMSE indicate good overall waveform fitting results, suggesting that the model is able to fit the majority of the LiDAR waveforms accurately. The optimization algorithm with no constrain yielded a fitting waveform with $\overline{R^2}$ of 0.978, $\overline{MAE}$ of 6.971, $\overline{MSE}$ of 104.740, and $\overline{RMSE}$ of 8.681. Compared with the constrained method, the average MAE, MSE, and RMSE values are higher for the unconstrained method, while the difference of average

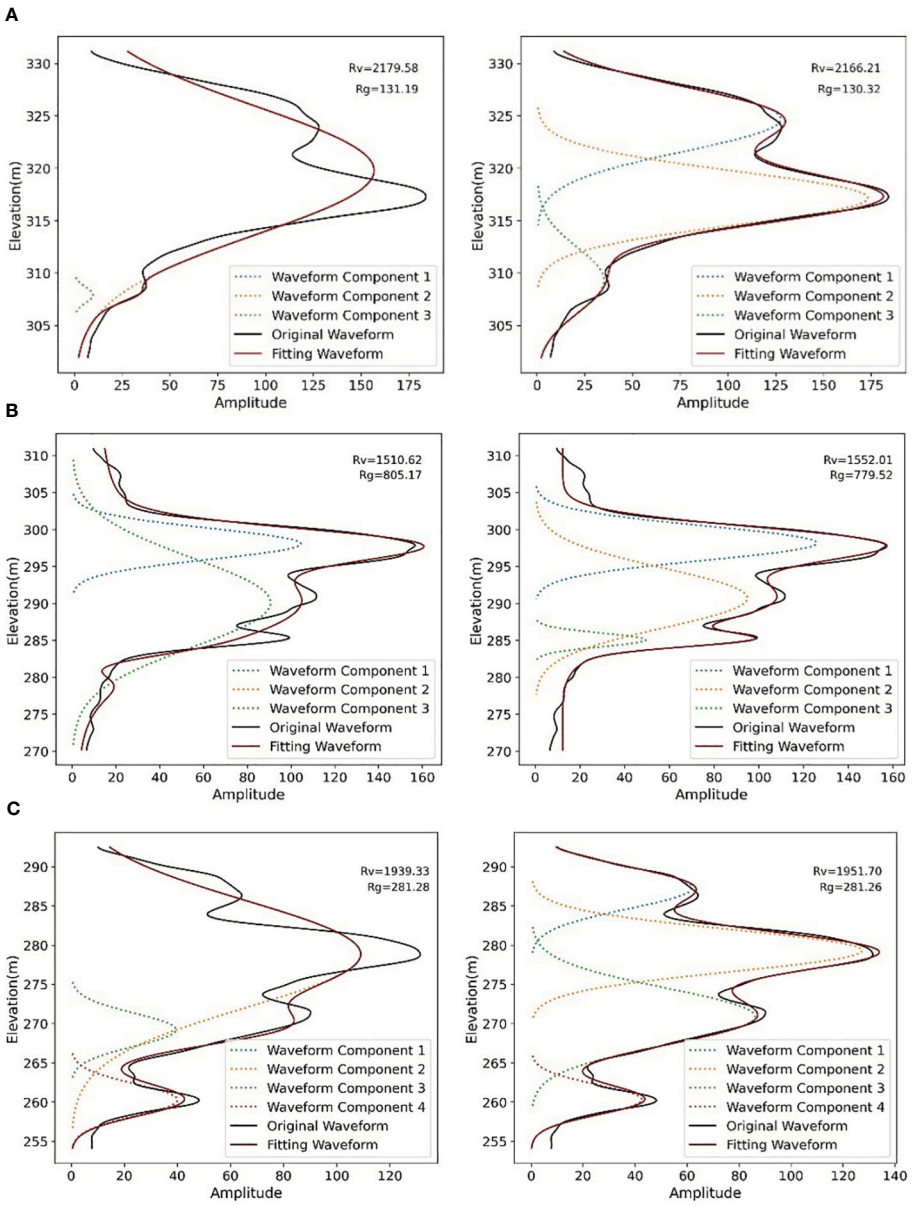


FIGURE 5
Waveform fitting and waveform energy parameters results using two different optimization methods. **(A)** with two canopy echoes and one ground echo; **(B)** with two canopy echoes and one ground echo; **(C)** with three canopy echoes and one ground echo (The left column: the original Expectation-Maximization algorithm, the right column: the Expectation-Maximization algorithm with peaks boundary constraints).

TABLE 3 Rg and Rv of typical waveforms based on different methods.

Typical waveforms	Parameter	Constraint-based EM	EM	Actual value
(a)	*Rv*	2166.21	2179.58	2166.87
	Rg	130.32	131.19	128.82
(b)	*Rv*	1552.01	1510.62	1562.44
	Rg	779.52	805.17	775.86
(c)	*Rv*	1951.70	1939.33	1945.25
	Rg	292.26	291.28	294.82

$R^2$is not significant. This suggests that the overall fitting accuracy of the unconstrained method is lower than that of the constrained method.

To further analyze the waveform decomposition result using different methods and its effects on the value of *Rg* and *RV*, we extracted some waveforms with significant differences in inversion accuracy using two different optimization methods (Figure 5). The left column shows results obtained by EM method, while the right column shows results obtained by constraint-based EM method. The waveform energy parameters result of the three waveforms are shown in Table 3. Waveform (a) contains two canopy echoes and one ground echo. The waveform decomposition using EM method identifies only one canopy echo, and the calculated energy of the canopy echo is 2179.58, which differs significantly from the actual value of 2166.87. The constraint-based EM method identifies two canopy echoes, and the calculated value of 2166.21 is almost the same as the actual value. Waveform (b) contains two canopy echoes and one ground echo, and the second canopy echo and ground echo are combined into one waveform component using the EM method, resulting in a large difference between the canopy and ground energy calculation results of 1510.62 and 805.17 and the actual 1562.44 and 775.86, while the added constraint method does not show this phenomenon. Waveform (c) contains three canopy echoes and one ground echo, which are partially combined by the EM method due to the close distance of each canopy echo, resulting in a large difference between the calculated canopy energy (1939.33) and the actual value (1945.25). The results reveal that the waveform energy parameters calculated by constraint-based EM method are more precise than the unconstrained EM method. The waveform diagram shows that the main reason for the difference in accuracy between the two waveform decomposition methods is that after constraining the peak position of each waveform component, the merge of the closed waveform echoes can be avoided.

These results provide evidence that compared to optimization algorithms without parameter boundary constraints, the proposed

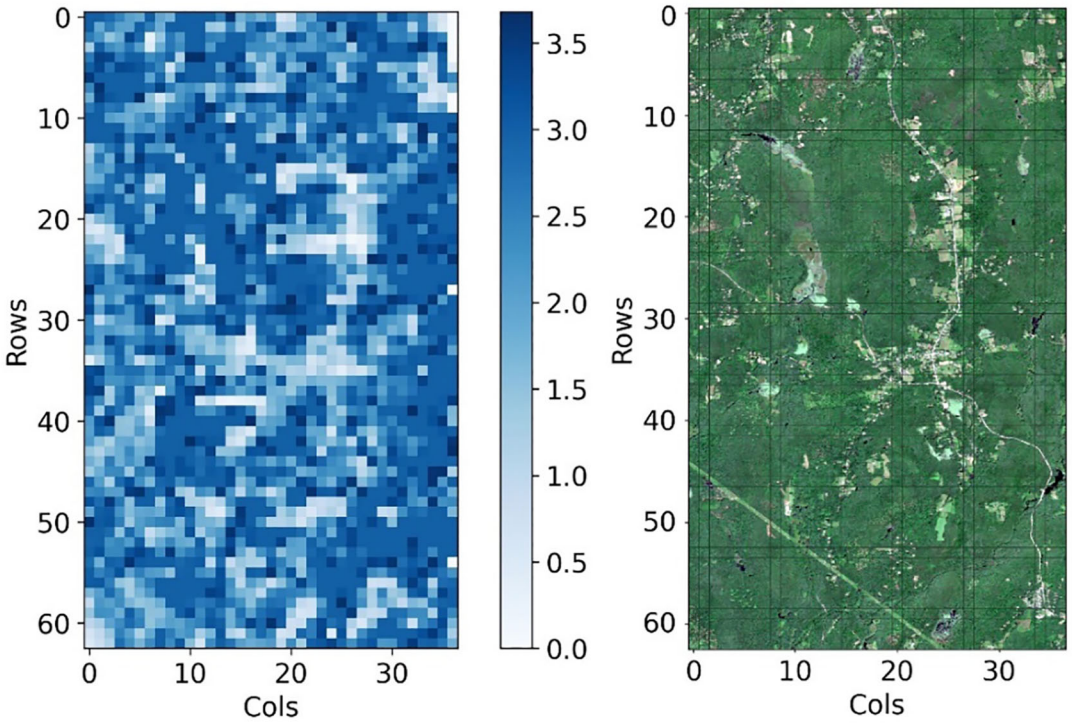


FIGURE 6
The heat map of gridded canopy/ground reflectivity ratios in the study area (left) (The right image shows the Sentinel-2 RGB image of the study area).

TABLE 4 The statistics of the canopy/ground reflectivity ratio of each grid.

Parameters	Value
Number of grids	2331
Max	3.68
Min	0.00
Average	2.45
Standard deviation	0.80

constraint-based EM method can perform waveform decomposition more accurately and, to a certain extent, avoid decomposition errors caused by waveform components being too close to each other. Based on the waveform decomposition method, more precise *RV* and *Rg* can be obtained, providing accurate values for subsequent LAI inversion.

4.2 The gridded ρ_v/ρ_g result

After gridding the study area, the ρ_v/ρ_g value of each grid is shown in Figure 6, and the statistical result is shown in the Table 4. Comparing the value heatmap (Figure 6 left) with the RGB images (Figure 6 right) of the study area, it can be found that the ρ_v/ρ_g is correlated with the degree of vegetation coverage. The value of ρ_v/ρ_g in non-vegetated and sparsely vegetated areas is generally lower than that of areas with higher vegetation coverage. The spatial distribution of ρ_v/ρ_g coincided with the actual vegetation distribution, indicating the accuracy of the calculation method we proposed. The statistical results show that the final ratio of 2331 grids in the study area is within the range of [0, 3.68], the average value is 2.45, and the root mean square is 0.80.

In previous studies, Lefsky et al. (1999) suggested using a constant (ρ_v/ρ_g = 2) for 1064 nm. Tang et al. (2012) obtained the ρ_v/ρ_g value of 2.5 at 1064 nm and used it as the mean value for the whole study area. In addition to determining the ratio by empirical field measurements, extracting from LiDAR waveforms using statistical methods (Ni-Meister et al., 2010; Armston et al., 2013) is also a way to calculate this ratio. The value of canopy/ground reflectivity ratio is basically between [0, 3]. Due to the lack of measured data, and no research has used the method of combining spectral and LiDAR data to calculate ρ_v/ρ_g, it is currently impossible to accurately demonstrate the accuracy of the calculated ratios. However, the average and mean square deviation results show that most of the reflectance ratios are [0, 3] with only a few abnormal values, it can be proved that accurate gridded ρ_v/ρ_g of the study area can be obtained by this strategy. These provide a new idea for the calculation of ρ_v/ρ_g.

4.3 Comparison of LAI inversion results based on different data and inversion methods

In this section, we explored the LAI retrieval results based on different datasets and different methods for calculating ρ_v/ρ_g. The accuracy is compared with the existing field measurement data.

Using the fusion strategy we proposed, we obtained the LAI map using LiDAR waveform and multispectral image (Figure 7). The result shows that directly using the data of entire study area without land cover classification can estimate LAI well. For example, LAI values of the road in the lower left of this research area are 0. We can clearly distinguish the vegetation and non-vegetation areas from the LAI map. In the vegetated area, the LAI values are generally around 3-7, and reach above 7 in a few dense areas, indicating that the area is relatively heavily forested, which is basically consistent with the actual situation. Comparing the inversion results with the true LAI value provided by Harvard Forest HEM plots, the correlation, R^2 and RMSE are 0.81, 0.65, 1.01 respectively (Figure 8), which shows that the LAI map obtained have high accuracy. The method we proposed to invert LAI by fusing LiDAR and spectral data is feasible.

We conducted LAI estimation based on four different strategies for comparative analysis. These strategies encompassed two that exclusively utilized LiDAR data, one that solely relied on spectral data, and one that integrated both LiDAR and spectral data. Using only LiDAR data, we reproduced the methods used by Tang et al. (2012) and Armston et al. (2013) respectively. They both performed the inversion based on the GORT model, only some of the parameters in the model were determined in different ways. The LAI estimation maps (Figures 9A, B) were performed according to the parameters they set. Using only spectral data, we adopted the most traditional empirical model to construct the linear

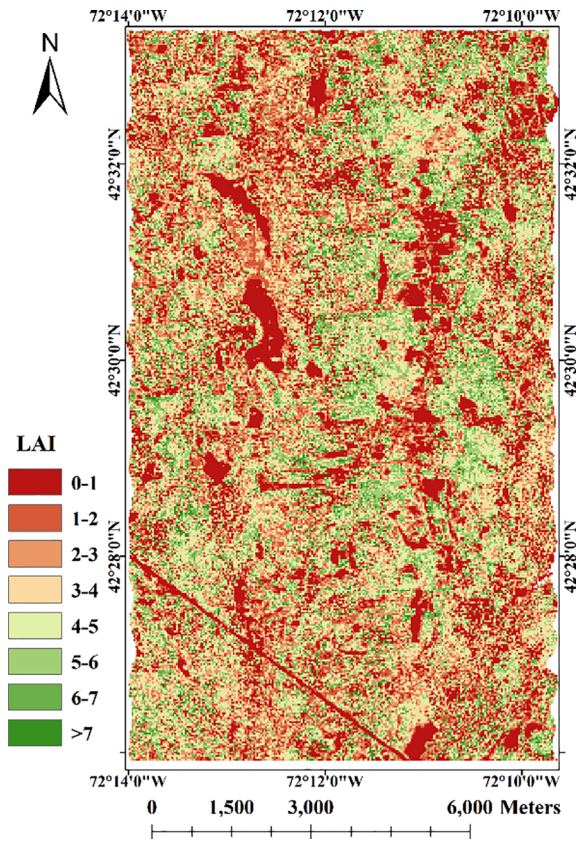


FIGURE 7
LAI map of the study area using both spectral and LiDAR data.

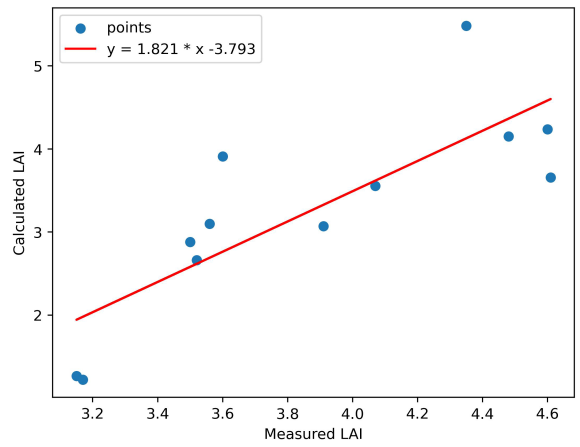


FIGURE 8
Scatterplots of field-observed LAI against estimated LAI using both spectral and LiDAR data.

relationship between NDVI-LAI for LAI inversion using the LAI true value of the Harvard Forest LPH flux tower's 36 plots (Figure 9C). Using LiDAR waveform and spectral data, we employed the method proposed by Yang et al. (2019) and the obtained result was depicted in Figure 9D. Based on the true LAI value of Harvard Forest, the accuracy evaluation of the estimation results obtained by various methods was carried out. The overall accuracy results are shown in Table 5.

Figures 9A, B indicates the LAI inversion results using only LiDAR data. Similar to LAI map obtained by the method proposed in this study, the results show obvious differentiation between vegetation and non-vegetation areas. However, the difference lies in that, LAI values of invention are too high in places with dense vegetation only using LiDAR data. By Tang et al. (2012) 's method, the LAI inversion result shows the correlation of 0.63, R^2 of 0.40 and RMSE of 2.01 with ground plot LAI, indicating a moderate correlation between the two. By Armston et al. (2013) 's method, the inverted LAI is weakly correlated with the true value, with a correlation coefficient of 0.53, R^2 of 0.28 and RMSE of 2.85. It shows that the accuracy of the two LiDAR-only inversion methods is lower than that of the fusion of LiDAR and spectral data. The LAI map inverted by the empirical model is shown in Figure 9C. The results show that there is a large difference between the LAI inversion results obtained by using only spectral data and those obtained by fusing the two data. Based only on spectral data, the inverted LAI is an underestimate with the highest value being only 4.8. It shows a correlation of 0.48, R^2 of 0.25 and RMSE of 2.72 with ground plot LAI, which is poor compared to the inversion result that combines the two data. Figure 9D illustrates a low accuracy in LAI estimation (the correlation of 0.51, R^2 of 0.30 and RMSE of 2.76), with LAI values consistently underestimated. This discrepancy may be attributed to the unsuitability of the model for the tree species and data sources in this study area. The original LiDAR waveform used by Yang et al. (2019) is acquired from a large-footprint LiDAR system (70m), which significantly differs from the experimental data used in this study (25m). Comparing various estimation strategies, it is evident that the fusion of both active and passive remote sensing data contributes to improved LAI estimation accuracy. The enhancements we have proposed for the GORT model further enhance LAI estimation precision.

We successfully estimate LAI based on the GORT model combining LiDAR and spectral data with a correlation of 0.81 and R^2 of 0.65, which shows a large accuracy improvement compared to both LiDAR data alone and spectral data alone. These improvements can be attributed to the addition of spectral data to improve the parameters of the model. Comparing the three inversion methods based on the GORT model, the main difference between them is the ρ_v/ρ_g value of the model, which may be the main reason for the difference in inversion accuracy. The ρ_v/ρ_g of the study area is determined as a constant value of 2.5 by experience, which helps to reduce the amount of computation. But this empirical value may not necessarily be applicable to Harvard Forest and a fixed value cannot adequately represent the forest conditions of the entire study area, which may be the prime causes of the low accuracy. The use of least squares methods provides a new approach for the calculation of the ratio, which does not rely on manual measurements, but rather on the energy returned by the LiDAR. Based on this approach, we obtain ρ_v/ρ_g of 1.17 for the study area. The main reasons for the low accuracy can be attributed to two factors. Firstly, similar to using experienced value, calculating a single value for the entire Harvard Forest will cause abnormal results due to the complexity of the forest canopy. Secondly, the quality of the LiDAR can greatly affect the results. Such as areas in low point density in canopy that do not reflect enough energy (Chauve et al., 2009). Combining the spaceborne spectral data and airborne LiDAR data to calculate the reflectance ratio can use high-quality spectral data to a certain extent to eliminate the abnormal phenomenon of reflectance ratio caused by the abnormal collection of some LiDAR footprints. Compared with the above two methods, the gridded ρ_v/ρ_g calculation method we proposed considers the influence of these factors. By combining spaceborne spectral data and airborne LiDAR waveform to calculate the ρ_v/ρ_g, the influence of abnormal waveforms on the value of ρ_v/ρ_g can be eliminated to some extent. At the same time, laborious ground truth measurements of reflectance are no longer needed (Yang et al., 2006). Furthermore, dividing the study area into grids and calculating the average ρ_v/ρ_g in each grid can not only further eliminate abnormal values through statistical method, but also calculate different ρ_v/ρ_g in view of the canopy heterogeneity in large-scale complex forest. It can be found from the LAI maps of the three methods that the method of taking a constant value would lead to higher LAI values in the areas of dense vegetation, which is due to the fact that the actual ρ_v/ρ_g in these areas are higher than the determined value of the model. LAI results obtained by the method we proposed are basically within 7, with only a few outliers. The results also confirm the validity of the proposed method.

In addition to the inversion based on the GORT physical model, the LAI inversion based on the empirical NDVI-LAI relationship (Turner et al., 1999) is also carried out based on the spectral data. The

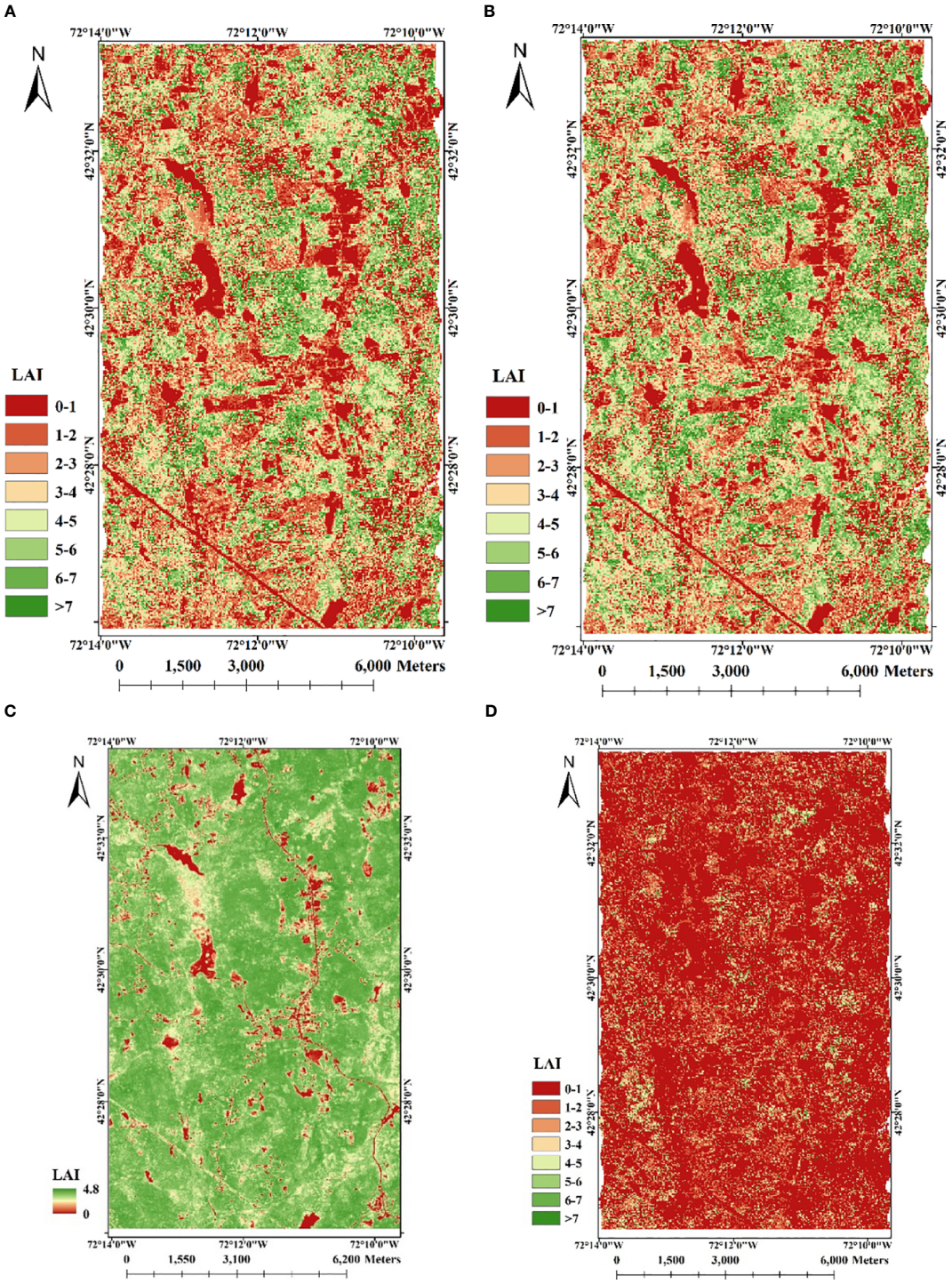


FIGURE 9
LAI maps of the study area using different strategies. **(A)** by Tang et al. (2012) 's method; **(B)** by Armston et al. (2013) 's method; **(C)** by NDVI-LAI relationship model; D: by Yang et al. (2019) 's method.

results show that the inversion accuracy of the empirical model using spectral data (correlation of 0.48 and R^2 of 0.25) is much lower than that of the physical model using LiDAR waveforms. The empirical model needs a certain amount of truth values to ensure the accuracy of the inversion equation. However, there are only 48 small plots in the study area, which in theory can cause large errors when used for empirical model construction and verification. Also, the existing plots of Harvard Forest HEM and LPH Towers are concentrated in a small area, which cannot well represent the NDVI-LAI relationship of the entire study area. It is the main reason for the lower precision., Our

TABLE 5 The accuracy of LAI retrieval results using different methods.

Data	Reference	Model	ρ_v/ρ_g calculation strategy	ρ_v/ρ_g	Correlation	R^2	RMSE
LiDAR waveform and Spectral imagery	\	GORT	Fusion strategy (proposed)	Change value	0.81	0.65	1.01
LiDAR waveform	Tang et al. (2012)	GORT	Set experienced value	2.50	0.63	0.40	2.01
LiDAR waveform	Armston et al. (2013)	GORT	Least squares	1.07	0.53	0.28	2.85
LiDAR waveform and Spectral imagery	Yang et al. (2019)	Gap fraction-based model	Set experienced value	2.00	0.51	0.30	2.76
Spectral imagery	\	NDVI-LAI relationship	\	\	0.48	0.25	2.72

\, indicates that the grid value is empty.

results show that the strategy we proposed in this study is viable for predicting forest LAI. The combined multispectral imagery and LiDAR waveform can improve the input parameter ρ_v/ρ_g of the GORT model and contribute to prediction accuracy of LAI.

4.4 Limitations

While this study has successfully achieved the joint inversion of spectral and LiDAR data for LAI estimation based on a physical model, there are still some limitations. Firstly, due to the limited number of ground measurement points in our study area, we are unable to achieve the fusion of LiDAR and spectral data for LAI retrieval based on empirical models, and compare it with the results from physical models. Nevertheless, this also partly demonstrates the advantages of developing data fusion inversion based on physical models, which helps us reduce reliance on ground measurements, lower manual labor costs, and facilitates the widespread application of large-scale regions. Secondly, while both LiDAR data and spectral data are commonly employed for retrieving canopy cover, there remains a disparity between the values obtained through these two methods (Smith et al., 2009; Li et al., 2023). This discrepancy, though overlooked in this experiment, may cause errors in inversion. It might be one of the contributing factors to the limited precision in LAI inversion. Research is warranted in future experiments to address this issue and enhance the inversion accuracy. Additionally, due to time and resource constraints, we do not validate the effectiveness of this method in regional scale. In future research, we will further explore the contribution of data fusion to LAI based on theoretical analysis.

5 Conclusion

Spectral imagery and full-waveform LiDAR data can provide reflectance information and echo energy information reflecting the vertical structure of the forest canopy respectively. Joint active and passive remote sensing data has great potential for accurate inversion of forest canopy LAI. Our research is one of the few attempts to derive LAI using both spectral imagery and LiDAR waveform based on physical model retrieval rather than through empirical methods. We proposed a useful data-joint LAI inversion strategy based on the GORT model using LiDAR waveform and spectral data. For the large-scale heterogeneous forest, we further accurately extracted the waveform energy parameters as the model input data and optimized the model input parameter canopy/ground reflectivity ratio to improve the inversion accuracy. The results show that comparing with only using LiDAR or spectral imagery, the LAI calculated by the proposed strategy using both LiDAR waveform and spectral imagery has a higher accuracy, indicating the effectiveness of the proposed strategy. Overall, our study confirms that optimizing the input parameter and data of the model for the study area can help improve the inversion accuracy, and the combined LiDAR waveform and multispectral imagery have potential for improving prediction accuracies of LAI.

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

Conceptualization, ZS and SS. Methodology, ZS and SS. Formal analysis, ZS and LX. Investigation, ZS and BW. Data curation, BC and QX. Writing—original draft preparation, ZS. Writing—review and editing, JS and SS. Supervision, WG. Project administration, WG. Funding acquisition, WG and SS. All authors contributed to the article and approved the submitted version.

Funding

This work is supported by the National Natural Science Foundation of China (Grant No.41971307), Fundamental Research

Funds for the Central Universities (Grant No.2042022kf1200, 2042023kf0217), State Key Laboratory of Geo-Information Engineering (Grant No.SKLGIE2023-Z-3-1), Wuhan University Specific Fund for Major School-level Internationalization Initiatives, and LIESMARS Special Research Funding.

Acknowledgments

We gratefully acknowledge the Harvard Forest, National Snow & Ice Data Center (NSIDC) and Copernicus SENTINEL-2 mission for providing public data and the open access to the python, SNAP. The authors would also like to thank reviewers for many constructive comments on 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.

References

Armston, J., Disney, M., Lewis, P., Scarth, P., Phinn, S., Lucas, R., et al. (2013). Direct retrieval of canopy gap probability using airborne waveform lidar. *Remote Sens. Environ.* 134, 24–38. doi: 10.1016/j.rse.2013.02.021

Barclay, H. J., and Goodman, D. (2000). Conversion of total to projected leaf area index in conifers. *Can. J. Botany-Revue Can. Botanique* 78, 447–454. doi: 10.1139/cjb-78-4-447

Blair, J. B., and Hofton., M. (2020). *LVIS Classic L1B Geolocated Return Energy Waveforms, Version 1* (Boulder, Colorado USA: NASA National Snow and Ice Data Center Distributed Active Archive Center). doi: 10.5067/O8UCOA2D6ZE3

Blair, J. B., Rabine, D. L., and Hofton, M. A. (1999). The Laser Vegetation Imaging Sensor: a medium-altitude, digitisation-only, airborne laser altimeter for mapping vegetation and topography. *Isprs J. Photogrammetry Remote Sens.* 54, 115–122. doi: 10.1016/S0924-2716(99)00002-7

Chauve, A., Vega, C., Durrieu, S., Bretar, F., Allouis, T., Deseilligny, M. P., et al. (2009). Advanced full-waveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest. *Int. J. Remote Sens.* 30, 5211–5228. doi: 10.1080/01431160903023009

Chen, J. M. (1996). Optically-based methods for measuring seasonal variation of leaf area index in boreal conifer stands. *Agric. For. Meteorol.* 80, 135–163. doi: 10.1016/0168-1923(95)02291-0

Chen, J. M., and Black, T. A. (1992). Defining leaf-area index for non-flat leaves. *Plant Cell Environ.* 15, 421–429. doi: 10.1111/j.1365-3040.1992.tb00992.x

Clevers, J., and vanLeeuwen, H. J. C. (1996). Combined use of optical and microwave remote sensing data for crop growth monitoring. *Remote Sens. Environ.* 56, 42–51. doi: 10.1016/0034-4257(95)00227-8

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., et al. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* 120, 25–36. doi: 10.1016/j.rse.2011.11.026

Fang, H. L., Baret, F., Plummer, S., and Schaepman-Strub, G. (2019). An overview of global leaf area index (LAI): methods, products, validation, and applications. *Rev. Geophys.* 57, 739–799. doi: 10.1029/2018RG000608

Fiala, A. C. S., Garman, S. L., and Gray, A. N. (2006). Comparison of five canopy cover estimation techniques in the western Oregon Cascades. *For. Ecol. Manage.* 232, 188–197. doi: 10.1016/j.foreco.2006.05.069

Fieber, K. D., Davenport, I. J., Tanase, M. A., Ferryman, J. M., Gurney, R. J., Walker, J. P., et al. (2014). Effective LAI and CHP of a single tree from small-footprint full-waveform liDAR. *IEEE Geosci. Remote Sens. Lett.* 11, 1634–1638. doi: 10.1109/LGRS.2014.2303500

Gastellu-Etchegorry, J. P., Lauret, N., Yin, T. G., Landier, L., Kallel, A., Malenovsky, Z., et al. (2017). DART: recent advances in remote sensing data modeling with atmosphere, polarization, and chlorophyll fluorescence. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 10, 2640–2649. doi: 10.1109/JSTARS.2017.2685528

Gastellu-Etchegorry, J. P., Yin, T. G., Lauret, N., Grau, E., Rubio, J., Cook, B. D., et al. (2016). Simulation of satellite, airborne and terrestrial LiDAR with DART (I): Waveform simulation with quasi-Monte Carlo ray tracing. *Remote Sens. Environ.* 184, 418–435. doi: 10.1016/j.rse.2016.07.010

Houborg, R., Soegaard, H., and Boegh, E. (2007). Combining vegetation index and model inversion methods for the extraction of key vegetation biophysical parameters using Terra and Aqua MODIS reflectance data. *Remote Sens. Environ.* 106, 39–58. doi: 10.1016/j.rse.2006.07.016

Hu, Q., Yang, J., Xu, B., Huang, J., Memon, M. S., Yin, G., et al. (2020). Evaluation of global decametric-resolution LAI, FAPAR and FVC estimates derived from sentinel-2 imagery. *Remote Sens.* 12, 912. doi: 10.3390/rs12060912

Jiang, H. L., Cheng, S. Y., Yan, G. J., Kuusk, A., Hu, R. H., Tong, Y. Y., et al. (2022). Clumping effects in leaf area index retrieval from large-footprint full-waveform liDAR. *IEEE Trans. Geosci. Remote Sens.* 60, 1–20. doi: 10.1109/TGRS.2021.3118925

Kennedy, B. E., King, D. J., and Duffe, J. (2020). Comparison of empirical and physical modelling for estimation of biochemical and biophysical vegetation properties: field scale analysis across an arctic bioclimatic gradient. *Remote Sens.* 12 (18), 3073. doi: 10.3390/rs12183073

Korhonen, L., Hadi, Packalen, P., and Rautiainen, M. (2017). Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote Sens. Environ.* 195, 259–274. doi: 10.1016/j.rse.2017.03.021

Korhonen, L., Korpela, I., Heiskanen, J., and Maltamo, M. (2011). Airborne discrete-return LIDAR data in the estimation of vertical canopy cover, angular canopy closure and leaf area index. *Remote Sens. Environ.* 115, 1065–1080. doi: 10.1016/j.rse.2010.12.011

Lefsky, M. A., Harding, D., Cohen, W. B., Parker, G., and Shugart, H. H. (1999). Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sens. Environ.* 67, 83–98. doi: 10.1016/S0034-4257(98)00071-6

Li, L. Y., Mu, X. H., Jiang, H. L., Chianucci, F., Hu, R. H., Song, W. J., et al. (2023). Review of ground and aerial methods for vegetation cover fraction (fCover) and related quantities estimation: definitions, advances, challenges, and future perspectives. *Isprs J. Photogrammetry Remote Sens.* 199, 133–156. doi: 10.1016/j.isprsjprs.2023.03.020

Li, X. W., Strahler, A. H., and Woodcock, C. E. (1995). A hybrid geometric optical-radiative transfer approach for modeling albedo and directional reflectance of discontinuous canopies. *IEEE Trans. Geosci. Remote Sens.* 33, 466–480. doi: 10.1109/TGRS.1995.8746028

Liang, L., Di, L. P., Zhang, L. P., Deng, M. X., Qin, Z. H., Zhao, S. H., et al. (2015). Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method. *Remote Sens. Environ.* 165, 123–134. doi: 10.1016/j.rse.2015.04.032

Liu, K., Zhou, Q. B., Wu, W. B., Xia, T., and Tang, H. J. (2016). Estimating the crop leaf area index using hyperspectral remote sensing. *J. Integr. Agric.* 15, 475–491. doi: 10.1016/S2095-3119(15)61073-5

Luo, S. Z., Wang, C., Li, G. C., and Xi, X. H. (2013). Retrieving leaf area index using ICESat/GLAS full-waveform data. *Remote Sens. Lett.* 4, 745–753. doi: 10.1080/2150704X.2013.790573

Luo, S. Z., Wang, C., Xi, X. H., Nie, S., Fan, X. Y., Chen, H. Y., et al. (2019). Combining hyperspectral imagery and LiDAR pseudo-waveform for predicting crop LAI, canopy height and above-ground biomass. *Ecol. Indic.* 102, 801–812. doi: 10.1016/j.ecolind.2019.03.011

Ma, H., Song, J. L., and Wang, J. D. (2015). Forest canopy LAI and vertical FAVD profile inversion from airborne full-waveform liDAR data based on a radiative transfer model. *Remote Sens.* 7, 1897–1914. doi: 10.3390/rs70201897

Ma, H., Song, J. L., Wang, J. D., Xiao, Z. Q., and Fu, Z. (2014). Improvement of spatially continuous forest LAI retrieval by integration of discrete airborne LiDAR and remote sensing multi-angle optical data. *Agric. For. Meteorol.* 189, 60–70. doi: 10.1016/j.agrformet.2014.01.009

Ma, Q., Su, Y. J., and Guo, Q. H. (2017). Comparison of canopy cover estimations from airborne liDAR, aerial imagery, and satellite imagery. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 10, 4225–4236. doi: 10.1109/JSTARS.2017.2711482

Mallet, C., and Bretar, F. (2009). Full-waveform topographic lidar: State-of-the-art. *Isprs J. Photogrammetry Remote Sens.* 64, 1–16. doi: 10.1016/j.isprsjprs.2008.09.007

Mananze, S., Pocas, I., and Cunha, M. (2018). Retrieval of maize leaf area index using hyperspectral and multispectral data. *Remote Sens.* 10 (12), 1942. doi: 10.3390/rs10121942

Moran, C. J., Kane, V. R., and Seielstad, C. A. (2020). Mapping forest canopy fuels in the western United States with liDAR–landsat covariance. *Remote Sens.* 12 (6), 1000. doi: 10.3390/rs12061000

Morsdorf, F., Kotz, B., Meier, E., Itten, K. I., and Allgower, B. (2006). Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sens. Environ.* 104, 50–61. doi: 10.1016/j.rse.2006.04.019

Myneni, R. B., Nemani, R. R., and Running, S. W. (1997). Estimation of global leaf area index and absorbed par using radiative transfer models. *IEEE Trans. Geosci. Remote Sens.* 35, 1380–1393. doi: 10.1109/36.649788

Neinavaz, E., Skidmore, A. K., Darvishzadeh, R., and Groen, T. A. (2016). Retrieval of leaf area index in different plant species using thermal hyperspectral data. *Isprs J. Photogrammetry Remote Sens.* 119, 390–401. doi: 10.1016/j.isprsjprs.2016.07.001

Ni, W. G., Li, X. W., Woodcock, C. E., Caetano, M. R., and Strahler, A. H. (1999). An analytical hybrid GORT model for bidirectional reflectance over discontinuous plant canopies. *IEEE Trans. Geosci. Remote Sens.* 37, 987–999. doi: 10.1109/36.752217

Ni-Meister, W., Jupp, D. L. B., and Dubayah, R. (2001). Modeling lidar waveforms in heterogeneous and discrete canopies. *IEEE Trans. Geosci. Remote Sens.* 39, 1943–1958. doi: 10.1109/36.951085

Ni-Meister, W., Lee, S. Y., Strahler, A. H., Woodcock, C. E., Schaaf, C., Yao, T. A., et al. (2010). Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. *J. Geophys. Research-Biogeosciences* 115. doi: 10.1029/2009JG000936

Orwig, D., and Hadley, J. (2022). "Leaf Area Index at Harvard Forest HEM and LPH Towers since 1998 ver 23. EDI (Environmental Data Initiative) Data Portal. doi: 10.6073/pasta/912d4da0d326da63d82e93de68ca5ad4

Pan, M., Wood, E. F., Wojcik, R., and Mccabe, M. F. (2008). Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation. *Remote Sens. Environ.* 112, 1282–1294. doi: 10.1016/j.rse.2007.02.039

Persson, Å., Söderman, U., Töpel, J., and Ahlberg, S. (2005). Visualization and analysis of full-waveform airborne laser scanner data. *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.* 36, 103–108.

Pope, G., and Treitz, P. (2013). Leaf area index (LAI) estimation in boreal mixedwood forest of ontario, Canada using light detection and ranging (LiDAR) and worldView-2 imagery. *Remote Sens.* 5, 5040–5063. doi: 10.3390/rs5105040

Potithep, S., Nagai, S., Nasahara, K. N., Muraoka, H., and Suzuki, R. (2013). Two separate periods of the LAI-VIs relationships using in *situ* measurements in a deciduous broadleaf forest. *Agric. For. Meteorol.* 169, 148–155. doi: 10.1016/j.agrformet.2012.09.003

Qu, Y. H., Han, W. C., and Ma, M. G. (2015). Retrieval of a temporal high-resolution leaf area index (LAI) by combining MODIS LAI and ASTER reflectance data. *Remote Sens.* 7, 195–210. doi: 10.3390/rs70100195

Silvan-Cardenas, J. L., and Wang, L. (2010). Retrieval of subpixel Tamarix canopy cover from Landsat data along the Forgotten River using linear and nonlinear spectral mixture models. *Remote Sens. Environ.* 114, 1777–1790. doi: 10.1016/j.rse.2010.04.003

Smith, A. M. S., Falkowski, M. J., Hudak, A. T., Evans, J. S., Robinson, A. P., and Steele, C. M. (2009). A cross-comparison of field, spectral, and lidar estimates of forest canopy cover. *Can. J. Remote Sens.* 35, 447–459. doi: 10.5589/m09-038

Sun, Y. H., Qin, Q. M., Ren, H. Z., and Zhang, Y. (2022). Decameter cropland LAI/FPAR estimation from sentinel-2 imagery using google earth engine. *IEEE Trans. Geosci. Remote Sens.* 60, 4400614. doi: 10.1109/TGRS.2021.3052254

Tang, H., Dubayah, R., Brolly, M., Ganguly, S., and Zhang, G. (2014). Large-scale retrieval of leaf area index and vertical foliage profile from the spaceborne waveform lidar (GLAS/ICESat). *Remote Sens. Environ.* 154, 8–18. doi: 10.1016/j.rse.2014.08.007

Tang, H., Dubayah, R., Swatantran, A., Hofton, M., Sheldon, S., Clark, D. B., et al. (2012). Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. *Remote Sens. Environ.* 124, 242–250. doi: 10.1016/j.rse.2012.05.005

Thomas, V., Noland, T., Treitz, P., and Mccaughey, J. H. (2011). Leaf area and clumping indices for a boreal mixed-wood forest: lidar, hyperspectral, and Landsat models. *Int. J. Remote Sens.* 32, 8271–8297. doi: 10.1080/01431161.2010.533211

Tseng, Y. H., Lin, L. P., and Wang, C. K. (2016). Mapping CHM and LAI for heterogeneous forests using airborne full-waveform liDAR data. *Terrestrial Atmospheric Oceanic Sci.* 27, 537–548. doi: 10.3319/TAO.2016.01.29.04(ISRS)

Turner, D. P., Cohen, W. B., Kennedy, R. E., Fassnacht, K. S., and Briggs, J. M. (1999). Relationships between leaf area index and Landsat TM spectral vegetation indices across three temperate zone sites. *Remote Sens. Environ.* 70, 52–68. doi: 10.1016/S0034-4257(99)00057-7

Verrelst, J., Camps-Valls, G., Munoz-Mari, 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. Photogrammetry Remote Sens.* 108, 273–290. doi: 10.1016/j.isprsjprs.2015.05.005

Verrelst, J., Malenovsky, Z., van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J. P., Lewis, P., et al. (2019). Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods. *Surveys Geophys.* 40, 589–629. doi: 10.1007/s10712-018-9478-y

Wagner, W., Ullrich, A., Ducic, V., Melzer, T., and Studnicka, N. (2006). Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *Isprs J. Photogrammetry Remote Sens.* 60, 100–112. doi: 10.1016/j.isprsjprs.2005.12.001

Wang, Y., and Fang, H. L. (2020). Estimation of LAI with the liDAR technology: A review. *Remote Sens.* 12 (20), 3457. doi: 10.3390/rs12203457

Wang, Y., Fang, H. L., Zhang, Y. H., Li, S. J., Pang, Y., Ma, T., et al. (2023). Retrieval and validation of vertical LAI profile derived from airborne and spaceborne LiDAR data at a deciduous needleleaf forest site. *Giscience Remote Sens.* 60 (1), 2214987. doi: 10.1080/15481603.2023.2214987

Weiss, M., Jacob, F., and Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* 236, 111402. doi: 10.1016/j.rse.2019.111402

Xu, J., Quackenbush, L. J., Volk, T. A., and Im, J. (2020). Forest and crop leaf area index estimation using remote sensing: research trends and future directions. *Remote Sens.* 12 (18), 2934. doi: 10.3390/rs12182934

Xu, L., Shi, S., Gong, W., Shi, Z. X., Qu, F. F., Tang, X. T., et al. (2022). Improving leaf chlorophyll content estimation through constrained PROSAIL model from airborne hyperspectral and LiDAR data. *Int. J. Appl. Earth Observation Geoinformation* 115, 103128. doi: 10.1016/j.jag.2022.103128

Yang, W. Z., Tan, B., Huang, D., Rautiainen, M., Shabanov, N. V., Wang, Y., et al. (2006). MODIS leaf area index products: From validation to algorithm improvement. *IEEE Trans. Geosci. Remote Sens.* 44, 1885–1898. doi: 10.1109/TGRS.2006.871215

Yang, X. B., Wang, C., Pan, F. F., Nie, S., Xi, X. H., and Luo, S. Z. (2019). Retrieving leaf area index in discontinuous forest using ICESat/GLAS full-waveform data based on gap fraction model. *Isprs J. Photogrammetry Remote Sens.* 148, 54–62. doi: 10.1016/j.isprsjprs.2018.12.010

Yang, G. J., Zhao, C. J., Liu, Q., Huang, W. J., and Wang, J. H. (2011). Inversion of a radiative transfer model for estimating forest LAI from multisource and multiangular optical remote sensing data. *IEEE Trans. Geosci. Remote Sens.* 49, 988–1000. doi: 10.1109/TGRS.2010.2071416

Zhang, F., Hassanzadeh, A., Kikkert, J., Pethybridge, S. J., and Van Aardt, J. (2022). Evaluation of leaf area index (LAI) of broadacre crops using UAS-based liDAR point clouds and multispectral imagery. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 15, 4027–4044. doi: 10.1109/JSTARS.2022.3172491

Zhang, Z. J., Xie, H., Tong, X. H., Zhang, H. W., Liu, Y., and Li, B. B. (2020). Denoising for satellite laser altimetry full-waveform data based on EMD-Hurst analysis. *Int. J. Digital Earth* 13, 1212–1229. doi: 10.1080/17538947.2019.1698665

Zhang, Y., Yang, Y. Z., Zhang, Q. W., Duan, R. Q., Liu, J. Q., Qin, Y. C., et al. (2023). Toward multi-stage phenotyping of soybean with multimodal UAV sensor data: A comparison of machine learning approaches for leaf area index estimation. *Remote Sens.* 15 (1), 7. doi: 10.3390/rs15010007

Zhao, F., Yang, X. Y., Schull, M. A., Roman-Colon, M. O., Yao, T., Wang, Z. S., et al. (2011). Measuring effective leaf area index, foliage profile, and stand height in New England forest stands using a full-waveform ground-based lidar. *Remote Sens. Environ.* 115, 2954–2964. doi: 10.1016/j.rse.2010.08.030

Zheng, G., and Moskal, L. M. (2009). Retrieving leaf area index (LAI) using remote sensing: theories, methods and sensors. *Sensors* 9, 2719–2745. doi: 10.3390/s90402719

Zhou, G. Q., Deng, R. H., Zhou, X., Long, S. H., Li, W. H., Lin, G. C., et al. (2022). Gaussian inflection point selection for liDAR hidden echo signal decomposition. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/LGRS.2021.3107438

Zhou, G. Q., Long, S. H., Xu, J. S., Zhou, X., Song, B., Deng, R. H., et al. (2021). Comparison analysis of five waveform decomposition algorithms for the airborne liDAR echo signal. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 14, 7869–7880. doi: 10.1109/JSTARS.2021.3096197

Zhou, X. J., Wang, P. X., Tansey, K., Zhang, S. Y., Li, H. M., and Tian, H. R. (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 Review
PUBLISHED 04 October 2023
DOI 10.3389/fpls.2023.1219673

Check for updates

OPEN ACCESS

EDITED BY
Bo Li,
Syngenta, United Kingdom

REVIEWED BY
Chenglong Huang,
Huazhong Agricultural University, China
Jun Liu,
Shandong Provincial University Laboratory for Protected Horticulture, China

*CORRESPONDENCE
Malcolm J. Hawkesford
malcolm.hawkesford@rothamsted.ac.uk
Toby W. Waine
t.w.waine@cranfield.ac.uk
Fady Mohareb
f.mohareb@cranfield.ac.uk

RECEIVED 09 May 2023
ACCEPTED 07 September 2023
PUBLISHED 04 October 2023

CITATION
Cudjoe DK, Virlet N, Castle M, Riche AB, Mhada M, Waine TW, Mohareb F and Hawkesford MJ (2023) Field phenotyping for African crops: overview and perspectives.
Front. Plant Sci. 14:1219673.
doi: 10.3389/fpls.2023.1219673

COPYRIGHT
© 2023 Cudjoe, Virlet, Castle, Riche, Mhada, Waine, Mohareb and Hawkesford. 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.

Field phenotyping for African crops: overview and perspectives

Daniel K. Cudjoe[1,2], Nicolas Virlet[1], March Castle[1], Andrew B. Riche[1], Manal Mhada[3], Toby W. Waine[2*], Fady Mohareb[2*] and Malcolm J. Hawkesford[1*]

[1]Sustainable Soils and Crops, Rothamsted Research, Harpenden, United Kingdom, [2]School of Water, Energy and Environment, Cranfield University, Cranfield, Bedfordshire, United Kingdom, [3]AgroBiosciences Department, Mohammed VI Polytechnic University (UM6P), Benguérir, Morocco

Improvements in crop productivity are required to meet the dietary demands of the rapidly-increasing African population. The development of key staple crop cultivars that are high-yielding and resilient to biotic and abiotic stresses is essential. To contribute to this objective, high-throughput plant phenotyping approaches are important enablers for the African plant science community to measure complex quantitative phenotypes and to establish the genetic basis of agriculturally relevant traits. These advances will facilitate the screening of germplasm for optimum performance and adaptation to low-input agriculture and resource-constrained environments. Increasing the capacity to investigate plant function and structure through non-invasive technologies is an effective strategy to aid plant breeding and additionally may contribute to precision agriculture. However, despite the significant global advances in basic knowledge and sensor technology for plant phenotyping, Africa still lags behind in the development and implementation of these systems due to several practical, financial, geographical and political barriers. Currently, field phenotyping is mostly carried out by manual methods that are prone to error, costly, labor-intensive and may come with adverse economic implications. Therefore, improvements in advanced field phenotyping capabilities and appropriate implementation are key factors for success in modern breeding and agricultural monitoring. In this review, we provide an overview of the current state of field phenotyping and the challenges limiting its implementation in some African countries. We suggest that the lack of appropriate field phenotyping infrastructures is impeding the development of improved crop cultivars and will have a detrimental impact on the agricultural sector and on food security. We highlight the prospects for integrating emerging and advanced low-cost phenotyping technologies into breeding protocols and characterizing crop responses to environmental challenges in field experimentation. Finally, we explore strategies for overcoming the barriers and maximizing the full potential of emerging field phenotyping technologies in African agriculture. This review paper will open new windows and provide new perspectives for breeders and the entire plant science community in Africa.

KEYWORDS

African crops, phenotypes, field phenotyping, high-throughput phenotyping, phenotyping infrastructures, low-cost phenotyping, African agriculture, precision agriculture

1 Introduction

The global demand for food is projected to increase in the coming decades, driven by population growth, climate change, pandemics, shifts in food consumption and biofuel use (Tilman et al., 2011; Godfray and Robinson, 2015; van Dijk et al., 2021). Ensuring that crop production is sufficient to meet future goals is a challenge for plant and agricultural sciences.

In Africa, agricultural crops provide food and income for smallholder farmers and consumers. Despite the huge agricultural potential, agricultural productivity in African countries continues to remain the lowest in the world (Bjornlund et al., 2020). Many studies have indicated that yields of several important staple crops may be stagnating or even declining across the continent (Roudier et al., 2011; Knox et al., 2012; Ray et al., 2012; Parkes et al., 2018). This is the case for key staple crops such as maize, rice, wheat, millet, sorghum, cowpea, cassava and yam, which together account for a large portion of the population's diet. Therefore, food supply systems would be negatively affected if yield gains in these crops continue to slow due to environmental stresses and production constraints.

Addressing food security in Africa is a vast challenge that needs to be tackled in many complementary directions. Infrastructure development adapted to local needs, good farming practices, management, and political will are some of the major axes of development for food security. Improving crop performance and tolerance/resistance to biotic and abiotic conditions is the challenge facing the scientific community and innovative methods are needed.

Advanced field phenotyping, e.g. using digital approaches, has developed substantially over the past decade and provides means for real-time monitoring of response to environmental stresses and nutrition, and aids unravelling the relationships between yield and complex genotypic traits. The identification of genotypes with superior traits of agricultural interest remains one of the major targets for the genetic improvement of crops (Varshney et al., 2021).

The genomes of many agricultural crops such as rice (Matsumoto et al., 2005), sorghum (Paterson et al., 2009), maize (Schnable et al., 2009), soybean (Schmutz et al., 2010) and recently wheat (Appels et al., 2018) have been sequenced. However, the advances made in genomic approaches such as maker-assisted selection and high-throughput sequencing (Crossa et al., 2017; Scheben et al., 2018) are yet to be complemented with accurate field phenotyping methods (Minervini et al., 2015). Most of the traits of agronomic relevance (e.g., yield) are complex, and quantitative, requiring tools for their phenotypic assessment in the field (Reynolds et al., 2020). Furthermore, open field rather than controlled environment measurements are more likely to be useful in identifying genotypes that will perform better in farming practice, especially when large plots that mimic real farm conditions (i.e., environmental and management conditions) are employed (Rebetzke et al., 2014).

In addition, precision agriculture (PA) is becoming increasingly important in today's technologically advanced world (Langemeier and Boehlje, 2021; Gobezie and Biswas, 2023) and PA remains one of the cardinal principles of field phenotyping. The PA farming management concept relies on modern digital techniques to monitor and optimize agricultural production processes to improve crop performance (Hedley, 2015; Gokool et al., 2023). Despite PA's contributions to sustainable agriculture, its use in resource-constrained smallholder farming environments, particularly in Sub-Saharan Africa (SSA), has been very limited (Gobezie and Biswas, 2023). Recent developments in sensor technologies, machine vision, and higher-resolution digital cameras, in tandem with advanced data processing power and other portable tools have paved the way for high-throughput plant phenotyping in the field to benefit crop breeding programs (Deery et al., 2014; Zhang et al., 2016; Araus et al., 2022; Ahmed et al., 2023). From the field phenotyping perspective, these emerging technologies are enabling automated intensive data collection and increasing the ability to investigate plant function and structure through non-invasive methods with high accuracy. Such field phenotyping methods will aid crop improvement efforts to meet the expected demand for food and agricultural products in the future.

The development and application of these high-throughput tools for field phenotyping are currently focused on the main staple crops grown in the most developed agricultural regions. Over the decades, breeders and agronomists in Africa have used traditional phenotyping based on manual methods either for selecting traits or for improving yields through changes in agronomic practices (Iizumi and Sakai, 2020). However, traditional phenotyping in breeding is time-consuming, laborious and data collection is insufficient to fulfil the needs of plant breeders which impedes breeding progress. Therefore, further advances in phenotyping methods and appropriate implementation are required to increase the effectiveness of selection in breeding programs, speed up genetic gains, reduce costs and enable monitoring of plant status more efficiently than is currently feasible. The sophistication and cost of current plant phenotyping equipment (Reynolds et al., 2019) have restricted them from being widely applied in the developing world and especially in Africa. Additionally, insufficient technical, operational, regulatory restrictions and conceptual capacity in the plant science community have further limited implementation. Therefore, it is timely to begin to apply these technologies more widely, both geographically and with respect to target crops in Africa. Affordable high-throughput phenotyping aims to achieve reasonably priced solutions for all the components comprising the phenotyping pipeline which will promote their adoption for the breeding of African crops (Whalen and Yuhas, 2019; Bongomin et al., 2022).

Few studies have covered the use of modern field phenotyping approaches employing remote sensing in Africa (e.g., Mutanga et al., 2016; Chivasa et al., 2017; Buchaillot et al., 2019; Bongomin et al., 2022; Kassim et al., 2022). For instance, Bongomin et al. (2022) recently reviewed the status of field phenotyping in Uganda with focus on the application of drones and image analytics.

In this review, we provide a background on African agriculture and cover the concept of digital field phenotyping, focused on traits that may be measured by emerging technologies and which could be applicable to African crops. The current developments of field phenotyping in Africa, including initiatives, implementation

challenges and prospects are comprehensively reviewed. We observed that the lack of suitable field phenotyping infrastructures and approaches using digital technologies is limiting the development of improved crop cultivars and will negatively affect the agricultural industry and food security in Africa. We emphasize the potential for incorporating cutting-edge, low-cost phenotyping tools (i.e., portable field sensors, UAVs) into breeding schemes and for identifying agricultural crop responses to environmental constraints through field experimentation. Finally, we consider policy directions for tackling the implementation challenges (i.e., practical, financial, geographical and political) of digital field phenotyping and realizing the full potential of available field phenotyping resources (i.e., technologies, tools and know-how) appropriate for African crops.

2 African crops and the challenges to production

African countries are important producers of major crops with diverse agro-climatic and ecological conditions, and cultural diversity (Leakey et al., 2022). Sub-Saharan West Africa is composed of a wide variety of ecosystems and an equally high number of production systems (https://www.fao.org/3/AC349E/ac349e04.htm). Generally, crop production is concentrated in areas with a favourable combination of agro-bioclimatic conditions. In the Sahelian zone, cereals such as millet and sorghum are the predominant crops with annual rainfall (200-600 mm), transitioning to maize, groundnuts and cowpeas farther south in the Sudanian savannah zone (the so-called "Middle Belt"). These food crops are among the top five harvested crops in the Sahelian countries – Burkina Faso, Senegal, Mauritania, Mali, Chad and Niger. According to FAOSTAT (2018a) data, maize is the major essential staple food in sub-Saharan Africa, accounting for nearly 20% of total calorie intake. The same source indicates that in Sub-Saharan West Africa, millet and sorghum account for roughly 64% of total cereal production. Across the rainy forests of the Guinean zone (1200-2200 mm of rainfall per year) crops are predominantly root and tuber crops such as cassava and yams which are mostly cultivated in Ghana, Nigeria, Côte d'Ivoire and Sierra Leone. Yam is the second most important crop in Africa in terms of production after cassava (FAOSTAT, 2018a). Rice, on the other hand, is one of the most widely harvested crops in this humid zone, ranking first in Guinea, Liberia and Sierra Leone in terms of area harvested (Soullier et al., 2020; Duvallet et al., 2021).

Crop production in West Africa is mostly rainfed and crop production is vulnerable to climate change, which manifests itself in unpredictably high temperatures and erratic rainfall patterns (Sultan and Gaetani, 2016; Affoh et al., 2022). The five principal crops in West Africa in terms of harvested area (in millions of hectares per year on average in the last decade) are cassava (81), maize (19), millet (10), sorghum (12), yam (57) (FAOSTAT, 2022). Major cash crops are cocoa, coffee and cotton. Declining soil fertility and unpredictable climate change impacts (among other factors) have made it difficult to maintain the yields of these major crops (Shimeles et al., 2018). Over the last three decades, the agricultural sector in West Africa has been characterized by strong production growth in some major staple crops culminating in increased production volumes for both domestic and export markets (Blein et al., 2008; FAO, 2015). Similarly to West Africa, Central Africa's principal food crops include cassava, peanuts, sorghum, millet, maize, sesame and plantains. Additionally major cash crops for export include cotton, coffee and tobacco (Ochieng et al., 2020).

In Northern Africa, particularly Morocco, crop production is regionally diverse owing to different climatic conditions, agro-ecological zones, land-crop tenure and farming systems (Ouraich and Tyner, 2018). This geographical diversity results in varied agriculture, with crops ranging from cereals and vegetables to fruits and nuts, grains, legumes, etc., that contribute significantly towards the country's agricultural sustainability and food security. Cereal production accounts for 65% of cultivable agricultural areas (Ouraich and Tyner, 2018). Most cereal production occurs under rainfed conditions. As a result, productivity performance is influenced by precipitation levels. For instance, 7.3 million tonnes of wheat were produced in 2018 making it the 20th largest producer in the world and 2.8 million tonnes of barley being the 15th largest producer in the world (FAOSTAT, 2018b). However, drought is a persistent threat to crop production especially the lowlands where cereals are grown are particularly at risk because of the wide variations in annual precipitation (Verner et al., 2018; Meliho et al., 2020). In recent years, quinoa has sparked particular attention in Morocco (Choukr-Allah et al., 2016; Hirich et al., 2021). It remains one of the most nutrient-dense crops and is recognized as a 'Superf Food' due to its nutritional benefits. Thus, Morocco is one of the few North African countries capable of achieving self-sufficiency in food production (Saidi and Diouri, 2017).

Grains and cereals (e.g., maize, wheat, barley, oats and sorghum) are South Africa's most important crops occupying more than 60% of the acreage under cultivation (FAO, 2022). Together, these crops account for one of the largest agricultural industries contributing more than 30% to the total gross value of agricultural production (FAO, 2022). Maize, the country's most important crop and largest locally produced field crop, is a dietary staple supplying most of the carbohydrate needs, a source of livestock feed and is an export crop (Epule et al., 2022).

The country has emerged as the largest maize producer and exporter in the Southern African Development Community (SADC) region and Africa as a whole (Fisher et al., 2015; FAO, 2022). According to the FAO, 2022, in 2021 South Africa produced 17 million metric tonnes of maize, making it the 9th largest producer in the world. Moreover, it produced 2.6 million metric tonnes of potato and 2.3 million metric tonnes of wheat. Largely, South Africa has a semi-arid climate characterized by summer and winter rainfall seasons. Unpredictable weather conditions due to climate change have a severe impact on maize and wheat production which accounts for more than 36% of the total value of field crops (Bradshaw et al., 2022).

Smallholder farmers dominate agriculture in East African countries, contributing up to 90% of total agricultural production (Salami et al., 2010; Livingston et al., 2011). A cereal-legume mixed cropping pattern is the dominant system that includes maize, millet,

sorghum and wheat (Van Duivenbooden et al., 2000). Over 40% of the region is covered by the maize mixed cropping system, which is followed by the pastoral (14%), root crop (12%) and cereal-root crop mixed system (11%) (Adhikari et al., 2015). Teff is a significant crop in the Ethiopian highlands, while other significant crops in the area include cassava, bananas and rice. The mixed cropping system in East Africa is based on millet in the drier regions and on maize and cassava in the humid regions (Adhikari et al., 2015). The main cash crops in most of the East African countries in SSA are coffee, tea, cotton, tobacco and sugarcane. Rainfall variability negatively impact on crop production in East African countries (Palmer et al., 2023). Generally, the major challenges to crop production in Africa are unproductive soils, pests and diseases, drought, and poor crop management (Tadele, 2017). The distribution of major crops in each sub-region except Northern Africa is summarized in Figure 1.

3 Digital and image-based field phenotyping

Experiments with repeated trials in diverse environments are often necessary to screen plants for desirable traits. This becomes problematic when there is the need to screen a large panel of genotypes for valuable traits (i.e., yield potential or abiotic and biotic stress tolerance) to assess genotype, environment, and management (G × E× M) interactions (Araus and Cairns, 2014). Over the years, the measurement of individual plants in controlled conditions has dominated most of the phenotyping research. However, controlled environments often do not accurately mimic plant growth and development in field conditions (White et al., 2012). Field phenotyping is becoming more widely recognized as the approach that gives the most accurate representation of traits in real-world cropping systems (Tariq et al., 2020). Thus, field phenotyping is an important component of crop improvement to assess how the genotype, the environment, and their interaction (G × E) influence quantitative traits in a complex and dynamic manner (Fiorani and Schurr, 2013; Araus and Cairns, 2014; Neilson et al., 2015). Furthermore, field phenotyping is employed to discover novel traits, identify new germplasm carrying relevant but complex traits for breeding, and for testing proof of concept to validate traits (Watt et al., 2020). Traditionally, destructive sampling has been used to quantify certain observable plant traits, including laboratory analysis to characterize phenotypes based on their genetic and physiological functions. Digital phenotyping approaches seek to reduce this need (Tripodi et al., 2022; Virlet et al., 2022).

Different measurement approaches including novel technologies such as non-invasive imaging, robotics and sensor positioning systems have been incorporated in well-designed field phenotyping installations for high-throughput phenotyping (e.g., Araus and Cairns, 2014; Kirchgessner et al., 2017; Shakoor et al., 2017; Virlet et al., 2017; Pieruschka and Schurr, 2022). These significant strides in field phenotyping have fostered a major international collaborative effort directed toward data and protocol standardization (Pieruschka and Schurr, 2019; Lorence and Jimenez, 2022). The appeal of these platforms is the increased throughput and objectivity in data collection compared to traditional field approaches.

Non-invasive portable devices, ground-wheeled, motorized gantry scanalyzer systems, agricultural robots and aerial vehicles that deploy a wide range of cameras and sensors, together with high-performance computing are currently required to conduct field phenotyping in a timely and economical manner (Figure 2). Together, these platforms are able to phenotype plant characteristics throughout the season in field environments (White et al., 2012; Fritsche-Neto and Borém, 2015; Jimenez-Berni et al., 2018; Furbank et al., 2019; Li et al., 2021).

In recent years, manned and unmanned aerial vehicle (UAV) remote sensing platforms have emerged as convenient high-throughput tools for field phenotyping (Pajares, 2015; Shi et al.,

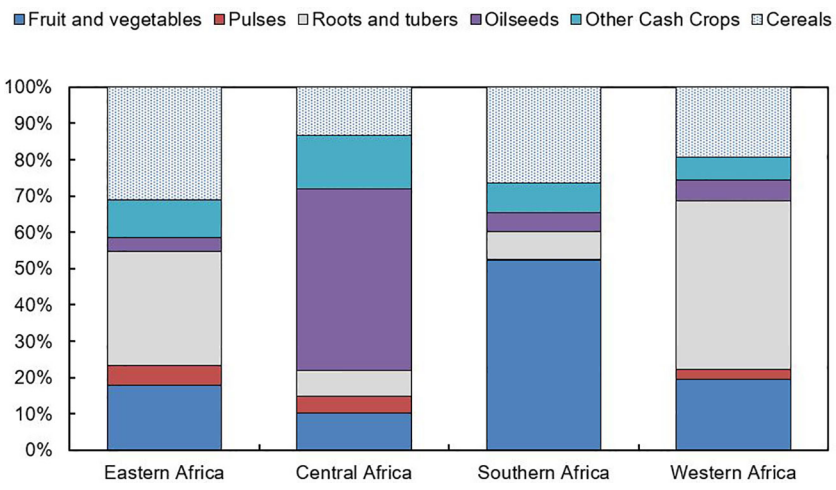


FIGURE 1
Major crop distribution in Sub-Saharan African region based on average production values between 2011-13. Adapted from FAOSTAT (2016). FAO, http://faostat3.fao.org/.

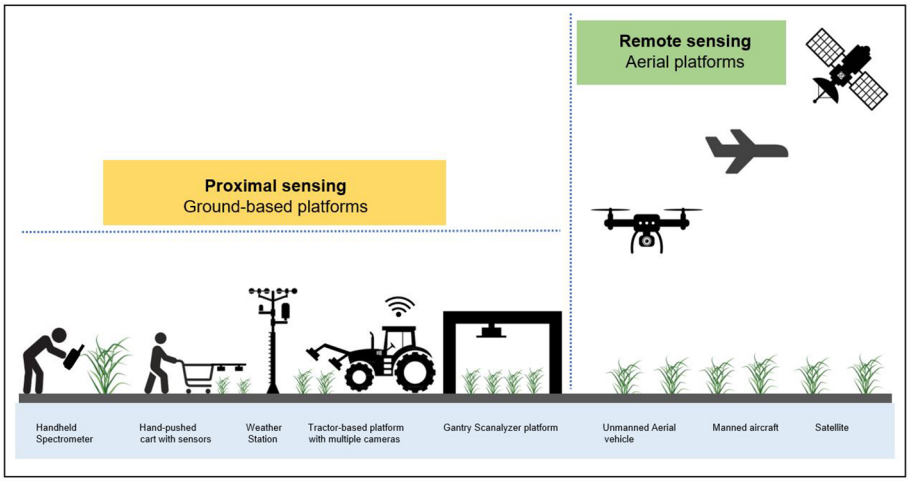


FIGURE 2
Overview of the most common field phenotyping systems and approaches at proximal and remote sensing scales. The proximal sensing approach is based on ground-based platforms such as handheld spectrometers, hand-pushed carts equipped with sensors, tractor-based platforms fitted with multiple cameras and gantry scanalyzer systems that collect spectral information of crops in close range or contact. On the other hand, the remote sensing technique is based on aerial platforms including unmanned aerial vehicles (i.e., drones), manned aircraft and satellites that acquire spectral imagery of crops without making physical contact but at a distance. Figure 2 was modified from (Pineda et al., 2021).

2016; Feng et al., 2021). These remote sensing approaches, particularly UAVs enable quick and non-destructive high throughput phenotyping, with the benefit of adaptable and convenient operation (Yang et al., 2017a). These phenotyping platforms can combine multiple sensors such as digital cameras, infrared thermal imagers, light detection and ranging (LiDAR), multispectral cameras and hyperspectral sensors for various assessments of morphological and physiological plant traits (Gonzalez-Dugo et al., 2015; Yang et al., 2017a; Camino et al., 2018; Roitsch et al., 2019).

Alternatively, field phenotyping can be accomplished on the ground utilizing a fully automated fixed-site phenotyping platform (e.g., Kirchgessner et al., 2017; Virlet et al., 2017; Bai et al., 2019), hand-held sensors, portable spectroradiometers, hand-pushed carts or high-clearance tractors carrying multiple high-resolution sensors to measure phenotypic features non-destructively (Comar et al., 2012; Andrade-Sanchez et al., 2014; Crain et al., 2016). The use of rapid non-invasive portable devices that carry sensors for crop status monitoring has advanced field data collection due to their applicability and ease of operation (Parks et al., 2012; Yang et al., 2014; Condorelli et al., 2018). Recently, field phenotyping has become more flexible by integrating ground-based and aerial platforms (Potgieter et al., 2018; Furbank et al., 2019; Ninomiya, 2022). Table 1 summarizes the diverse ground-based and aerial field phenotyping platforms, their applications, advantages, and limitations.

4 Traits assessed by sensor platforms and their relevance for field phenotyping

For field phenotyping, traits that have been evaluated by sensors in the field have been reviewed recently by Watt et al. (2020) and include for example; (a) plant morphological development (i.e., including seed establishment and growth of the crop, the timing, and dynamics of flower and fruit development); (b) functional traits that are related to the photosynthetic capacity and carbon uptake during the phenological growth phase; (c) traits related to biotic and abiotic stress resistance/tolerance; (d) traits that determine crop water status (e.g., water uptake and transpiration and water-use efficiency) of plants; (e) yield-related traits and harvest quality of crops (i.e., biomass yield) and (f) the structural and functional root traits (i.e., root architecture). These traits have been previously classified into morphometric and physiological parameters (Qiu et al., 2018). Traits such as plant height, stem diameter, leaf area or leaf area index, leaf angle, stalk length and in-plant space are morphometric parameters. Physiological parameters include traits such as photosynthetic rate, chlorophyll content, water stress, leaf water content, biomass, and salt resistance, which together can impact plant growth. It should be emphasized that different phenotypic traits have specific time frames within the phenological cycle of the plant when they are relevant for the breeder and farmer. Currently, the most researched crops in field phenotyping are economic crops, such as wheat, maize, barley, sorghum, tomato, bean and grape because they have significant economic value for agricultural development. A challenge is to extend phenotyping into the vast range of African crops, some of which may be of only local importance.

Field phenotyping makes use of a variety of sensors due to the large number of phenotypic traits that must be measured. Several conventional and novel sensors such as digital cameras, range cameras, depth cameras, spectral sensors, lidar or laser sensors, thermal sensors, fluorescence sensors, multispectral cameras, hyperspectral cameras and others are employed and integrated for plant trait measurement in field phenotyping (Qiu et al., 2018; Roitsch et al., 2019; Xie and Yang, 2020).

TABLE 1 Applications and limitations of field phenotyping platforms.

Phenotyping platform	Examples	Applications	Advantages	Limitations	References
Ground-based platforms					
Fixed-site systems	Field scanalyzers (i.e Rothamsted field scanalyzer, Maricopa field scanalyzer)	Ground cover, canopy height, plant geometry, growth, growth stages, vegetation indices, chlorophyll fluorescence parameters	Unmanned continuous operation with good repeatability, deploy a wide range of sensors, fully automated. Not limited by soil conditions	Expensive, monitor a limited number of plots, limited by weather conditions	Virlet et al., 2017; Burnette et al., 2018
Permanent platforms based on a cable-suspended multi-sensor system	The ETH field phenotyping platform, the University of Nebraska phenotyping system	Monitor canopy cover, canopy height, and traits related to thermal and multi-spectral imaging with selected examples from winter wheat, maize, and soybean	Produce precise, high-resolution images, deploy a wide range of sensors, fully automated	Monitor a limited area of crop, difficult to move, expensive, and limited by weather conditions	Kirchgessner et al., 2017; Bai et al., 2019
Handheld sensors	Point spectroradiometers, thermal sensors, chlorophyll meters, imagers	Estimate chlorophyll fluorescence, canopy temperature, nitrogen status, leaf area, plant height, yield	Ground truth reference to validate aerial measurements (UAVs) and airplanes, low-cost and easy to use	Labour intensive and time-consuming, limited plot coverage, measurement bias	Yang et al., 2014; Andrianto et al., 2017; Garriga et al., 2017
In-field mobile platforms	Phenocart, proximal sensing cart, phenomobiles, manned buggies	Estimate biomass, leaf area index, counting plants, plant height, early vigour, and plant maturity	Manually operated, low-cost, easier to construct, multiple traits evaluations, deploy more sensors, flexibility with payload and view angle geometry; very adaptable	The motorized platforms are costly to construct and run, need technical expertise, hard operation for large-scale experiments. Limited by weather and soil conditions	White and Conley, 2013; Andrade-Sanchez et al., 2014; Deery et al., 2014; Crain et al., 2016
Aerial platforms					
Unmanned aerial vehicles (UAVs)	Broadly classified into Rotocopters, fixed wing systems, parachutes, and blimps	Traits such as canopy cover, canopy height, crop lodging, growth indices, and canopy temperature can be estimated from the imagery	Rotocopters (i.e drones) can deploy a wide range of sensors, including thermal, multispectral, and hyperspectral cameras, high hovering capabilities, better flight time	Lower speeds for image stitching, lens distortion, and overlap of the acquired images can affect orthomosaic, battery use and flying time may be limited by the payload, and operability is limited in windy, wet, dull, variable light, or cold conditions	Sankaran et al., 2015; Zaman-Allah et al., 2015; Chawade et al., 2019; Holman, 2020
Satellite imaging	Digital Globe WorldView-2 satellite, WorldView-3 satellite, RADARSAT-2	In precision agriculture for germplasm evaluation, multi-location yield trials, field observation of crop biophysical parameters, weather predictions	Evaluation of moderate to large-sized trial, multi-location evaluation; provides automated coverage of isolated field trials across a larger geographical area	Affected by weather conditions, resolution, frequency of imaging, takes a long time from image acquisition to access, costly, higher frequency of satellite revisits, cloud cover can interfere with imaging	Tattaris et al., 2016; Yang et al., 2017b; Yang, 2018

Modified from Li et al., 2014 and Deery et al., 2014.

Since plants develop rapidly during their early growth stages, frequent measurements during their establishment are a prerequisite for the quantitative selection of vigour phenotypes. Drones fitted with conventional RGB (red-green-blue) cameras, in combination with advanced image processing pipelines, can automatically detect crop stands (single plants) and determine seed emergence, germination rates and timing under extreme climatic events in the field (Liu et al., 2017).

Most plants display strong morphological changes during their phenological development, which is greatly influenced by the availability of resources and changes in abiotic and biotic factors. Therefore, the development of robust, automated, and precise methods to measure morphological plant traits in field conditions is still required (Gibbs et al., 2017).

The leaf is one of the important components of a plant. It plays a major role in plant growth given that its growing status influences the efficiency of the direct solar energy utilization by plants. Hence, it is a significant parameter in plant phenotyping. Measurements of morphometric parameters of the leaf and other canopy features (i.e., leaf area, stem height, number of tillers, and inflorescence architecture) have been evaluated using non-destructive multi-sensor approaches (Busemeyer et al., 2013; Fiorani and Schurr, 2013; Rahaman et al., 2015). However, the most frequently used geometric measure of plant canopy is the green leaf area index (GLAI), which relates the one-sided green leaf area per unit projected ground area (Chen and Black, 1992). For instance, UAV multispectral imagery has been used to characterize GLAI

dynamics of a large maize panel under contrasted environmental conditions and thus holds great potential for yield predictions in breeding programs (Blancon et al., 2019). LAI can also be evaluated, indicating plant coverage, from spectral images (Dammer et al., 2016; Schirrmann et al., 2016).

Plant canopy architecture and other morphological traits of plant organs have been measured concurrently with 3D proximal sensing techniques. A body of recent reviews has compared the performances of the most common 3D sensors for high-throughput plant phenotyping (Li et al., 2014; Qiu et al., 2019). The 3D acquisition devices and approaches commonly used are LiDAR time-of-flight cameras, mono, and multi-view stereo vision and structure-from-motion. The LiDAR sensors can scan and extract morphological traits of plant organs from 3D point clouds. For example, LiDAR was used to estimate plant height, ground cover and above-ground biomass in wheat (Jimenez-Berni et al., 2018). However, LiDAR sensors are expensive (Li et al., 2014), take significant time and there is a need to increase scanning time to increase the spatial resolution. Deploying a UAV-based system may reduce this challenge.

Plant height is a key indicator for canopy structure, yield, carbohydrate storage capacity and lodging occurrence (Holman et al., 2016; Hassan et al., 2019). Additionally, it has significant applications in predicting biomass, identifying plant cultivars, plant stress and phenological stages (Aasen et al., 2015). The traditional method of measuring height using a metre rule is labor-intensive, cumbersome and low throughput. In recent years, the development of drones and imaging sensors that capture high-resolution images has enabled high-throughput plant height estimation. For instance, Holman et al. (2016) estimated wheat height using UAV-based RGB images and terrestrial LiDAR.

Chlorophyll is a vital plant trait because it is strongly related to crop physiological status and may be indicative of photosynthetic rate, crop stress, nutrition status, yield, and plant productivity (Peng et al., 2011; Maimaitijiang et al., 2017). The most popular tools for evaluating vegetation health using visible and near-infrared light are spectral sensors. Chlorophyll meters such as the SPAD-502 are frequently used instruments to measure the relative chlorophyll content. Handheld chlorophyll meters and fluorescence meters have been used to assess plant nitrogen status, photosynthesis, yield and its components in crops (Yang et al., 2014; Andrianto et al., 2017; Fernández-Calleja et al., 2020). Additionally, chlorophyll can be measured using NDVI sensors and portable spectrometers in the field (Bai et al., 2016).

Crop nitrogen content can serve as a proxy for soil fertilizer availability, assisting farmers in precision nitrogen application to the soil. UAV-based hyperspectral imaging and ground-level optical sensors (SPAD-502, Duplex, and Multiplex) have been employed to estimate nitrogen fertilization status in maize (Quemada et al., 2014). In another study, Zaman-Allah et al. (2015) used a UAV equipped with a multispectral sensor (Green, Red, and NIR) to assess low nitrogen stress tolerance in corn. Additionally, vegetation indices (VIs) derived from spectral reflectance data captured by sensors devices such as the CropScan multispectral radiometer (Zhu et al., 2008), handheld spectroradiometers and the FieldSpec (Fitzgerald et al., 2006; Tilling et al., 2007; Feng et al., 2008), Tec5 (Erdle et al., 2013) can accurately measure nitrogen status in wheat and rice.

The above-ground biomass reflects light use efficiency and growth and is vital for carbon stock accumulation and monitoring (Swinfield et al., 2019). Brocks and Bareth (2018) estimated the biomass in barley using RGB images collected by UAV. Thermal infrared sensors are mostly used to detect crop water stress since they can provide temperature information for the crop (Park et al., 2017; Poblete et al., 2018; Bian et al., 2019). Thermal infrared sensors enable the estimation of canopy temperature which is a reflection of plant transpiration and plant water status. Kumar et al. (2020) used a proximal phenotyping cart (phenocart) mounted with low-cost consumer-grade digital cameras to characterize wheat germplasm for drought tolerance under field conditions. Plant yield has been considered an important agronomic trait for field phenotyping. Bascon et al. (2022) estimated rice yield using multispectral images.

The features of the sensors (e.g., spectral resolution, spatial resolution, specificity, and cost) should be considered according to the specific applications, phenotyping needs and context. In the African context, low-cost sensors and analysis pipelines which are not complex would benefit a broader user base for plant phenotypic trait assessments. The most successful trait assessment approach incorporates in time (throughout the crop cycle) and space (at the canopy level) the performance of the crop with respect to capturing resources (e.g., radiation, water and nutrients) and the efficiency of resource utilization (Araus et al., 2008). The aforementioned traits are discussed here with specific examples of sensors and automated measurement approaches used for their evaluation in the field (see Table 2). The advantages and limitations of each type of sensor are indicated.

5 Overview of the status of field phenotyping in Africa

Despite the recent advances in high-throughput field phenotyping based on the non-destructive analysis of plant traits, Africa has yet to consolidate the gains of these cutting-edge technologies for research into agricultural productivity. In terms of the deployment of high-end field phenotyping tools and approaches, Africa cannot keep pace with many regions, even in the era of artificial intelligence (AI), 'internet-of-things' (IoT) and technological advancements, although more affordable and lean phenotyping systems are now becoming available. Community-wide surveys and exchanges conducted by the International Plant Phenotyping Network (IPPN) and European Infrastructure for Multi-Scale Plant Phenomics and Simulation (EMPHASIS) within the growing phenotyping community in recent years have identified focus areas to assess the status of global plant phenotyping and crucial bottlenecks in the emerging field.

The major bottlenecks for developing field phenotyping in Africa were non-invasive phenotyping approaches, data management and cost among others (IPPN, 2016; Rosenqvist et al., 2019). This survey further reveals that in terms of using

TABLE 2 Emerging high-throughput phenotyping techniques and integrated sensor platforms applicable for plant trait assessment for field phenotyping.

Sensor	Examples	Crop species	Trait/phenotypic parameter	Applications	Advantages	Limitation	References
Hyperspectral sensor	VNIR, SWIR	Rice, Wheat	Nitrogen status, nitrogen use efficiency, water content, yield estimation, canopy components	Evaluate spectral properties, explore hyperspectral bands, estimate indices for fertilizer accumulated in plant organs, early detection of plant stress	Accurate estimation of nitrogen content and other biochemical or physiological status	Update model for new crop species, image processing is challenging, sensors are costly, large data size	Seiffert et al., 2010; Deery et al., 2014; Sadeghi-Tehran et al., 2021; Wang et al., 2021
Thermal sensor	Thermal infrared sensor, near-infrared camera, FLIR sensor	Wheat	Canopy temperature, drought tolerance, water use efficiency	Monitor crop temperature for abiotic stresses e.g., drought tolerance	Low-cost, precise and reliable in repeated experiments	Environmental factors have an impact on performance, very small temperature variations are undetectable, and cameras with higher resolution are heavier	Costa et al., 2013; Deery et al., 2019; Sagan et al., 2019
Visible light sensor	RGB sensor, visible light camera	Rice	Shoot growth, phenology, greenness, plant vigour, leaf area	Visible phenotype parameters, classification of crop organs, greenness, growth and health, time series of vegetation indices	Affordable sensors are available	Visual spectral bands and properties are limited, Changes in illumination conditions cause image blur and noise errors	Kipp et al., 2014; Guo et al., 2015
3D sensor	LIDAR (Light Detection and Ranging) sensor, 3D laser scanner	Maize, Wheat	Plant height, canopy cover, above-ground biomass, crop architecture	Extract morphological traits of plants organs from 3D point clouds; measuring crop height and volume	3D plant information can be quickly captured through close-range observation	LIDAR can be sensitive to small variations in path length, field applications can be challenging	Müller-Linow et al., 2015; Guo et al., 2018; Jimenez-Berni et al., 2018; Qiu et al., 2019
Fluorescence sensor	Fluorescence camera, LIFT fluorometer	Wheat	Photosynthetic capacity, chlorophyll content, quantum yield	Measure photosynthesis, chlorophyll, water stress	Automatic and rapid measurement of photosynthetic parameters	Limited for UAV imagery, can be affected by background noise, difficult to use in the field	Chaerle and Van Der Straeten, 2000; Zendonadi dos Santos et al., 2021
Multispectral sensor		Sorghum, Maize	Disease resistance, nutrient use efficiency, N content, biomass, grain yield	Multiple plant responses to nutrient deficiency, water stress, diseases, etc.,	High-resolution, fast	Sensors can be expensive, limited to a few spectral bands	Zaman-Allah et al., 2015; Zhao et al., 2021
Spectrometer		Maize	Water content, seed composition, yield	Leaf and canopy growth, disease evaluation, leaf area, chlorophyll content, canopy temperature, and crop responses	Handy and easy to use, inexpensive	The quality of the data may be affected by soil background, spectral mixing could occur, and sensor calibration required	Cozzolino, 2014; Andrianto et al., 2017; Chivasa et al., 2020; Cavaco et al., 2022

Modified from Zhao et al., 2019.

high-intensity field approaches (e.g., automation, robotics, image analysis and data storage management) for field phenotyping, Africa ranks lowest around the world. A recent survey conducted in the framework of the IPPN and EMPHASIS projects in 2020 (IPPN, 2020) which is reported by Yang et al. (2020) and Fahrner et al. (2021), indicated that Africa is still behind in the implementation of high-throughput field phenotyping. This highlights the need for a broader deployment of high-throughput field phenotyping techniques, which are essential enablers or resources for agricultural sciences and breeding to address upcoming crop production challenges.

The IPPN over the years has been promoting the idea of strengthening modern plant phenotyping in African countries by giving travel grants to Africa and inviting students and researchers for International Plant Phenotyping symposia and internships. However, only a few institutional members are identified for collaboration in the region. In recent times, there have been some high-throughput field phenotyping research and initiatives in African countries such as South Africa, Ghana, Senegal, Morocco, Nigeria, Ethiopia, Kenya, Egypt, and Zimbabwe which is encouraging for the emerging field and will be highlighted in this review (see section 5.2 and Table 3).

Like in many developing countries, field phenotyping in African countries is mostly based on conventional and traditional methodologies which rely heavily on manually recorded measurements of phenotypic data or visual assessment of plant parameters. It entails manually inspecting crops and measuring several crop characteristics that affect yield traits, including plant

TABLE 3 Summary of some major characteristics of field phenotyping activities implemented in some African countries.

Region	Country	Area of high-throughput field phenotyping research	Prospects	Reference/web link
West Africa	Ghana	Exploration of digital agriculture, deployment of low-cost sensors and technologies for breeding, exploration of remote sensing for precision agriculture, GIS	Digital agriculture, low-cost precision agriculture, and breeding, use of high-throughput tools	Hall et al., 2018; Kpienbaareh et al., 2019; Kassim et al., 2022; Sie et al., 2022; https://ftfpeanutlab.caes.uga.edu/Research/variety-development/high-throughput-phenotyping-in-senegal–ghana-and-uganda.html
	Senegal	Exploration of digital agriculture, exploration of UAV imagery, multi-spectral imaging, GIS	Development of high-throughput approaches, digital agriculture, low-cost precision breeding	Dingkuhn et al., 2015; Gano et al., 2021; https://www.devdiscourse.com/article/other/523595-senegals-embrace-of-the-digital-revolution-in-agriculture-marks-the-way-forward-for-africa
	Nigeria	Use of field mobile agricultural robots, digital imaging, remote sensing, machine learning, GIS, site-specific analytics, drone imagery	Development of high-throughput approaches, digital agriculture, low-cost precision breeding, deployment of digital technologies and innovations	Ifeanyieze et al., 2014; Daniel et al., 2016; Ejikeme et al., 2017; Iseki and Matsumoto, 2019; Alabi et al., 2022; Izuogu et al., 2023; https://nitda.gov.ng/wp-content/uploads/2020/11/Digital-Agriculture-Strategy-NDAS-In-Review_Clean.pdf
North Africa	Morocco	High-throughput phenotyping, precision field-based phenotyping platform for drought/heat tolerance, development of quinoa phenotyping methodologies, expanding the precision and prediction value of phenotyping/genotypic data for new germplasm emerging from the wheat, adding an HTPP system for wheat abiotic stresses	Expanding phenotyping capabilities, low-cost precision breeding, deployment of digital technologies and innovations, expansion in remote sensing capabilities	Bijaber et al., 2018; Danzi et al., 2019; Bouras et al., 2020; Jabir and Falih, 2020; Laachrate et al., 2020; Quahir et al., 2022; https://www.fao.org/in-action/plant-breeding/nuestrosasociados/africa/morocco/es/; https://www.icarda.org/research/projects/precision-field-based-phenotyping-platform-droughtheat-tolerance-morocco-pwpp
	Egypt	High-throughput precision phenotyping for improvement of drought and salt tolerance in wheat genotypes, implementation of digital technology (mobile applications) for field phenotyping, AI-enabled system to enhance agriculture process, satellite imagery for crop monitoring	Expanding phenotyping capacities, low-cost precision agriculture, and breeding	El-Shirbeny et al., 2014; Elsayed et al., 2017; Shokr, 2020; Bahn et al., 2021; Elmetwalli et al., 2022; Elsafty and Atallah, 2022; Abdelnabby and Khalil, 2023; Mahdy and Ahmad, 2023; Sayed et al., 2023; https://globalrust.org/geographic/egypt; https://www.fao.org/e-agriculture/news/egypt-turns-fao-digital-transformation-agriculture; https://dailynewsegypt.com/2021/12/07/government-launches-ai-enabled-system-to-enhance-agriculture-process/
Southern Africa	South Africa	Deployment of field scanalyzer (FieldScan) for spectral crop measurement, remote sensing for precision agriculture	Expanding phenotyping capacities, low-cost precision agriculture, and breeding, advancing remote sensing capabilities	Mutanga et al., 2016; Brewer et al., 2022a, b; Buthelezi et al., 2023

(Continued)

TABLE 3 Continued

Region	Country	Area of high-throughput field phenotyping research	Prospects	Reference/web link
	Zimbabwe	UAV-based high-throughput phenotyping, multispectral remote sensing in maize varietal response to maize streak virus (MSV) disease, high-throughput phenotyping of maize performance under phosphorus fertilization, remote sensing methodologies for crop monitoring under conservation agriculture	Expanding phenotyping capacities, low-cost precision agriculture, and breeding	Kefauver et al., 2015; Zaman-Allah et al., 2015; Gracia-Romero et al., 2018; Musungwini, 2018; Buchaillot et al., 2019; Chivasa et al., 2020; Gracia-Romero et al., 2020; Shonhe and Scoones, 2022; Parwada and Marufu, 2023;
East Africa	Kenya	Satellite-based assessment of maize yield variations in smallholder farms, GIS and remote sensing capabilities	Expanding phenotyping capacities, low-cost precision agriculture, and breeding	Kefauver et al., 2015; Kotikot and Onywere, 2015; Zaman-Allah et al., 2015; Burke and Lobell, 2017; Manzi and Gweyi-Onyango, 2021
	Ethiopia	GIS for precision agriculture, imaging technologies for crop trait analysis	Expanding phenotyping capacities, low-cost precision agriculture, and breeding	Alemaw and Agegnehu, 2019; Bontpart et al., 2020; Debesa et al., 2020; Beyene et al., 2022; Debalke and Abebe, 2022

height, number of tillers, leaf color, leaf shape, leaf area index (LAI), chlorophyll content, growth stages, above-ground biomass and stress tolerance (Gedil and Menkir, 2019; Bongomin et al., 2022; Badu-Apraku et al., 2023). In practice, in traditional field phenotyping, breeders or research evaluators inspect the trial fields and rate the plots according to how they feel, taste, smell, and appear (Kim, 2020). Such phenotyping methods have several disadvantages such as being low-throughput, time-consuming, laborious, expensive and error-prone (Chapu et al., 2022; Xiao et al., 2022). Although these methods have been beneficial in developing new crop cultivars and improved yields, it is crucial that more effective phenotyping methods be used to increase the accuracy of data collection.

In parallel, field phenotyping is undertaken to evaluate the agronomic performance of crops in breeding programs, germplasm collections and in biotechnology programs to deliver improved cultivars that can cope with environmental stresses (e.g., Asare-Bediako et al., 2019; Gedil and Menkir, 2019; Rezende et al., 2020; Kavhiza et al., 2022). These phenotyping research targets are focused on key crops for food security but are predominantly low-throughput phenotyping based on field trials. In sub-Saharan Africa, breeding programs championed by the Alliance for a Green Revolution in Africa (AGRA) have been dedicated to priority crops such as rice, maize, cassava, yam, beans, cowpea and vegetables under various regional breeding networks for improved varieties and seed systems (FAO, 2011; AGRA, 2019).

Previous studies have used a variety of calibration data, including ground-based survey methods and crop model simulations, to predict yield in smallholder systems (Burke and Lobell, 2017; Ogutu et al., 2018). However, there has been emerging evidence in SSA suggesting inaccurate farmer-reported crop production estimates in smallholder production systems (World Bank, 2010; Gourlay et al., 2017; Abay et al., 2019; Wahab, 2020). These anomalies in crop yield estimation at smallholder, country and regional levels can cause price fluctuations (i.e., inflation), wrong national policy decisions and food insecurity among others. High-throughput and/or digital phenotyping might offer a better estimation of regional and national crop production.

Recent advances in sensor technology and the availability of free high-resolution (spatial and temporal) multispectral satellite images have also presented an opportunity to predict the yield of maize (Chivasa et al., 2017) and detect leaf spot diseases in groundnut (Sie et al., 2022), adaptation responses to early drought stress in sorghum (Gano et al., 2021) as well as mapping spatial distribution on a near real-time basis for a region, which hitherto was not feasible.

5.1 Field phenotyping initiatives and programs in Africa

Despite the low implementation of high-throughput field phenotyping in Africa, there are some efforts by research organizations to adopt the technology in some countries. Prominent among these initiatives is a global network for precision field-based wheat phenotyping. (https://globalrust.org/content/global-network-precision-field-based-wheat-phenotyping). Based on a global network of wheat partners, field phenotyping platforms are being developed with the support of the CGIAR research program on wheat and co-investing national agricultural research centers around the world, including some African countries such as Kenya, Ghana, Nigeria, Ethiopia, and Morocco.

The main goal of this network is to generate high-quality phenotypic data to assist plant breeders in developing disease and drought-resistant, high-yielding wheat varieties with a broad genetic base and maximizing the potential of new genotyping technologies. Additional but vital goals are to share knowledge and germplasm to accelerate new germplasm development and dissemination as well as develop capacities of breeders and plant scientists in precision field phenotyping. Some examples of these field phenotyping interventions being implemented include the development and application of precise phenotyping approaches, standardized protocols and novel tools for heat stress assessment in

Sudan, *Septoria tritici* blotch in durum wheat in Tunisia (Ben M'Barek et al., 2022), *Septoria tritici* blotch in durum wheat and wheat rusts in Ethiopia (Kidane et al., 2017; https://globalrust.org/content/sources-resistance-septoria-tritici-blotch-identified-ethiopian-durum-wheat), heat and drought tolerance in spring wheat in Morocco, yield potential in Egypt and Zimbabwe and drought and yield potential in Kenya (https://globalrust.org/content/global-network-precision-field-based-wheat-phenotyping).

Additionally, low-cost high-throughput phenotyping tools for field selection for disease, drought and crop variety performance are currently being developed. These tools will be used in breeding programs in Senegal, Ghana and Uganda and will serve as "centers of excellence for peanut breeding" in West and Eastern Africa (https://ftfpeanutlab.caes.uga.edu/Research/variety-development/high-throughput-phenotyping-in-senegal–ghana-and-uganda.html).

In West Africa, the field phenotyping network, since its inception in 2016 in the sub-region, has implemented high-throughput UAV (drone-based) phenotyping methodologies which are functional for sorghum, cowpea, pea nut and pearl millet (Gano et al., 2021; Audebert et al., 2022). The network is advancing breeding activities through 'fine phenotyping', varietal evaluations in diverse environments to identify hot spots for specific stresses, including farmers' fields to test promising breeding lines in participating countries such as Senegal, Ghana, Mali and Burkina Faso.

The establishment of the network has facilitated infrastructure development, equipment acquisition, data management paired with long-term training of dedicated students, technicians and breeders capable of doing both breeding and carrying out high-throughput phenotyping measurements. In the subregion, three sites have been chosen as prospective hubs for high throughput phenotyping. Each hub including Bambey (ISRA research station, Senegal), Sotouba (IER research station, Bamako, Mali) and Farako-ba (INERA research Station, Bobo Dioulasso, Burkina Faso) exemplifies the diversity of soil and climate conditions in the region. According to Audebert et al. (2022), the network setup in Senegal is the most advanced while Mali and Burkina Faso lag behind mainly due to limited phenotyping equipment and funding challenges.

Similarly, the Regional Study Centre for the Improvement of Drought Adaptation (CERAAS) in complementing the field phenotyping initiatives of the West African field phenotyping network, has developed robust UAV imagery-based data collection and spatial modelling methodologies to accurately measure key traits of cereal crops to advance plant breeding programs. UAVs equipped with a multispectral imaging system coupled with a fully automated image processing pipeline can indirectly measure agronomic and phenological characteristics of cereal crops in agricultural field trials (Mbaye et al., 2022).

Moreover, to advance the promotion and advancement of precision agriculture (PA) in Africa, the African Association for Precision Agriculture (AAPA), an initiative of the African Plant Nutrition Institute (APNI) is spearheading this goal (https://paafrica.org/AAPA). Since its establishment in 2020, the AAPA has worked in partnership with academia, research institutions, agri-food industry, financial institutions, and public and private sector organizations to develop and scale up PA strategies and innovations through sustainable integration into African agriculture to address food security (i.e., reduce yield gaps) climate change, and land degradation challenges.

5.2 Field phenotyping research in African countries

5.2.1 The case in Ghana

Digitalization of Agriculture is a new trend facilitated by digital platforms aimed at transforming small scale agriculture by providing agricultural services to smallholder farmers in Ghana (Atanga, 2020; Abdulai et al., 2023). These digital platforms include simple devices such as mobile phones or radio to a more sophisticated devices (e.g., field sensors, GIS, drones, field sensors, machinery sensors and diagnostics precision systems).

In Ghana some of the notable digital platforms transforming the small-scale farming sector include the TROTRO Tractor Limited (an agritech company) that combines mechanization with IoT and technology to make agricultural machinery (i.e., tractors and combined harvesters) available, accessible, and affordable to farmers thereby enhancing their efficiency and productivity (https://www.trotrotractor.com). The use of remote sensing as a decision support system (DSS) tool to optimize irrigation and farm management towards increasing yields has also been demonstrated (Kpienbaareh et al., 2019). These innovations primarily address the numerous issues smallholders and rural farmers confront in the present food systems, such as climate change, low access to inputs and restricted access to information (Degila et al., 2023).

As in many African countries breeding and field phenotyping is mostly based on conventional manual methods. However, to evaluate crop performance and improve breeding competitiveness, modern technologies using high-throughput techniques are being implemented but at a slow pace (e.g., Hall et al., 2018; Kassim et al., 2022; Sie et al., 2022). For instance, the responses of two populations of groundnut genotypes with various maturities to early and late leaf spot diseases were assessed under field conditions using UAV imagery (Kassim et al., 2022). In another breeding program, a smartphone-based RGB images detected leaf spot resistance and predicted yield in groundnut (Sie et al., 2022). In a resource constraint economy, Ghana is faced with numerous challenges such as lack of research funding, phenotyping infrastructures and technical personnel among others that can advance rapid characterization of agriculturally relevant traits (e.g., growth, yield, stress resistance). To increase its phenotyping capabilities will require a concerted effort from all stakeholders across the crop production value chain.

5.2.2 The case in Senegal

Senegal is making strides in precision agriculture by employing digital tools to address crop production challenges (https://www.apni.net/wp-content/uploads/2020/02/WAFPA-Tine.pdf). Even though advancement in modern breeding and field phenotyping methodologies has been slower and predominantly based on conventional methods (e.g., Dingkuhn et al., 2015), the use of drones for agricultural monitoring (i.e., stress detection, disease

surveillance, crop performance) aided by high-throughput phenotyping has been exploited thanks to initiatives by the CERAAS and West African field phenotyping network. For instance, UAV multi-spectral imaging has been employed for the estimation of shoot biomass, leaf area index (LAI) and plant height of West African sorghum varieties under severe drought conditions (Gano et al., 2021). The drone-based field phenotyping approach developed could help identify essential traits and cultivars for drought tolerance in sorghum breeding. The main challenges confronting crop field phenotyping in Senegal are lack of equipment, technical personnel and funding (Audebert et al., 2022). However, Senegal being a hub for field phenotyping in West Africa, has the potential to increase its field phenotyping capabilities in the future.

5.2.3 The case in Nigeria

According to a recent review by Izuogu et al. (2023), the digitalization of agriculture in Nigeria has reduced middlemen's participation in agriculture, offered small-holder farmers opportunities to improve their productivity and markets, and strengthened the connections between extension and research facilities. The authors demonstrated that for effective digitalization of agriculture, training was required in the areas of skills development, use of demand-driven digital services, digital privacy, and security issues. The challenges of digitalization of agriculture identified were lack of technical expertise, inadequate infrastructure, and high purchase and maintenance costs. The use of remote sensing techniques for precision crop production and monitoring has been implemented but to a lesser extent. Ifeanyieze et al. (2014) have previously reviewed the remote sensing techniques needed for the smooth implementation of precision crop management by farmers as a climate change adaptation strategy in Nigeria. Few research groups have utilized remote sensing techniques for field phenotyping. For instance, Ejikeme et al. (2017) used a satellite-based crop prediction model to estimate crop statistics of major crops including rice, cassava, yam, and maize. Recently, the Institute of Tropical Agriculture (IITA) through its collaborative soybean breeding programs has implemented machine learning (ML) models and multispectral high-resolution UAV imagery to aid rapid high-throughput phenotypic workflow for soybean yield estimation (Alabi et al., 2022). Other breeding programs used manual field evaluation coupled with digital imaging analysis for phenotyping tomato breeding population (Daniel et al., 2016).

The use of a handheld optical NDVI sensor for the evaluation of shoot biomass in field-grown staking yam has been implemented (Iseki and Matsumoto, 2019). Altogether, Nigeria has great potential for improving its field phenotyping capabilities.

5.2.4 The case in Morocco

Morocco is among the few African countries well-positioned for widespread agricultural digitalization for precision agriculture and field phenotyping to increase crop production and cope with adverse environmental conditions such as drought. Jabir and Falih (2020), recently reviewed the state of digital agriculture in Morocco and highlighted the opportunities and challenges that need to be addressed. The design and implementation of a wireless sensor network (WSN) and decision support tools (i.e., drones) for monitoring the agricultural environment have been demonstrated (Jabir and Falih, 2020). Nevertheless, challenges such as sensor deployment and inadequate software analytics still exist (Kobo et al., 2017). Morocco is home to the International Centre for Agricultural Research in the Dry Areas (ICARDA's) phenotyping facilities (ICARDA phenotyping platforms in Morocco), including a precision phenotyping platform at Sidi el Aidi (Settat) (Figure 3) and a phenomobile system (PhenoBuggy) situated at the main research station in Marchouch (Rabat) designed for drought and heat stress tolerance studies (https://www.cgiar.org/news-events/news/icardas-phenotyping-facilities-a-game-changing-solution-for-abiotic-stress-tolerance-in-crops/). The PhenoMA is another high-throughput phenotyping platform currently installed in Benguerir (Quahir et al., 2022). Field phenotyping using various remote sensing techniques has been deployed for drought monitoring (Bijaber et al., 2018; Bouras et al., 2020; Laachrate et al., 2020), and grain yield prediction (Belmahi et al., 2023).



FIGURE 3
The ICARDA's precision field phenotyping platforms installed at Sidi el Aidi (Settat) in Morocco. Images are in courtesy of Andrea Visioni of ICARDA-Morocco.

5.2.5 The case in Egypt

Digital agriculture appears promising in addressing the major challenges facing the agri-food sector in Egypt and across the Middle East and North Africa (MENA) countries (Bahn et al., 2021). Available evidence indicates that the adoption of digital and precision agriculture technologies is still in its infancy and is typically driven by high-value agricultural production (Elsafty and Atallah, 2022; Sayed et al., 2023). However, Egypt has made strides in the utilization of modern technologies for agricultural crop management employing big data in tandem with cloud support systems, IoT, UAVs, satellite imagery, AI, machine learning, and remote sensing (Shokr, 2020; Abdelnabby and Khalil, 2023; Sayed et al., 2023). Typical high-throughput field phenotyping methodologies has been implemented in various crops for quantifying wheat characteristics in the Nile Delta (Elmetwalli et al., 2022) and estimating the growth performance and yield of soybean exposed to different drip irrigation regimes under arid conditions (Elmetwalli et al., 2020). Additionally, remote sensing techniques based on thermal imaging and passive reflectance have been used to estimate the crop water status and grain yield in wheat (El-Shirbeny et al., 2014; Elsayed et al., 2017).

5.2.6 The case in South Africa

The agricultural sector in South Africa has been developing and moving towards becoming a knowledge-intensive enterprise due to new innovations and technologies incorporated in the digital economy (Baumüller and Kah, 2019; Born et al., 2021; Smidt and Jokonya, 2022). Due to this transformation, conventional production methods have gradually been replaced with more advanced, efficient and innovative systems (e.g., remote sensing) for crop breeding and phenotyping (Mutanga et al., 2016).

Field phenotyping using modern high-throughput infrastructures and precision agriculture techniques is better developed in South Africa compared to other countries on the continent (Nyaga et al., 2021; Mukhawana et al., 2023). Some research groups are making efforts in championing field phenotyping and precision agriculture through workshops and implementation of UAV remote sensing applications and other approaches for agricultural monitoring (stress detection, nutrient and irrigation management) (https://www.fabinet.up.ac.za/index.php/research-groups/remote-sensing). For example, the Forestry and Agricultural Biotechnology Institute (FABI) and the Agricultural Research Council (ARC) (https://www.arc.agric.za/Pages/Home.aspx) are committed to building phenotyping infrastructures and disseminating emerging technologies for agricultural development.

Various remote sensing applications have been employed targeted at different scales of crop monitoring (e.g., crop water use efficiency) in precision agriculture (e.g., Munghemezulu et al., 2023; Wellington, 2023). For instance, foliar temperature and stomatal conductance have been used as indicators of water stress in maize based on optical and thermal imagery acquired using a UAV platform (Brewer et al., 2022a). The utility of multispectral UAV imagery as proxy for predicting chlorophyll content of maize at various growth stages in smallholder farming systems has been reported (Brewer et al., 2022b). The physiological processes of the maize canopy are intimately tied to and influenced by LAI, which is closely related to its productivity (Peng et al., 2021). Another study has focused on estimating the LAI of maize in smallholder farms across the growing season using UAV-derived multi-spectral data (Buthelezi et al., 2023). Maize is a major crop in South Africa, therefore, significant research on the crop using high-throughput techniques will aid in developing improved cultivars for farmers. South Africa has a great potential for becoming the field phenotyping hub of Africa due to the massive investment in modern technologies.

5.2.7 The case in Zimbabwe

In Zimbabwe, the implementation of digitalized agriculture is low and tilted toward commercial farmers than smallholder community farmers (Parwada and Marufu, 2023). Specifically, highly literate, and resource-rich farming communities tend to use digitalized agriculture more frequently than farmers with lesser resources. At the communal level, farmers use mobile phones to obtain farming information relating to crop management, climate, and weather information (Musungwini, 2018; Zimbabwe Centre For High Performance Computing, 2021). The application of modern digital agriculture tools and infrastructure (i.e., sensors, robotics, AI, UAVs, and other advanced machinery is common in a few well-resourced commercial farms notably, those managed by large multinational companies (Shonhe and Scoones, 2022). Parwada and Marufu (2023) recently reviewed the challenges and opportunities for digitalization of the Zimbabwean agriculture. Key challenges such as lack of high-throughput infrastructures, digital illiteracy, and strict regulations for drone deployment among others have been highlighted for limiting digital agriculture applications. However, according to the authors, Zimbabwe have the potential for improving its digital agriculture for crop management, yield prediction, disease detection, climate forecasting, and soil management through precision agriculture. In recent years, few high-throughput phenotyping has been implemented in Zimbabwe using RGB picture vegetation indexes (Kefauver et al., 2015), and multi-spectral imaging for field phenotyping of maize (Zaman-Allah et al., 2015). Other studies include remote sensing methodologies for crop monitoring under conservation agriculture (Gracia-Romero et al., 2018; Gracia-Romero et al., 2020), affordable UAV-based RGB phenotyping techniques for evaluating maize performance under low nitrogen conditions (Buchaillot et al., 2019), and accelerating crop improvement in response to changing climate conditions employing UAV-based multispectral phenotyping for disease resistance in maize (Chivasa et al., 2020). Zimbabwe is among the few African countries capable of advancing its field phenotyping capabilities in the future.

5.2.8 The case in Kenya

Although there are several technologies currently available to Kenya's agricultural sector they have not yet become widely used

(Osiemo et al., 2021). Large-scale adoption of digital solutions is hampered by a lack of digital literacy and infrastructure. Only a few research groups are skilled in using and maintaining back-end service operations like data management, blockchain, machine learning, IoT, GIS, and drones (Osiemo et al., 2021). However, the application of GIS and remote sensing techniques have been used to map frost hotspots for mitigating agricultural losses (Kotikot and Onywere, 2015), climate-smart crop management (Manzi and Gweyi-Onyango, 2021), and assessment of yield variations and its determinants in smallholder systems (Burke and Lobell, 2017). Similarly, high-throughput phenotyping platforms based on multi-spectral imaging and RGB vegetation indices have been implemented for field phenotyping of maize (Kefauver et al., 2015; Zaman-Allah et al., 2015). Kenya has the potential of expanding its phenotyping capacities through low-cost precision agriculture and breeding.

5.2.9 The case in Ethiopia

Digital agricultural innovations in precision agriculture have the potential to increase productivity while minimizing harmful environmental impacts along the value chains of agriculture and the food systems in Ethiopia (Alemaw and Agegnehu, 2019; Tamene and Ashenafi, 2022). In recent years, there have been some improvement in digital infrastructure in Ethiopia (Abdulai, 2022). However, the majority of Ethiopia's smallholder farmers have limited access to digital farming technologies (Tamene and Ashenafi, 2022). According to Tamene and Ashenafi (2022), several challenges such as inadequate technological capacity, limited funding to develop and disseminate digital tools and lack of data sharing channels hampers the development of digital agriculture in Ethiopia. These barriers restrict the deployment of modern technologies for crop breeding and field phenotyping. Field phenotyping has relied largely on conventional methods as in the studies of eco-geographic adaptation and phenotypic diversity of Ethiopian teff across its cultivation range (Woldeyohannes et al., 2020) and genetic diversity in Ethiopian Durum Wheat (Mengistu et al., 2018). Field phenotyping using high-throughput techniques has been introduced in recent times. Remote sensing and GIS based methods has been used as crop yield predictors in wheat and maize (Beyene et al., 2022; Debalke and Abebe, 2022) as well as physical land suitability analysis for major cereal crops (Debesa et al., 2020). In essence, Ethiopia has the potential to accelerate its phenotyping capabilities. Table 3 summarizes some key field phenotyping activities that exist in the African countries discussed in this review.

5.3 Current developing field phenotyping platforms in Africa

UAVs have been selected as the technical solution that is most suited for deployment across sites and trials throughout the several initiatives that made it possible for the West African field phenotyping network to get started (Audebert et al., 2022). For instance, in Senegal, the UAV platform comprises a FeHexaCopterV2 hexaCopter UAV system (Flying Eye Ltd., Sophia Antipolis, France) fitted with three cameras mounted on a two-axis gimbal pointing vertically downward. The camera consists of an RGB ILCE-6000 digital camera (Sony Corporation, New York, NY, USA), AIRPHEN multispectral camera (Hiphen, Avignon, France), and infrared thermographic camera Tau 2 (Flir system, Oregon, USA) that collects spectral imagery of crops such as sorghum, pearl millet and peanut and cowpea (Gano et al., 2021).

The Agricultural Research Council (ARC) of South Africa has installed a Phenospex planteye multispectral 3D laser scanner (i.e., the first of its kind in Africa) in the field (https://phenospex.com/products/plant-phenotyping/fieldscan-high-throughput-field-phenotyping/fieldscan-3d-spectral-plant-measurements-in-the-field-south-africa/). This state-of-the-art facility is fully automated, carrying a high-resolution sensor that combines the strength of 3D vision with the power of multispectral imaging. It captures plant data non-destructively and delivers precise and accurate plant parameters in real-time. Plant phenotypic features such as digital biomass, plant height, 3D leaf area, projected leaf area, leaf area index, leaf inclination, etc., can be measured. The spectral information allows for the quantification of plant health, disease, senescence, N-content, chlorophyll levels, etc. Therefore, this phenotyping facility could assist in the characterization and development of varieties with improved biotic and abiotic stress resistance for key crops such as grapefruit, sunflower, green maize and other cereals in Southern Africa.

Recently, a unique close-to-field high-throughput plant phenotyping platform "PhenoMA" has been installed in Benguerir, in the arid region of Morocco by the Mohammed VI Polytechnic University. PhenoMA consists of a 1440 fully automated lysimetric mini-plots system that can track the dynamics of water use and simulate drought scenarios. A critical component is a fully autonomous phenotyping robot (Hiphen PhenoMobile) that enables plant measurements at the canopy scale, using a range of sensors including RGB, multispectral, infrared (IR), and LiDAR cameras to monitor canopy development (Quahir et al., 2022).

Overall, due to the rich agricultural biodiversity of Africa, phenotyping in Africa has great potential to contribute to the development of improved crop varieties and enhanced food security. The utilization of high-throughput tools can boost the elucidation of new agriculturally proven traits and catalogue these phenotypes in their natural environment.

6 Challenges limiting the application of high-throughput field phenotyping in Africa and the way forward

The application of emerging field phenotyping technologies has the potential to accelerate plant breeding efforts and crop production in Africa. On the other hand, most of these approaches reviewed here are at best relatively new or unknown to some of the plant science community in Africa. Field phenotyping is a critical component of crop improvement but remains a major bottleneck in African agriculture, as is the case globally. Some of the key challenges limiting the application of high-throughput field phenotyping in Africa are highlighted below.

6.1 Lack of appropriate high-throughput field phenotyping approaches

Phenotypic analysis has become a major limiting factor in genetic and physiological analyses in plant sciences as well as in plant breeding in Africa. The inadequate phenotyping infrastructures and software analytical tools that can be used by agricultural practitioners to make sense of simple to complicated phenotypic datasets have contributed to the low implementation of high-throughput phenotyping. The operational complexity to support both data acquisition and analysis has limited the use of these platforms for research activities worldwide (Chapman et al., 2014), including developing continents like Africa. To this end, training in image analytics, software, and computer vision to provide a new generation of skilled personnel must be implemented by African governments, universities, and the private sector. Phenotyping advancement is critical for current breeding progress for crop improvement in Africa. While the development of efficient high-throughput field phenotyping remains a challenge for future breeding progress, the growing interest in low-cost solutions for remote-sensing approaches, machine vision, as well as data management, may facilitate technological adoption.

6.2 Cost of phenotyping infrastructures and maintenance

As a developing continent comprising highly indebted poor countries (HIPC) (Henri, 2019) and faced with multi-faceted economic hardships, the major limitation to the adoption and implementation of high-tech field phenotyping in Africa is the perceived high entry costs associated with the longer-term footprint of prototypical platforms (Reynolds et al., 2019). In several African countries, especially those discussed in this review, basic phenotyping tools and infrastructure even for the simplest field measurements and experimentation are scarce.

This prevents many research organizations in Africa such as IITA, CIAT, and Africa Rice, from implementing demand-driven approaches due to a lack of investment budget or avoiding the significant follow-up costs on maintenance of large phenotyping infrastructures. For instance, the use of ground vehicles, aerial vehicles and gantries may require huge investment costs (Pauli et al., 2016; Vergara-Díaz et al., 2016).

Therefore, the requirements for such specialized equipment may be a bottleneck for widespread use in breeding programs in poor countries. To alleviate this challenge, low-cost concepts and methods of high-throughput phenotyping platforms (HTPPs) (e.g., sensors and platforms) that rely on easy-to-use technology must be disseminated in Africa by identifying demands, and relevance, and adopting the required approach given the current financial constraints. For instance, conventional digital cameras (i.e., digital photography) could provide a more convenient method since they are more affordable, portable, and easy to use (Casadesús and Villegas, 2014).

6.3 Limited investment and funding

Limited investments in science, technology, and innovation (STI) on the part of African governments, research institutions (e.g., academia) and the private sector have partly contributed to the poor implementation of high-throughput field phenotyping. The budgetary allocations dedicated to research, development, and innovation are small. For example, in Ghana, a minimum of 1% of gross domestic product (GDP) is applied towards research and development (https://mesti.gov.gh/government-increase-research-funding/). Similarly, in Morocco, the percentage of GDP to research as of 2010 was 0.63% (Hamidi and Benabdeljalil, 2013). This research funding gap is pervasive across the African continent.

Whereas research institutions and universities in developed economies, such as Europe (see https://eppn2020.plant-phenotyping.eu/EPPN2020_installations#/), Australia, North America and Asia, have in recent years invested heavily in large-scale research infrastructure for automated and high-throughput field phenotyping, the same cannot be said for Africa. These large investments for plant phenotyping include funding, research hours and high-throughput installations (Costa et al., 2019; Rosenqvist et al., 2019; https://eppn2020.plant-phenotyping.eu/).

Furthermore, crops grown in Africa are frequently too local to attract international research funding for field phenotyping. Only a few essential African crop commodities, such as cassava and sweet potatoes, are funded solely by extrabudgetary sources. Most of the main staple crops are exclusively funded for phenotyping exploitation outside of Africa.

In addition to the above considerations, African governments and the Science Granting Councils Initiative (SGCI) in sub-Saharan African countries mandated to support the Science Granting Councils (SGCs), must dedicate enough funding for low-cost plant phenotyping research infrastructure in the sub-region in the short to medium term. This could be achieved by developing financing mechanisms and collaborating with private sector partners. Donor support to Africa for agriculture and food security research should also consider projects in modern plant phenotyping and digital agriculture.

6.4 Lack of Skilled technical personnel

A serious deficit of skilled technical personnel in the plant sciences and phenotyping ecosystem is evident in African countries. The building up of such competencies and the development of human resource capacity is necessary to operate simple-to-sophisticated equipment to accelerate breeding efforts through high-throughput phenotyping techniques. Another major barrier is the loss of talented and skilled personnel who were trained in developed nations and have contributed to the brain drain due to inadequate job prospects in Africa. Mostly, funds to pay salaries and absorb project operating costs

are either limited or insufficient, resulting in a reduction of skilled personnel. Furthermore, due to the inadequacies in research and infrastructure in many African nations, training acquired overseas is sometimes unsuited to local demands. To address this constraint, digital agricultural competencies and sensor technologies should be integrated into undergraduate and postgraduate learning curricula to allow students to specialize in digital agriculture through their projects. This will create a plethora of career opportunities for competent skilled personnel who can adapt to the emerging technologies for field phenotyping.

6.5 Regulations controlling emerging technologies

Emerging technologies such as UAVs offer the advantages of being flexible, real-time and non-destructive for agricultural phenotyping, but they must adhere to strict operational standards to ensure their safe use. Strict airspace regulations in many jurisdictions around the world and particularly in African countries due to impact of political instability and military governments on UAV deployment may prohibit their use or make them unfeasible in practice (Gago et al., 2015; Yang et al., 2017a; Ayamga et al., 2021). For instance, authorization from regulatory authorities, such as the air force, civil aviation and police, are required to undertake UAV flight campaigns, which mostly take time to be approved causing issues in time-critical data collection applications. According to Ayamga et al. (2021), in Africa, countries with regulations include Ghana, South Africa, Zimbabwe, Nigeria, Cameroon, Benin, Gabon, Senegal, Botswana, Namibia, Malawi, Tanzania, Zambia, Madagascar, Rwanda and Kenya. However, the lack of proper regulation and enforcement continues to limit the widespread adoption of drones. Unfortunately, these regulations combine to mean that most high-throughput techniques can only be implemented by multinational research institutions, even in those organizations, deployment of systems is limited to a few high-priority projects. Commitment of African governments and relevant stakeholders is crucial in the implementation and enforcement of regulations. The widespread deployment of drones stands to benefit farmers hence concerted effort need to be made to sustain its adoption by promoting public digital literacy on the technology, skill development for potential users and farmers on drone operation and developing the necessary policy framework with regulatory agencies to increase the safety and acceptability of using agricultural drones in Africa.

6.6 Weakness of phenotyping linkages

At the regional and continental levels, networking is a powerful tool for increasing scientific collaboration and fostering information sharing. There seems to be weak collaborations between the African plant science community and international phenotyping partners which hampers technological transfer and adoption. As high-throughput field phenotyping initiatives have started in Africa, there is a need to strengthen national and institutional efforts within the continent for the development and application of accurate and high-throughput field phenotyping capabilities. The West Africa field phenotyping network should be strengthened and better resourced to carry out their mandate. Similar initiatives such as the EMPHASIS (https://emphasis.plant-phenotyping.eu) should be experimented to provide a more practical use of the available phenotyping data.

The IPPN should spread its operations to Africa to develop programs and establish synergies geared towards face-lifting plant phenotyping projects in the continent. Again, African governments and their partners should invest in building a center of excellence or shared facilities for African plant scientists. Finally, a more urgent challenge is, however, that the international phenotyping community needs to bridge the gap between advanced economies and developing regions of the world such as Africa to benefit from the huge research efforts made internationally.

7 Concluding remarks and future perspectives

This review provides an overview of high-throughput field phenotyping and its implications for African crops. It highlights the prospects of emerging high-throughput phenotyping techniques and integrated sensor platforms for plant trait assessment for field phenotyping that could apply to African crops. High-throughput field phenotyping has superior advantages that facilitate quick, non-destructive, and high-throughput detection, thus overcoming the shortcomings of conventional approaches. The readiness and the potential adoption of high-throughput field phenotyping for practical implementation in Africa are of paramount interest and should be demonstrated.

Field phenotyping solutions of immediate to long-term feasibility for African crops will likely rely on a combination of available techniques or prototypes of low-cost sensors and imaging approaches to study crop performance. Manual methods dominate the field phenotyping ecosystem with only a few countries beginning to explore high-throughput techniques through digital and precision agriculture. Notably, high-throughput phenotyping cannot yet completely replace manual measurements but should be promoted. The implementation of high-throughput phenotyping in general, and low-cost methods for field evaluation, is still fraught with challenges in Africa. Challenges identified by this present review include the high upfront cost of the prototypical platforms, huge funding gap, lack of conceptual and technical capacity, lack of technology transfer infrastructure and methodological approaches, lack of phenotyping network on the continent and the needed legislation in some cases, amongst others.

Lack of financial resources, a problem pervasive in African countries needs to be tackled holistically. Public-private partnerships could support resolving these financial and investment challenges to foster political will. Although in some countries, this public-private drive is already being implemented through close collaboration between universities and agricultural research organizations, these

efforts need to be stepped up. In parallel, African governments should dedicate enough funding, incentives, and tools to breeders to advance research and innovations regarding high-end plant breeding. We suggest that donor support to Africa for agriculture and food security research should also consider projects in modern plant phenotyping to cope with current and projected climate change.

This will open the possibility of investing more in current sensor and imaging technologies for field data collection and the use of cost-effective phenotyping technologies that are already available to increase the throughput, quantity and quality of phenotypic data. The wide range of applications for these phenotyping technologies makes them good candidates and feasible choices for adoption in Africa which hitherto were prohibitive in terms of cost and deployment. The advantages of improved sensor-platform integration have facilitated the development of complete phenotyping systems that can gather, integrate and store data for many subsystems concurrently in a structured, efficient and cost-effective way. Such platforms have been widely adopted by research groups in developed countries and are gradually adopted by plant breeders in Africa as the technology develops and the benefits are proven.

In addition to the adoption of high throughput field phenotyping approaches in African countries, precision agriculture will also greatly benefit and revitalize the establishment of closer interaction between breeders and farmers to develop protocols mutually for the optimal use of improved crop varieties. The tools and knowledge exchange are expected to spur a second green revolution to meet the agricultural challenges to feed the ever-increasing African population. In terms of advancing field crop phenotyping in Africa for agricultural crop sustainability, we propose that breeding priority should be given to the most important staple crops such as maize, wheat, yam, cassava, cowpea, sorghum, etc. These crops form the backbone for food security and hence their improvement is crucial in the wake of prevailing climate change and production constraints. We suggest that each country selects traits that are of high demand and relevance by farmers and consumers when designing breeding strategies. In parallel, high-throughput phenotyping should be incorporated into national agricultural research policies and prioritize the practical implementation of field phenotyping. By and large, these could be achieved when governmental and private sector participation, as well as financial support, is readily available.

To overcome the challenges with the deployment of phenotyping tools and the integration of software to deliver accurate data acquisition, processing, analysis and management, a multidisciplinary team of expert-level skills and competencies may be required. This will necessitate deliberate training and capacity improvement of African plant scientists and students in software engineering and computer science domains, including AI, demanding true interdisciplinary partnerships to provide meaningful results and inform decision-making, while addressing the issue of training cost and related risks. In this instance, we recommend technological adoption rather than complete technological development considering the financial constraints and the low-level expertise in software and equipment development. However, as the plant phenotyping industry develops the development of new technologies from scratch may be feasible in Africa.

Furthermore, we propose encouraging collaborations between the African plant science community with their international counterparts to foster collaborative research, effective technological transfer and adoption. This review recommends close collaboration with the IPPN and similar phenotyping networks to benefit from the unprecedented investments made in field phenotyping infrastructures globally. Consequently, crop scientists may leverage ground-breaking advancements in high-throughput field data collection, image analysis and data management. Efforts should be made to foster synergies among different African countries by establishing transnational interdisciplinary networks that incorporate expertise in all aspects of plant breeding.

To address the limited investments in science, technology and innovation (STI), a commitment for expanded and long-term funding of agricultural research and development is essential. At the policy and operational levels, barriers must be overcome to allow the smooth establishment of public-private partnerships for transformational change in research and demand-driven technologies for breeders and farmers. There is renewed interest both from private and public institutions in developed countries to support African agriculture. Hence, African agricultural institutions need to develop strategies and synergies that include building partnerships that must be implemented to tackle the challenges, especially in the face of climate change and food insecurity.

The widespread adoption of high-throughput field phenotyping techniques in African countries could only be made possible in plant breeding programs if it can be proven as something worthwhile in terms of genetic gains attained with resources invested. Hence, costs must be reasoned in relation to the precision, repeatability, heritability, cost per unit plot or trait, prevailing climatic and economic condition, etc., required in a particular phenotyping activity. Given what has been said, to ensure that such implementation of field phenotyping can be translated into yield gains, low-cost phenotyping tools must be adopted. On this basis, affordable, easy-to-handle, reliable tools, and phenotyping infrastructures for small to large-scale field phenotyping may become a strategic choice and pave the way for practical implementation. Such technologies applicable to phenotyping methodologies should be available soon due to the high demands and efforts by the phenotyping community in Africa.

Author contributions

DC, NV, TW, FM, and MH provided the conceptualization of the manuscript. DC drafted, wrote, and edited the manuscript. NV, MC, AR, MM, TW, FM, and MH substantially reviewed the manuscript. TW, FM, and MH provided supervision. All authors read and approved the final manuscript. All authors contributed to the article.

Funding

This research is funded by OCP S.A under the Mohammed VI Polytechnic University (UM6P), Rothamsted Research, and Cranfield University program. Rothamsted Research receives grant-aided support from the Biotechnology and Biological

Sciences Research Council (BBSRC) through the Designing Future Wheat program [BB/P016855/1].

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

Aasen, H., Burkart, A., Bolten, A., and Bareth, G. (2015). Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: from camera calibration to quality assurance. *ISPRS J. Photogrammetry Remote Sens.* 108, 245–259. doi: 10.1016/j.isprsjprs.2015.08.002

Abay, K. A., Abate, G. T., Barrett, C. B., and Bernard, T. (2019). Correlated non-classical measurement errors, 'Second best' policy inference, and the inverse size-productivity relationship in agriculture. *J. Dev. Economics* 139, 171–184. doi: 10.1016/j.jdeveco.2019.03.008

Abdelnabby, M., and Khalil, T. (2023). Assessing precision agriculture applicability in agriculture sector in Egypt. *Fayoum J. Agric. Res. Dev.* 37 (1), 54–62. doi: 10.21608/fjard.2023.281055

Abdulai, A. R. (2022). Toward digitalization futures in smallholder farming systems in Sub-Sahara Africa: A social practice proposal. *Front. Sustain. Food Syst.* 6. doi: 10.3389/fsufs.2022.866331

Abdulai, A. R., Kc, K. B., and Fraser, E. (2023). What factors influence the likelihood of rural farmer participation in digital agricultural services? experience from smallholder digitalization in Northern Ghana. *Outlook Agric.* 52 (1), 57–66. doi: 10.1177/00307270221144641

Adhikari, U., Nejadhashemi, A. P., and Woznicki, S. A. (2015). Climate change and eastern Africa: a review of impact on major crops. *Food Energy Secur.* 4 (2), 110–132. doi: 10.1002/fes3.61

Affoh, R., Zheng, H., Dangui, K., and Dissani, B. M. (2022). The impact of climate variability and change on food security in sub-saharan Africa: Perspective from panel data analysis. *Sustainability* 14 (2), 759. doi: 10.3390/su14020759

AGRA. (2019). *A handbook of plant breeding: Case studies from Africa 2019*. Available at: https://agra.org/wp-content/uploads/2022/09/Handbook-of-Plant-Breeding-Case-Studies-from-Africa-2019.pdf (Accessed July 15, 2023).

Ahmed, H. G. M. D., Zeng, Y., Fiaz, S., and Rashid, A. R. (2023). "Applications of high-throughput phenotypic phenomics," in *Sustainable agriculture in the era of the OMICs revolution*. Eds. C. S. Prakash, S. Fiaz, M. A. Nadeem, F. S. Baloch and A. Qayyum (Cham: Springer). doi: 10.1007/978-3-031-15568-0_6

Alabi, T. R., Abebe, A. T., Chigeza, G., and Fowobaje, K. R. (2022). Estimation of soybean grain yield from multispectral high-resolution UAV data with machine learning models in West Africa. *Remote Sens. Applications: Soc. Environ.* 27, 100782. doi: 10.1016/j.rsase.2022.100782

Alemaw, G., and Agegnehu, G. (2019). Precision agriculture and the need to introduce in Ethiopia. *Ethiopian J. Agric. Sci.* 29 (3), 139–158.

Andrade-Sanchez, P., Gore, M. A., Heun, J. T., Thorp, K. R., Carmo-Silva, A. E., French, A. N., et al. (2014). Development and evaluation of a field-based high-throughput phenotyping platform. *Funct. Plant Biol.* 41 (1), 68–79. doi: 10.1071/FP13126

Andrianto, H., Suhardi, S., and Faizal, A. (2017). "Measurement of chlorophyll content to determine nutrition deficiency in plants: A systematic literature review," in *2017 International Conference on Information Technology Systems and Innovation, ICITSI 2017 - Proceedings*, 2018-January. (Piscataway, New Jersey, United States: Institute of Electrical and Electronics Engineers (IEEE)), pp. 392–397. doi: 10.1109/ICITSI.2017.8267976

Appels, R., Eversole, K., Feuillet, C., Keller, B., Rogers, J., Stein, N., et al. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* 361 (6403). doi: 10.1126/science.aar7191

Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends Plant Sci.* 19 (1), 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 (2), 592–618. doi: 10.1111/jipb.13191

Araus, J. L., Slafer, G. A., Royo, C., and Serret, M. D. (2008). Breeding for yield potential and stress adaptation in cereals. *Crit. Rev. Plant Sci.* 27 (6), 377–412. doi: 10.1080/07352680802467736

Asare-Bediako, E., Taah, K. J., van der Puije, G., Amenorpe, G., Appiah-Kubi, A., and Akuamoa-Boateng, S. (2019). Evaluation of maize (Zea mays L.) genotypes for high grain yield and resistance to maize streak virus infections under diverse agro-ecological zones. *Res. J. Plant Pathol.* 2 (2), 11. doi: 10.36648/plantpathology.2.2.11

Atanga, S. N. (2020). Digitalization of agriculture: how digital technology is transforming small-scale farming in Ghana. *Agrarian, Food and Environmental Studies (AFES). A research paper submitted to Erasmus University in partial fulfilment of the requirements for the degree of Master of Arts in Development Studies at the International Institute of Social Studies.*

Audebert, A., Luquet, D., Vadez, V., Fonceka, D., and Kane, N. A. (2022). *Toward a regional field phenotyping network in West Africa*. Available at: https://agritrop.cirad.fr/604058/2/ID604058.pdf.

Ayamga, M., Tekinerdogan, B., and Kassahun, A. (2021). Exploring the challenges posed by regulations for the use of drones in agriculture in the African context. *Land* 10 (2), 164. doi: 10.3390/land10020164

Badu-Apraku, B., Fakorede, M. A., Nelimor, C., Osuman, A. S., Bonkoungou, T. O., Muhyideen, O., et al. (2023). Recent advances in breeding maize for drought, heat and combined heat and drought stress tolerance in sub-saharan Africa. *CABI Rev*. Volume 2023. doi: 10.1079/cabireviews.2023.0011

Bahn, R. A., Yehya, A. A. K., and Zurayk, R. (2021). Digitalization for sustainable agri-food systems: Potential, status, and risks for the MENA Region. *Sustainability* 13, 3223. doi: 10.3390/su13063223

Bai, G., Ge, Y., Scoby, D., Leavitt, B., Stoerger, V., Kirchgessner, N., et al. (2019). NU-Spidercam: a large-scale, cable-driven, integrated sensing and robotic system for advanced phenotyping, remote sensing, and agronomic research. *Comput. Electron. Agric.* 160, 71–81. doi: 10.1016/j.compag.2019.03.009

Bai, G., Ge, Y., Hussain, W., Baenziger, P. S., and Graef, G. (2016). A multi-sensor system for high throughput field phenotyping in soybean and wheat breeding. *Computers and Electronics in Agriculture* 128, 181–192. doi: 10.1016/j.compag.2016.08.021

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. *Agric. (Switzerland)* 12 (8), 1141. doi: 10.3390/agriculture12081141

Baumüller, H., and Kah, M. M. (2019). Going digital: Harnessing the power of emerging technologies for the transformation of Southern African agriculture. *Transforming Agriculture in Southern Africa: Constraints, Technologies, Policies and Policies and Processes* (Routledge: Taylor and Francis Inc., Milton Park, Abingdon in the United Kingdom), 179–187.

Belmahi, M., Hanchane, M., Krakauer, N. Y., Kessabi, R., Bouayad, H., Mahjoub, A., et al. (2023). Analysis of relationship between grain yield and NDVI from MODIS in the fez-meknes region, Morocco. *Remote Sens.* 15 (11), 2707. doi: 10.3390/rs15112707

Ben M'Barek, S., Laribi, M., Kouki, H., Castillo, D., Araar, C., Nefzaoui, M., et al. (2022). Phenotyping mediterranean durum wheat landraces for resistance to zymoseptoria tritici in Tunisia. *Genes* 13 (2), 355. doi: 10.3390/genes13020355

Beyene, A. N., Zeng, H., Wu, B., Zhu, L., Gebremicael, T. G., Zhang, M., et al. (2022). Coupling remote sensing and crop growth model to estimate national wheat yield in Ethiopia. *Big Earth Data* 6 (1), 18–35. doi: 10.1080/20964471.2020.1837529

Bian, J., Zhang, Z., Chen, J., Chen, H., Cui, C., Li, X., et al. (2019). Simplified evaluation of cotton water stress using high resolution unmanned aerial vehicle thermal imagery. *Remote Sens.* 11 (3), 267. doi: 10.3390/rs11030267

Bijaber, N., El Hadani, D., Saidi, M., Svoboda, M. D., Wardlow, B. D., Hain, C. R., et al. (2018). Developing a remotely sensed drought monitoring indicator for Morocco. *Geosciences (Switzerland)* 8 (2), 55. doi: 10.3390/geosciences8020055

Bjornlund, V., Bjornlund, H., and Van Rooyen, A. F. (2020). Why agricultural production in sub-Saharan Africa remains low compared to the rest of the world–a historical perspective. *Int. J. Water Resour. Development* 36 (sup1), S20–S53, 1–34. doi: 10.1080/07900627.2020.1739512

Blancon, J., Dutartre, D., Tixier, M. H., Weiss, M., Comar, A., Praud, S., et al. (2019). A high-throughput model-assisted method for phenotyping maize green leaf area index

dynamics using unmanned aerial vehicle imagery. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00685

Blein, R., Soule, B. G., Dupaigre, B. F., and Yerima, B. (2008). "Agricultural potential of west Africa," in *Economic community of West African States (ECOWAS)* (Foundation pour l'agriculture et la ruralité dans le monde (FARM), 118. Available at: https://old.fondation-farm.org/IMG/pdf/potentialites_rapport_ang_mp.pdf.

Bongomin, O., Lamo, J., Guina, J., Okello, C., Ocen, G., Obura, M., et al. (2022). *Applications of Drones and Image Analytics in Field Phenotyping: A Potential Breakthrough in Uganda's Agricultural Research*. SSRN Electron. J. Available at SSRN 4158755. https://dx.doi.org/10.2139/ssrn.4158755.

Bontpart, T., Concha, C., Giuffrida, M. V., Robertson, I., Admkie, K., Degefu, T., et al. (2020). Affordable and robust phenotyping framework to analyse root system architecture of soil-grown plants. *Plant J.* 103 (6), 2330–2343. doi: 10.1111/tpj.14877

Born, L., Chirinda, N., Mabaya, E., Afun-Ogidan, O., Girvetz, E. H., Jarvis, A., et al. (2021). *Digital agriculture profile: South Africa* (FAO). Available at: http://www.fao.org/3/cb2506en/CB2506EN.pdf (Accessed August 4, 2023).

Bouras, E. H., Jarlan, L., Er-Raki, S., Albergel, C., Richard, B., Balaghi, R., et al. (2020). Linkages between rainfed cereal production and agricultural drought through remote sensing indices and a land data assimilation system: A case study in Morocco. *Remote Sens.* 12 (24), 4018. doi: 10.3390/rs12244018

Bradshaw, C. D., Pope, E., Kay, G., Davie, J. C., Cottrell, A., Bacon, J., et al. (2022). Unprecedented climate extremes in South Africa and implications for maize production. *Environ. Res. Lett.* 17 (8), 084028. doi: 10.1088/1748-9326/ac816d

Brewer, K., Clulow, A., Sibanda, M., Gokool, S., Naiken, V., and Mabhaudhi, T. (2022b). Predicting the chlorophyll content of maize over phenotyping as a proxy for crop health in smallholder farming systems. *Remote Sens.* 14 (3), 518. doi: 10.3390/rs14030518

Brewer, K., Clulow, A., Sibanda, M., Gokool, S., Odindi, J., Mutanga, O., et al. (2022a). Estimation of maize foliar temperature and stomatal conductance as indicators of water stress based on optical and thermal imagery acquired using an unmanned aerial vehicle (UAV) platform. *Drones* 6 (7), 169. doi: 10.3390/drones6070169

Brocks, S., and Bareth, G. (2018). Estimating barley biomass with crop surface models from oblique RGB imagery. *Remote Sens.* 10 (2), 268. doi: 10.3390/rs10020268

Buchaillot, M. L., Gracia-Romero, A., Vergara-Diaz, O., Zaman-Allah, M. A., Tarekegne, A., Cairns, J. E., et al. (2019). Evaluating maize genotype performance under low nitrogen conditions using RGB UAV phenotyping techniques. *Sensors (Switzerland)* 19 (8), 1815. doi: 10.3390/s19081815

Burke, M., and Lobell, D. B. (2017). Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proc. Natl. Acad. Sci. United States America* 114 (9), 2189–2194. doi: 10.1073/pnas.1616919114

Burnette, M., Kooper, R., Maloney, J. D., Rohde, G. S., Terstriep, J. A., Willis, C., et al. (2018). TERRA-REF data processing infrastructure. In *Proceedings of the Practice and Experience on Advanced Research Computing*, pp. 1–7. doi: 10.1145/3219104.3219152

Busemeyer, L., Mentrup, D., Möller, K., Wunder, E., Alheit, K., Hahn, V., et al. (2013). Breedvision - A multi-sensor platform for non-destructive field-based phenotyping in plant breeding. *Sensors (Switzerland)* 13 (3), 2830–2847. doi: 10.3390/s130302830

Buthelezi, S., Mutanga, O., Sibanda, M., Odindi, J., Clulow, A. D., Chimonyo, V. G., et al. (2023). Assessing the prospects of remote sensing maize leaf area index using UAV-derived multi-spectral data in smallholder farms across the growing season. *Remote Sens.* 15 (6), 1597. doi: 10.3390/rs15061597

Camino, C., González-Dugo, V., Hernández, P., Sillero, J. C., and Zarco-Tejada, P. J. (2018). Improved nitrogen retrievals with airborne-derived fluorescence and plant traits quantified from VNIR-SWIR hyperspectral imagery in the context of precision agriculture. *Int. J. Appl. Earth Observation Geoinformation* 70, 105–117. doi: 10.1016/j.jag.2018.04.013

Casadesús, J., and Villegas, D. (2014). Conventional digital cameras as a tool for assessing leaf area index and biomass for cereal breeding. *J. of Integr. Plant Biol.* 56 (1), 7–14. doi: 10.1111/jipb.12117

Cavaco, A. M., Utkin, A. B., da Silva, J. M., and Guerra, R. (2022). Making sense of light: The use of optical spectroscopy techniques in plant sciences and agriculture. *Appl. Sci. (Switzerland)* 12 (3), 997. doi: 10.3390/app12030997

Chaerle, L., and Van Der Straeten, D. (2000). Imaging techniques and the early detection of plant stress. *Trends Plant Sci.* 5 (11), 495–501. doi: 10.1016/S1360-1385(00)01781-7

Chapman, S. C., Merz, T., Chan, A., Jackway, P., Hrabar, S., Dreccer, M. F., et al. (2014). Pheno-copter: A low-altitude, autonomous remote-sensing robotic helicopter for high-throughput field-based phenotyping. *Agronomy* 4 (2), 279–301. doi: 10.3390/agronomy4020279

Chapu, I., Okello, D. K., Okello, R. C. O., Odong, T. L., Sarkar, S., and Balota, M. (2022). Exploration of alternative approaches to phenotyping of late leaf spot and groundnut rosette virus disease for groundnut breeding. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.912332

Chawade, A., Van Ham, J., Blomquist, H., Bagge, O., Alexandersson, E., and Ortiz, R. (2019). High-throughput field-phenotyping tools for plant breeding and precision agriculture. *Agronomy* 9 (5), 258. doi: 10.3390/agronomy9050258

Chen, J. M., and Black, T. A. (1992). Defining leaf area index for non-flat leaves. *Plant Cell Environ.* 15 (4), 421–429. doi: 10.1111/j.1365-3040.1992.tb00992.x

Chivasa, W., Mutanga, O., and Biradar, C. (2017). Application of remote sensing in estimating maize grain yield in heterogeneous african agricultural landscapes: A review. *Int. J. Remote Sens.* 38 (23), 6816–6845. doi: 10.1080/01431161.2017.1365390

Chivasa, W., Mutanga, O., and Biradar, C. (2020). UAV-based multispectral phenotyping for disease resistance to accelerate crop improvement under changing climate conditions. *Remote Sens.* 12 (15), 2445. doi: 10.3390/RS12152445

Choukr-Allah, R., Rao, N. K., Hirich, A., Shahid, M., Alshankiti, A., Toderich, K., et al. (2016). Quinoa for marginal environments: toward future food and nutritional security in MENA and Central Asia regions. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00346

Comar, A., Burger, P., De Solan, B., Baret, F., Daumard, F., and Hanocq, J. F. (2012). A semi-automatic system for high throughput phenotyping wheat cultivars in-field conditions: Description and first results. *Funct. Plant Biol.* 39 (11), 914–924. doi: 10.1071/FP12065

Condorelli, G. E., Maccaferri, M., Newcomb, M., Andrade-Sanchez, P., White, J. W., French, A. N., et al. (2018). Corrigendum: Comparative aerial and ground based high throughput phenotyping for the genetic dissection of NDVI as a proxy for drought adaptive traits in durum wheat (Front. Plant Sci. 9, 893, 10.3389/fpls.2018.00893). *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01885

Costa, J. M., Grant, O. M., and Chaves, M. M. (2013). Thermography to explore plant-environment interactions. *J. Exp. Bot.* 64 (13), 564–584. doi: 10.1093/jxb/ert029

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

Cozzolino, D. (2014). Use of infrared spectroscopy for in-field measurement and phenotyping of plant properties: Instrumentation, data analysis, and examples. *Appl. Spectrosc. Rev.* 49 (7), 564–584. doi: 10.1080/05704928.2013.878720

Crain, J. L., Wei, Y., Barker, J., Thompson, S. M., Alderman, P. D., Reynolds, M., et al. (2016). Development and deployment of a portable field phenotyping platform. *Crop Sci.* 56 (3), 965–975. doi: 10.2135/cropsci2015.05.0290

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., et al. (2017). Genomic selection in plant breeding: Methods, models, and perspectives. *Trends Plant Sci.* 22 (11), 961–975. doi: 10.1016/j.tplants.2017.08.011

Dammer, K. H., Dworak, V., and Selbeck, J. (2016). On-the-go phenotyping in field potatoes using camera vision. *Potato Res.* 59 (2), 113–127. doi: 10.1007/s11540-016-9315-y

Daniel, I. O., Atinsola, K. O., Ajala, M. O., and Popoola, A. R. (2016). Phenotyping a tomato breeding population by manual field evaluation and digital imaging analysis. *Int. J. Plant Breed. Genet.* 11 (1), 19–24. doi: 10.3923/ijpbg.2017.19.24

Danzi, D., Briglia, N., Petrozza, A., Summerer, S., Povero, G., Stivaletta, A., et al. (2019). Can high throughput phenotyping help food security in the mediterranean area? *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00015

Debalke, D. B., and Abebe, J. T. (2022). Maize yield forecast using GIS and remote sensing in Kaffa Zone, South-West Ethiopia. *Environ. Syst. Res.* 11 (1), 1. doi: 10.1186/s40068-022-00249-5

Debesa, G., Gebre, S. L., Melese, A., Regassa, A., and Teka, S. (2020). GIS and remote sensing-based physical land suitability analysis for major cereal crops in Dabo Hana district, South-West Ethiopia. *Cogent Food Agric.* 6 (1), 1780100. doi: 10.1080/23311932.2020.1780100

Deery, D., Jimenez-Berni, J., Jones, H., Sirault, X., and Furbank, R. (2014). Proximal remote sensing buggies and potential applications for field-based phenotyping. *Agronomy* 4 (3), 875. doi: 10.3390/agronomy4030349

Deery, D. M., Rebetzke, G. J., Jimenez-Berni, J. A., Bovill, W. D., James, R. A., Condon, A. G., et al. (2019). Evaluation of the phenotypic repeatability of canopy temperature in wheat using continuous-terrestrial and airborne measurements. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00875

Degila, J., Sodedji, F. A. K., Avakoudjo, H. G. G., Tahi, S. P. G., Houetohossou, S. C. A., Honfoga, A. C., et al. (2023). Digital agriculture policies and strategies for innovations in the agri-food systems – cases of five West African Countries. *Sustainability* 15 (12), 9192. doi: 10.3390/su15129192

Dingkuhn, M., Sow, A., Manneh, B., Radanielina, T., Raboin, L. M., Dusserre, J., et al. (2015). Field phenomics for response of a rice diversity panel to ten environments in Senegal and Madagascar. 1. Plant phenological traits. *Field Crops Res.* 183, 342–355. doi: 10.1016/j.fcr.2015.07.027

Duvallet, M., Dumas, P., Makowski, D., Boé, J., del Villar, P. M., and Ben-Ari, T. (2021). Rice yield stability compared to major food crops in West Africa. *Environ. Res. Lett.* 16 (12), 124005. doi: 10.1088/1748-9326/ac343a

Ejikeme, J. O., Ojiako, J. C., Onwuzuligbo, C. U., and Ezeh, F. C. (2017). Enhancing food security in Anambra state, Nigeria using remote sensing data. *Environ. Rev.* 6 (1), 27–44. Available at: http://erjournal.net/index.php/erjournal/article/viewFile/33/pdf4

Elmetwalli, A. H., Mazrou, Y. S. A., Tyler, A. N., Hunter, P. D., Elsherbiny, O., Yaseen, Z. M., et al. (2022). Assessing the efficiency of remote sensing and machine learning algorithms to quantify wheat characteristics in the nile delta region of Egypt. *Agriculture* 12 (3), 332. doi: 10.3390/agriculture12030332

Elmetwalli, A. H., El-Hendawy, S., Al-Suhaibani, N., Alotaibi, M., Tahir, M. U., Mubushar, M., et al. (2020). Potential of hyperspectral and thermal proximal sensing for estimating growth performance and yield of soybean exposed to different drip irrigation regimes under arid conditions. *Sensors* 20 (22), 6569. doi: 10.3390/s20226569

Elsafty, A., and Atallah, A. (2022). Factors influencing precision agriculture tools or technologies adoption in Egypt. *Business Manage. Stud.* 8 (2). doi: 10.11114/bms.v8i2.5598

Elsayed, S., Elhoweity, M., Ibrahim, H. H., Dewir, Y. H., Migdadi, H. M., and Schmidhalter, U. (2017). Thermal imaging and passive reflectance sensing to estimate the water status and grain yield of wheat under different irrigation regimes. *Agric. Water Manage.* 189, 98–110. doi: 10.1016/j.agwat.2017.05.001

El-Shirbeny, M. A., Ali, A., and Saleh, N. H. (2014). Crop water requirements in Egypt using remote sensing techniques. *J. Agric. Chem. Environ.* 3 (2), 57–65. doi: 10.4236/jacen.2014.32B010

Epule, T. E., Chehbouni, A., and Dhiba, D. (2022). Recent patterns in maize yield and harvest area across Africa. *Agronomy* 12 (2), 374. doi: 10.3390/agronomy12020374

Erdle, K., Mistele, B., and Schmidhalter, U. (2013). Spectral assessments of phenotypic differences in spike development during grain filling affected by varying N supply in wheat. *J. Plant Nutr. Soil Sc.* 176, 952–963. doi: 10.1002/jpln.201300247

Fahrner, S., Janni, M., Pieruschka, R., Vincenz-Donnelly, L., and von Gillhaussen, P. (2021). *Global plant phenotyping survey 2020/21* (Zenodo). doi: 10.5281/zenodo.4723409

FAO. (2011). Evolving a plant breeding and seed system in sub-Saharan Africa in an era of donor dependence. *A report for the Global Partnership Initiative for Plant Breeding Capacity Building (GIPB) of the Food and Agricultural Organisation*. FAO plant production and protection paper 210, pp. 21–30. Available at: https://www.fao.org/3/at535e/at535e.pdfb.

FAO. (2015). *FAOSTAT. Food and agricultural organisation of the United Nations*. Available at: http://www.fao.org/faostat/en/#data/.

FAO. (2022). *FAOSTAT. Database of the food and agriculture organization united nations*. Available at: www.fao.org/faostat/en/.

FAOSTAT. (2022). *Database of the Food and Agriculture Organization of the United Nations, Rome, Italy*. Available at: https://www.fao.org/faostat/en/. Accessed on August 12, 2023.

FAOSTAT. (2018a). *Database of the food and agriculture organization united nations*. Available at: www.fao.org/faostat/en/.

FAOSTAT. (2018b). *Morocco production in 2018*. Available at: https://www.fao.org/faostat/en/#data/QCL.

FAOSTAT. (2016). FAOSTAT statistical database. *Food and Agriculture Organization of the United Nations, Rome, Italy*. Accessed on August 12, 2023.

Feng, L., Chen, S., Zhang, C., Zhang, Y., and He, Y. (2021). A comprehensive review on recent applications of unmanned aerial vehicle remote sensing with various sensors for high-throughput plant phenotyping. *Comput. Electron. Agric.* 182, 106033. doi: 10.1016/j.compag.2021.106033

Feng, W., Yao, X., Zhu, Y., Tian, Y. C., and Cao, W. X. (2008). Monitoring leaf nitrogen status with hyperspectral reflectance in wheat. *Eur. J. Agron.* 28 (3), 394–404. doi: 10.1016/j.eja.2007.11.005

Fernández-Calleja, M., Monteagudo, A., Casas, A. M., Boutin, C., Pin, P. A., Morales, F., et al. (2020). Rapid on-site phenotyping via field fluorimeter detects differences in photosynthetic performance in a hybrid—parent barley germplasm set. *Sensors (Switzerland)* 20 (5), 1486. doi: 10.3390/s20051486

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

Fisher, M., Abate, T., Lunduka, R. W., Asnake, W., Alemayehu, Y., and Madulu, R. B. (2015). Drought tolerant maize for farmer adaptation to drought in sub-Saharan Africa: Determinants of adoption in eastern and southern Africa. *Climatic Change* 133, 283–299. doi: 10.1007/s10584-015-1459-2

Fitzgerald, G. J., Rodriguez, D., Christensen, L. K., Belford, R., Sadras, V. O., and Clarke, T. R. (2006). Spectral and thermal sensing for nitrogen and water status in rainfed and irrigated wheat environments. *Precis. Agric.* 7 (4), 233–248. doi: 10.1007/s11119-006-9011-z

Fritsche-Neto, R., and Borém, A. (2015). Phenomics: How next-generation phenotyping is revolutionizing plant breeding. In *Phenomics: How Next-Generation Phenotyping is Revolutionizing Plant Breeding*. (Switzerland: Springer International Publishing). doi: 10.1007/978-3-319-13677-6

Furbank, R. T., Jimenez-Berni, J. A., George-Jaeggli, B., Potgieter, A. B., and Deery, D. M. (2019). Field crop phenomics: enabling breeding for radiation use efficiency and biomass in cereal crops. *New Phytol.* 223 (4), 1714–1727. doi: 10.1111/nph.15817

Gago, J., Douthe, C., Coopman, R. E., Gallego, P. P., Ribas-Carbo, M., Flexas, J., et al. (2015). UAVs challenge to assess water stress for sustainable agriculture. *Agric. Water Manage.* 153, 9–19. doi: 10.1016/j.agwat.2015.01.020

Gano, B., Dembele, J. S. B., Ndour, A., Luquet, D., Beurier, G., Diouf, D., et al. (2021). Adaptation responses to early drought stress of West Africa sorghum varieties. *Agronomy* 11 (5), 850. doi: 10.3390/agronomy11050850

Garriga, M., Romero-Bravo, S., Estrada, F., Escobar, A., Matus, I. A., del Pozo, A., et al. (2017). Assessing wheat traits by spectral reflectance: Do we really need to focus on predicted trait-values or directly identify the elite genotypes group? *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00280

Gedil, M., and Menkir, A. (2019). An integrated molecular and conventional breeding scheme for enhancing genetic gain in maize in Africa. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01430

Gibbs, J. A., Pound, M., French, A. P., Wells, D. M., Murchie, E., and Pridmore, T. (2017). Approaches to three-dimensional reconstruction of plant shoot topology and geometry. *Funct. Plant Biol.* 44 (1), 62–75. doi: 10.1071/FP16167

Gobezie, T. B., and Biswas, A. (2023). The need for streamlining precision agriculture data in Africa. *Precis. Agric.* 24 (1), 375–383. doi: 10.1007/s11119-022-09928-w

Godfray, H. C. J., and Robinson, S. (2015). Contrasting approaches to projecting long-run global food security. *Oxf Rev. Economic Policy* 31 (1), 26–44. doi: 10.1093/oxrep/grv006

Gokool, S., Mahomed, M., Kunz, R., Clulow, A., Sibanda, M., Naiken, V., et al. (2023). Crop monitoring in smallholder farms using unmanned aerial vehicles to facilitate precision agriculture practices: a scoping review and bibliometric analysis. *Sustainability* 15 (4), 3557. doi: 10.3390/su15043557

Gonzalez-Dugo, V., Hernandez, P., Solis, I., and Zarco-Tejada, P. J. (2015). Using high-resolution hyperspectral and thermal airborne imagery to assess physiological condition in the context of wheat phenotyping. *Remote Sens.* 7 (10), 13586–13605. doi: 10.3390/rs71013586

Gourlay, S., Kilic, T., and Lobell, D. (2017). Could the debate be over? Errors in farmer-reported production and their implications for the inverse scale-productivity relationship in Uganda. *Technical Report, Policy Research Working Paper No. 8192*. (Washington, D C: World Bank). doi: 10.1596/1813-9450-8192

Gracia-Romero, A., Kefauver, S. C., Vergara-Díaz, O., Hamadziripi, E., Zaman-Allah, M. A., Thierfelder, C., et al. (2020). Leaf versus whole-canopy remote sensing methodologies for crop monitoring under conservation agriculture: a case of study with maize in Zimbabwe. *Sci. Rep.* 10 (1), 16008. doi: 10.1038/s41598-020-73110-3

Gracia-Romero, A., Vergara-Díaz, O., Thierfelder, C., Cairns, J. E., Kefauver, S. C., and Araus, J. L. (2018). Phenotyping conservation agriculture management effects on ground and aerial remote sensing assessments of maize hybrids performance in Zimbabwe. *Remote Sens.* 10 (2), 349. doi: 10.3390/rs10020349

Guo, W., Fukatsu, T., and Ninomiya, S. (2015). Automated characterization of flowering dynamics in rice using field-acquired time-series RGB images. *Plant Methods* 11 (1), 1–15. doi: 10.1186/s13007-015-0047-9

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 (3), 328–339. doi: 10.1007/s11427-017-9056-0

Hall, O., Dahlin, S., Marstorp, H., Archila Bustos, M. F., Öborn, I., and Jirström, M. (2018). Classification of maize in complex smallholder farming systems using UAV imagery. *Drones* 2 (3), 22. doi: 10.3390/drones2030022

Hamidi, S., and Benabdeljalil, N. (2013). National innovation systems: The Moroccan case. *Proc. - Soc. Behav. Sci.* 75, 119–128. doi: 10.1016/J.SBSPRO.2013.04.014

Hassan, M. A., Yang, M., Fu, L., Rasheed, A., Zheng, B., Xia, X., et al. (2019). Accuracy assessment of plant height using an unmanned aerial vehicle for quantitative genomic analysis in bread wheat. *Plant Methods* 15 (1), 1–12. doi: 10.1186/s13007-019-0419-7

Hedley, C. (2015). The role of precision agriculture for improved nutrient management on farms. *J. Sci. Food Agric.* 95 (1), 12–19. doi: 10.1002/jsfa.6734

Henri, A. O. (2019). Heavily indebted poor countries initiative (HIPC), debt relief, economic stability and economic growth in Africa. *Economic Change Restructuring* 52, 89–121. doi: 10.1007/s10644-017-9218-1

Hirich, A., Rafik, S., Rahmani, M., Fetouab, A., Azaykou, F., Filali, K., et al. (2021). Development of quinoa value chain to improve food and nutritional security in rural communities in Rehamna, Morocco: lessons learned and perspectives. *Plants* 10 (2), 301. doi: 10.3390/plants10020301

Holman, F. (2020). *Development and evaluation of unmanned aerial vehicles for high throughput phenotyping of field-based wheat trials*. (Doctoral dissertation, King's College London). Available at: https://repository.rothamsted.ac.uk/item/97q41/development-and-evaluation-of-unmanned-aerial-vehicles-for-high-throughput-phenotyping-of-field-based-wheat-trials

Holman, F. H., Riche, A. B., Michalski, A., Castle, M., Wooster, M. J., and Hawkesford, M. J. (2016). High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. *Remote Sens.* 8 (12), 1031. doi: 10.3390/rs8121031

ICARDA phenotyping platforms in Morocco. Available at: https://www.cgiar.org/news-events/news/icardas-phenotyping-facilities-a-game-changing-solution-for-abiotic-stress-tolerance-in-crops/.

Ifeanyieze, F. O., Ikehi, M. E., and Isiwu, E. (2014). Techniques in utilizing remote sensor technology for precision crop production by farmers as climate change adaptation strategy in Nigeria. *Agric. Sci.* 5 (14), 1476. doi: 10.4236/as.2014.514158

Iizumi, T., and Sakai, T. (2020). The global dataset of historical yields for major crops 1981–2016. *Sci. Data* 7 (1), 97. doi: 10.1038/s41597-020-0433-7

IPPN. (2016). The international plant phenotyping network. In: *Survey 2016*. Available at: http://www.plant-phenotyping.org/lw_resource/datapool/systemfiles/elements/files/ba090795-026e-11e7-8c78-dead53a91d31/live/document/Survey_2016_final_2.pdf (Accessed June 20, 2023).

IPPN. (2020). *Survey*. Available at: https://globalplantcouncil.org/global-plant-phenotyping-survey-2020-21/.

Iseki, K., and Matsumoto, R. (2019). Non-destructive shoot biomass evaluation using a handheld NDVI sensor for field-grown staking Yam (Dioscorea rotundata Poir.). *Plant Production Sci.* 22 (2), 301–310. doi: 10.1080/1343943X.2018.1540278

Izuogu, C. U., Njoku, L. C., Olaolu, M. O., Kadurumba, P. C., Azuamairo, G. C., and Agou, G. D. (2023). A review of the digitalization of agriculture in Nigeria. *J. Agric. Extension* 27 (2), 47–64. doi: 10.4314/jae.v27i2.5

Jabir, B., and Falih, N. (2020). Digital agriculture in Morocco, opportunities and challenges. In *Proceedings of the 2020 6th International Conference on Optimization and Applications (ICOA)*, Beni Mellal, Morocco. (Piscataway, New Jersey, United States: Institute of Electrical and Electronics Engineers (IEEE)), 1–5, IEEE. doi: 10.1109/ICOA49421.2020.9094450

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. doi: 10.3389/fpls.2018.00237

Kassim, Y. B., Oteng-Frimpong, R., Puozaa, D. K., Sie, E. K., Abdul Rasheed, M., Abdul Rashid, I., et al. (2022). High-throughput plant phenotyping (HTPP) in resource-constrained research programs: A working example in Ghana. *Agronomy* 12 (11), 2733. doi: 10.3390/agronomy12112733

Kavhiza, N. J., Zargar, M., Prikhodko, S. I., Pakina, E. N., Murtazova, K. M. S., and Nakhaev, M. R. (2022). Improving crop productivity and ensuring food security through the adoption of genetically modified crops in sub-saharan Africa. *Agronomy* 12 (2), 439. doi: 10.3390/agronomy12020439

Kefauver, S. C., El-Haddad, G., Vergara-Diaz, O., and Araus, J. L. (2015). RGB picture vegetation indexes for high-throughput phenotyping platforms (HTPPs). *Remote Sens. agriculture ecosystems hydrology XVII* 9637, 82–90. SPIE. doi: 10.1117/12.2195235

Kidane, Y. G., Hailemariam, B. N., Mengistu, D. K., Fadda, C., Pè, M. E., and Dell'Acqua, M. (2017). Genome-wide association study of Septoria tritici blotch resistance in Ethiopian durum Wheat Landraces. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01586

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

Kipp, S., Mistele, B., and Schmidhalter, U. (2014). Identification of stay-green and early senescence phenotypes in high-yielding winter wheat, and their relationship to grain yield and grain protein concentration using high-throughput phenotyping techniques. *Funct. Plant Biol.* 41 (3), 227–235. doi: 10.1071/FP13221

Kirchgessner, N., Liebisch, F., Yu, K., Pfeifer, J., Friedli, M., Hund, A., et al. (2017). The ETH field phenotyping platform FIP: A cable-suspended multi-sensor system. *Funct. Plant Biol.* 44 (1), 154–168. doi: 10.1071/FP16165

Knox, J., Hess, T., Daccache, A., and Wheeler, T. (2012). Climate change impacts on crop productivity in Africa and South Asia. *Environ. Res. Lett.* 7 (3), 034032. doi: 10.1088/1748-9326/7/3/034032

Kobo, H. I., Abu-Mahfouz, A. M., and Hancke, G. P. (2017). A survey on software-defined wireless sensor networks: Challenges and design requirements. *IEEE Access* 5, 1872–1899. doi: 10.1109/ACCESS.2017.2666200

Kotikot, S. M., and Onywere, S. M. (2015). Application of GIS and remote sensing techniques in frost risk mapping for mitigating agricultural losses in the Aberdare ecosystem, Kenya. *Geocarto Int.* 30 (1), 104–121. doi: 10.1080/10106049.2014.965758

Kpienbaareh, D., Kansanga, M., and Luginaah, I. (2019). Examining the potential of open-source remote sensing for building effective decision support systems for precision agriculture in resource-poor settings. *GeoJournal* 84 (6), 1481–1497. doi: 10.1007/s10708-018-9932-x

Kumar, D., Kushwaha, S., Delvento, C., Liatukas, Ž., Vivekanand, V., Svensson, J. T., et al. (2020). Affordable phenotyping of winter wheat under field and controlled conditions for drought tolerance. *Agronomy* 10 (6), 882. doi: 10.3390/agronomy10060882

Laachrate, H., Fadil, A., and Ghafiri, A. (2020). Drought monitoring of some Moroccan agricultural areas: Settat and Meknes using remote sensing techniques and a soil moisture-based drought index. In *Proceedings-2020 IEEE International Conference of Moroccan Geomatics*, MORGEO 2020. doi: 10.1109/Morgeo49228.2020.9121917

Langemeier, M., and Boehlje, M. (2021). What will be the capabilities and skills needed to manage the farm of the future? *Farmdoc Dly.* 11, 1–4. Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign.

Leakey, R. R., Tientcheu Avana, M. L., Awazi, N. P., Assogbadjo, A. E., Mabhaudhi, T., Hendre, P. S., et al. (2022). The future of food: Domestication and commercialization of indigenous food crops in Africa over the third decade, (2012–2021). *Sustainability* 14 (4), 2355. doi: 10.3390/su14042355

Li, D., Quan, C., Song, Z., Li, X., Yu, G., Li, C., et al. (2021). High-throughput plant phenotyping platform (HT3P) as a novel tool for estimating agronomic traits from the lab to the field. *Front. Bioengineering Biotechnol.* 8. doi: 10.3389/fbioe.2020.623705

Li, L., Zhang, Q., and Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors (Switzerland)* 14 (11), 20078–20111. doi: 10.3390/s141120078

Liu, T., Li, R., Jin, X., Ding, J., Zhu, X., Sun, C., et al. (2017). Evaluation of seed emergence uniformity of mechanically sown wheat with UAV RGB imagery. *Remote Sens.* 9 (12), 1241. doi: 10.3390/rs9121241

Livingston, G., Schonberger, S., and Delaney, S. (2011). Sub-Saharan Africa: The state of smallholders in agriculture. *Paper presented at the Conference on New Directions for Smallholder Agriculture*. (Rome: International Fund for Agricultural Development (IFAD)) 24, 25.

A. Lorence and K. M. Jimenez (Eds.) (2022). *High-throughput plant phenotyping: Methods and protocols* Vol. 2539 (New York, USA: Springer Nature). Available at: https://doi.org/10.1007/978-1-0716-2537-8

Mahdy, E. M. B., and Ahmad, H. (2023). A study of Corchorus L. diversity in Egypt using high-throughput phenotyping platform (HTPP): an Egyptian gene bank example. *Genet. Resour. Crop Evolution.* 70, 1–11. doi: 10.1007/s10722-023-01551-6

Maimaitijiang, M., Ghulam, A., Sidike, P., Hartling, S., Maimaitiyiming, M., Peterson, K., et al. (2017). Unmanned Aerial System (UAS)-based phenotyping of soybean using multi-sensor data fusion and extreme learning machine. *ISPRS J. Photogrammetry Remote Sens.* 134, 43–58. doi: 10.1016/j.isprsjprs.2017.10.011

Manzi, H., and Gweyi-Onyango, J. P. (2021). "Agro-ecological lower midland zones IV and V in Kenya using GIS and remote sensing for climate-smart crop management," in *African handbook of climate change adaptation*. Eds. N. Oguge, D. Ayal, L. Adeleke and I. da Silva (Cham: Springer). doi: 10.1007/978-3-030-45106-6_35

Matsumoto, T., Wu, J., Kanamori, H., Katayose, Y., Fujisawa, M., Namiki, N., et al. (2005). The map-based sequence of the rice genome. *Nature* 436 (7052), 793–800. doi: 10.1038/nature03895

Mbaye, M., Ndour, A., Gano, B., Dembele, J. S. B., Luquet, D., Beurier, G., et al. (2022). "UAV method based on multispectral imaging for field phenotyping," in *Crop adaptation and improvement for drought-prone environments*. Eds. N. A. Kane, K. Daniel and T. J. Dalton (Manhattan: New Prairie Press), 173–187, ISBN: .

Meliho, M., Khattabi, A., Jobbins, G., and Sghir, F. (2020). Impact of meteorological drought on agriculture in the Tensift watershed of Morocco. *J. Water Climate Change* 11 (4), 1323–1338. doi: 10.2166/wcc.2019.279

Mengistu, D. K., Kidane, Y. G., Fadda, C., and Pè, M. E. (2018). Genetic diversity in Ethiopian durum wheat (Triticum turgidum var durum) inferred from phenotypic variations. *Plant Genet. Resour.* 16 (1), 39–49. doi: 10.1017/S1479262116000393

Minervini, M., Scharr, H., and Tsaftaris, S. A. (2015).). Image analysis: The new bottleneck in plant phenotyping [applications corner]. *IEEE Signal Process. Magazine* 32 (4), 126–131. doi: 10.1109/MSP.2015.2405111

Mukhawana, M. B., Kanyerere, T., and Kahler, D. (2023). Review of in-Situ and remote sensing-based indices and their Applicability for integrated drought monitoring in South Africa. *Water* 15 (2), 240. doi: 10.3390/w15020240

Müller-Linow, M., Pinto-Espinosa, F., Scharr, H., and Rascher, U. (2015). The leaf angle distribution of natural plant populations: Assessing the canopy with a novel software tool. *Plant Methods* 11 (1), 1–16. doi: 10.1186/s13007-015-0052-z

Munghemezulu, C., Mashaba-Munghemezulu, Z., Ratshiedana, P. E., Economon, E., Chirima, G., and Sibanda, S. (2023). Unmanned aerial vehicle (UAV) and spectral datasets in South Africa for precision agriculture. *Data* 8 (6), 98. doi: 10.3390/data8060098

Musungwini, S. (2018). Mobile phone use by Zimbabwean smallholder farmers: A baseline study. *Afr. J. Inf. Communication* 22, 29–52. doi: 10.23962/10539/26171

Mutanga, O., Dube, T., and Ahmed, F. (2016). Progress in remote sensing: vegetation monitoring in South Africa. *South Afr. Geographical J.* 98 (3), 461–471. doi: 10.1080/03736245.2016.1208586

Neilson, E. H., Edwards, A. M., Blomstedt, C. K., Berger, B., Møller, B. L., and Gleadow, R. M. (2015). Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal crop plant to nitrogen and water deficiency over time. *J. Exp. Bot.* 66 (7), 1817–1832. doi: 10.1093/jxb/eru526

Ninomiya, S. (2022). High-throughput field crop phenotyping: current status and challenges. *Breed. Sci.* 72 (1), 3–18. doi: 10.1270/jsbbs.21069

Nyaga, J. M., Onyango, C. M., Wetterlind, J., and Söderström, M. (2021). Precision agriculture research in sub-Saharan Africa countries: A systematic map. *Precis. Agric.* 22, 1217–1236. doi: 10.1007/s11119-020-09780-w

Ochieng, J., Knerr, B., Owuor, G., and Ouma, E. (2020). Food crops commercialization and household livelihoods: Evidence from rural regions in Central Africa. *Agribusiness* 36 (2), 318–338. doi: 10.1002/agr.21619

Ogutu, G. E. O., Franssen, W. H. P., Supit, I., Omondi, P., and Hutjes, R. W. A. (2018). Probabilistic maize yield prediction over East Africa using dynamic ensemble seasonal climate forecasts. *Agric. For. Meteorology* 250–251, 243–261. doi: 10.1016/j.agrformet.2017.12.256

Osiemo, J., Girvetz, E. H., Hasiner, E., Schroeder, K., Treguer, D., Juergenliemk, A., et al. (2021). *Digital agriculture profile: Kenya* (Rome: FAO). Available at: https://www.fao.org/3/cb3958en/cb3958en.pdf (Accessed August 8, 2023).

Ouraich, I., and Tyner, W. E. (2018). Moroccan agriculture, climate change, and the Moroccan Green Plan: A CGE analysis. *Afr. J. Agric. Resource Economics* 13 (311-2019-681), 307–330. doi: 10.22004/ag.econ.284990

Pajares, G. (2015). Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVs). *Photogrammetric Eng. Remote Sens.* 81 (4), 281–330. doi: 10.14358/PERS.81.4.281

Palmer, P. I., Wainwright, C. M., Dong, B., Maidment, R. I., Wheeler, K. G., Gedney, N., et al. (2023). Drivers and impacts of Eastern African rainfall variability. *Nat. Rev. Earth Environ.* 4 (4), 254–270. doi: 10.1038/s43017-023-00397-x

Park, S., Ryu, D., Fuentes, S., Chung, H., Hernández-Montes, E., and O'Connell, M. (2017). Adaptive estimation of crop water stress in nectarine and peach orchards using high-resolution imagery from an unmanned aerial vehicle (UAV). *Remote Sens.* 9 (8), 828. doi: 10.3390/rs9080828

Parkes, B., DeFrance, D., Sultan, B., Ciais, P., and Wang, X. (2018). Projected changes in crop yield mean and variability over West Africa in a world 1.5K warmer than the pre-industrial era. *Earth System Dynamics* 9 (1), 119–134. doi: 10.5194/esd-9-119-2018

Parks, S. E., Irving, D. E., and Milham, P. J. (2012). A critical evaluation of on-farm rapid tests for measuring nitrate in leafy vegetables. *Scientia Hortic.* 134, 1–6. doi: 10.1016/j.scienta.2011.10.015

Parwada, C., and Marufu, H. (2023). Digitalisation of agriculture in Zimbabwe: Challenges and opportunities. *Int. J. Sustain. Agric. Res.* 10 (1), 32–41. doi: 10.18488/ijsar.v10i1.3280

Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature* 457 (7229), 551–556. doi: 10.1038/nature07723

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 Genet.* 6 (4), 865–879. doi: 10.1534/g3.115.023515

Peng, Y., Gitelson, A. A., Keydan, G., Rundquist, D. C., and Moses, W. (2011). Remote estimation of gross primary production in maize and support for a new paradigm based on total crop chlorophyll content. *Remote Sens. Environ.* 115 (4), 978–989. doi: 10.1016/j.rse.2010.12.001

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 (6), 1094. doi: 10.3390/rs13061094

Pieruschka, R., and Schurr, U. (2019). Plant phenotyping: Past, present, and future. *Plant Phenomics* 2019. doi: 10.34133/2019/7507131

Pieruschka, R., and Schurr, U. (2022). "Origins and drivers of crop phenotyping," in *Advances in plant phenotyping for more sustainable crop production*. Ed. A. Walter (Cambridge, UK: Burleigh Dodds Science Publishing Limited), 1–25.

Pineda, M., Barón, M., and Pérez-Bueno, M. L. (2021). Thermal imaging for plant stress detection and phenotyping. *Remote Sens.* 13 (1), 68. doi: 10.3390/rs13010068

Poblete, T., Ortega-Farías, S., and Ryu, D. (2018). Automatic co-registration algorithm to remove canopy shaded pixels in UAV-borne thermal images to improve the estimation of crop water stress index of a drip-irrigated cabernet sauvignon vineyard. *Sensors (Switzerland)* 18 (2), 397. doi: 10.3390/s18020397

Potgieter, A. B., Watson, J., Eldridge, M., Laws, K., George-Jaeggli, B., Hunt, C., et al. (2018). Determining crop growth dynamics in sorghum breeding trials through remote and proximal sensing technologies. *International Geoscience and Remote Sensing Symposium (IGARSS)*, Valencia, Spain. (Piscataway, New Jersey, United States: The Institute of Electrical and Electronics Engineers (IEEE)). doi: 10.1109/IGARSS.2018.8519296

Qiu, Q., Sun, N., Bai, H., Wang, N., Fan, Z., Wang, Y., et al. (2019). Field-based high-throughput phenotyping for maize plant using 3d LIDAR point cloud generated with a "phenomobile". *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00554

Qiu, R., Wei, S., Zhang, M., Li, H., Sun, H., Liu, G., et al. (2018). Sensors for measuring plant phenotyping: A review. *Int. J. Agric. Biol. Eng.* 11 (2), 1–17. doi: 10.25165/j.ijabe.20181102.2696

Quahir,, et al. (2022). *Drought research in the dry areas of Africa: the case of PhenoMA the high throughput phenotyping Platform in Morocco*. Available at: https://www.plantphenotyping.org/lw_resource/datapool/systemfiles/elements/files/0dec546e-673e-11ed-9086-dead53a91d31/current/document/abstract_Book_IPPS_2022.pdf.

Quemada, M., Gabriel, J. L., and Zarco-Tejada, P. (2014). Airborne hyperspectral images and ground-level optical sensors as assessment tools for maize nitrogen fertilization. *Remote Sens.* 6 (4), 2940–2962. doi: 10.3390/rs6042940

Rahaman, M. M., Chen, D., Gillani, Z., Klukas, C., and Chen, M. (2015). Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00619

Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C., and Foley, J. A. (2012). Recent patterns of crop yield growth and stagnation. *Nat. Commun.* 3, 1293. doi: 10.1038/ncomms2296

Rebetzke, G. J., Fischer, R. A., Van Herwaarden, A. F., Bonnett, D. G., Chenu, K., Rattey, A. R., et al. (2014). Plot size matters: Interference from intergenotypic competition in plant phenotyping studies. *Funct. Plant Biol.* 41 (2), 107–118. doi: 10.1071/FP13177

Reynolds, D., Baret, F., Welcker, C., Bostrom, A., Ball, J., Cellini, F., et al. (2019). What is cost-efficient phenotyping? Optimizing costs for different scenarios. *Plant Sci.* 282, 14–22. doi: 10.1016/j.plantsci.2018.06.015

Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., et al. (2020). Breeder friendly phenotyping. *Plant Sci.* 295, 110396. doi: 10.1016/j.plantsci.2019.110396

Rezende, W. S., Beyene, Y., Mugo, S., Ndou, E., Gowda, M., Sserumaga, J. P., et al. (2020). Performance and yield stability of maize hybrids in stress-prone environments in eastern Africa. *Crop J.* 8 (1), 107–118. doi: 10.1016/j.cj.2019.08.001

Roitsch, T., Cabrera-Bosquet, L., Fournier, A., Ghamkhar, K., Jiménez-Berni, J., Pinto, F., et al. (2019). Review: New sensors and data-driven approaches—A path to next generation phenomics. *Plant Sci.* 282, 2–10. doi: 10.1016/j.plantsci.2019.01.011

Rosenqvist, E., Großkinsky, D. K., Ottosen, C. O., and van de Zedde, R. (2019). The phenotyping dilemma—the challenges of a diversified phenotyping community. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00163

Roudier, P., Sultan, B., Quirion, P., and Berg, A. (2011). The impact of future climate change on West African crop yields: What does the recent literature say? *Global Environ. Change* 21 (3), 1073–1083. doi: 10.1016/j.gloenvcha.2011.04.007

Sadeghi-Tehran, P., Virlet, N., and Hawkesford, M. J. (2021). A neural network method for classification of sunlit and shaded components of wheat canopies in the field using high-resolution hyperspectral imagery. *Remote Sens.* 13 (5), 898. doi: 10.3390/rs13050898

Sagan, V., Maimaitijiang, M., Sidike, P., Eblimit, K., Peterson, K. T., Hartling, S., et al. (2019). UAV-based high resolution thermal imaging for vegetation monitoring, and plant phenotyping using ICI 8640 P, FLIR Vue Pro R 640, and thermomap cameras. *Remote Sens.* 11 (3), 330. doi: 10.3390/rs11030330

Saidi, A. S., and Diouri, M. (2017). Food self-sufficiency under the Green-Morocco Plan. *J. Exp. Biol. Agric. Sci.* 5 (Spl-1-SAFSAW), 33–40. doi: 10.18006/2017.5(Spl-1-SAFSAW).S33.S40

Salami, A., Kamara, A. B., and Schwidrowski, Z. B. (2010). *Smallholder agriculture in east Africa: Trends, constraints and opportunities* (Tunis, Tunisia: African Development Bank).

Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., et al. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *Eur. J. Agron.* 70, 112–123. doi: 10.1016/j.eja.2015.07.004

Sayed, S. A., Mahmoud, A. S., Farg, E., Mohamed, A. M., Saleh, A. M., AbdelRahman, M. A., et al. (2023). A comparative study of big data use in Egyptian agriculture. *J. Electrical Syst. Inf. Technol.* 10 (1), 21. doi: 10.1186/s43067-023-00090-5

Scheben, A., Batley, J., and Edwards, D. (2018). Revolution in genotyping platforms for crop improvement. *Adv. Biochem. Engineering/Biotechnology* 164, 37–52. doi: 10.1007/10_2017_47

Schirrmann, M., Hamdorf, A., Garz, A., Ustyuzhanin, A., and Dammer, K. H. (2016). Estimating wheat biomass by combining image clustering with crop height. *Comput. Electron. Agric.* 121, 374–384. doi: 10.1016/j.compag.2016.01.007

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463 (7278), 178–183. doi: 10.1038/nature08670

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: Complexity, diversity, and dynamics. *Science* 326 (5956), 1112–1115. doi: 10.1126/science.1178534

Seiffert, U., Bollenbeck, F., Mock, H. P., and Matros, A. (2010). Clustering of crop phenotypes by means of hyperspectral signatures using artificial neural networks. *2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHISPERS 2010 - Workshop Program*, Reykjavik, Iceland. (Piscataway, New Jersey, United: The Institute of Electrical and Electronics Engineers (IEEE)), 1–4. doi: 10.1109/WHISPERS.2010.5594947

Shakoor, N., Lee, S., and Mockler, T. C. (2017). High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. *Curr. Opin. Plant Biol.* 38, 184–192. doi: 10.1016/j.pbi.2017.05.006

Shi, Y., Alex Thomasson, J., Murray, S. C., Ace Pugh, N., Rooney, W. L., Shafian, S., et al. (2016). Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PLoS One* 11 (7), e0159781. doi: 10.1371/journal.pone.0159781

Shimeles, A., Verdier-Chouchane, A., and Boly, A. (2018). Introduction: Understanding the Challenges of the Agricultural Sector in Sub-Saharan Africa. In: A. Shimeles, A. Verdier-Chouchane and A. Boly (eds) *Building a Resilient and Sustainable Agriculture in Sub-Saharan Africa*. (Cham., Switzerland: Palgrave Macmillan), 1–12. doi: 10.1007/978-3-319-76222-7_1

Shokr, M. E. (2020). Environmental Applications of Remote Sensing in Egypt: A Review and an Outlook. In: S. Elbeih, A. Negm and A. Kostianoy (eds) *Environmental Remote Sensing in Egypt*. (Cham, Switzerland: Springer Geophysics. Springer), 95–126. doi: 10.1007/978-3-030-39593-3_4

Shonhe, T., and Scoones, I. (2022). Private and state-led contract farming in Zimbabwe: Accumulation, social differentiation and rural politics. *J. Agrarian Change* 22 (1), 118–138. doi: 10.1111/joac.12473

Sie, E. K., Oteng-Frimpong, R., Kassim, Y. B., Puozaa, D. K., Adjebeng-Danquah, J., Masawudu, A. R., et al. (2022). RGB-image method enables indirect selection for leaf spot resistance and yield estimation in a groundnut breeding program in Western Africa. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.957061

Smidt, H. J., and Jokonya, O. (2022). Factors affecting digital technology adoption by small-scale farmers in agriculture value chains (AVCs) in South Africa. *Inf. Technol. Dev.* 28 (3), 558–584. doi: 10.1080/02681102.2021.1975256

Soullier, G., Demont, M., Arouna, A., Lançon, F., and Del Villar, P. M. (2020). The state of rice value chain upgrading in West Africa. *Global Food Secur.* 25, 100365. doi: 10.1016/j.gfs.2020.100365

Sultan, B., and Gaetani, M. (2016). Agriculture in West Africa in the twenty-first century: climate change and impacts scenarios, and potential for adaptation. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01262

Swinfield, T., Lindsell, J. A., Williams, J. V., Harrison, R. D., Agustiono,, Habibi,, et al. (2019). Accurate measurement of tropical forest canopy heights and aboveground carbon using structure from motion. *Remote Sens.* 11 (8), 928. doi: 10.3390/rs11080955

Tadele, Z. (2017). Raising crop productivity in Africa through intensification. *Agronomy* 7 (1), 22. doi: 10.3390/agronomy7010022

Tamene, L. D., and Ashenafi, A. (2022). Digital agriculture profile: Ethiopia. *Addis Ababa (Ethiopia): Alliance of Biodiversity International and International Center for Tropical Agriculture (CIAT)*. CIAT publication. 20 p. Available at: https://hdl.handle.net/10568/119309

Tariq, M., Ahmed, M., Iqbal, P., Fatima, Z., and Ahmad, S. (2020). Crop phenotyping. In: M. Ahmed (eds) *Systems modeling* (Singapore: Springer), 45–60. doi: 10.1007/978-981-15-4728-7_2

Tattaris, M., Reynolds, M. P., and Chapman, S. C. (2016). A direct comparison of remote sensing approaches for high-throughput phenotyping in plant breeding. *Front. Plant Sci.* 7 (AUG2016). doi: 10.3389/fpls.2016.01131

Tilling, A. K., O'Leary, G. J., Ferwerda, J. G., Jones, S. D., Fitzgerald, G. J., Rodriguez, D., et al. (2007). Remote sensing of nitrogen and water stress in wheat. *Field Crops Res.* 104 (1–3), 77–85. doi: 10.1016/j.fcr.2007.03.023

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America.* 108 (50), 20260–20264. doi: 10.1073/pnas.1116437108

Tripodi, P., Nicastro, N., Pane, C., and Cammarano, D. (2022). Digital applications and artificial intelligence in agriculture toward next-generation plant phenotyping. *Crop Pasture Sci.* 18. doi: 10.1071/CP21387

Van Dijk, M., Morley, T., Rau, M. L., and Saghai, Y. (2021). A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. *Nature Food* 2 (7), 494–501. doi: 10.1038/s43016-021-00322-9

Van Duivenbooden, N., Pala, M., Studer, C., Bielders, C. L., and Beukes, D. J. (2000). Cropping systems and crop complementarity in dryland agriculture to increase soil water use efficiency: a review. *NJAS: Wageningen J. Life Sci.* 48 (3), 213–236. doi: 10.1016/S1573-5214(00)80015-9

Varshney, R. K., Bohra, A., Yu, J., Graner, A., Zhang, Q., and Sorrells, M. E. (2021). Designing future crops: Genomics-assisted breeding comes of age. *Trends Plant Sci.* 26 (6), 631–649. doi: 10.1016/j.tplants.2021.03.010

Vergara-Díaz, O., Zaman-Allah, M. A., Masuka, B., Hornero, A., Zarco-Tejada, P., Prasanna, B. M., et al. (2016).). A novel remote sensing approach for prediction of maize yield under different conditions of nitrogen fertilization. *Front. Plant Sci.* 7 (MAY2016). doi: 10.3389/fpls.2016.00666

Verner, D., Treguer, D., Redwood, J., Christensen, J., McDonnell, R., Elbert, C., et al. (2018). *Climate variability, drought, and drought management in Morocco's agricultural sector* (Washington, D. C., USA: World BankWorld Bank). Available at: https://openknowledge.worldbank.org/handle/10986/30603.

Virlet, N., Lyra, D. H., and Hawkesford, M. J. (2022). Digital phenotyping and genotype-to-phenotype (G2P) models to predict complex traits in cereal crops. in: A Walter (ed.) *Advances in Plant Phenotyping for More Sustainable Crop Production*. (Cambridge: Burleigh Dodds Science Publishing), p. 40. doi: 10.19103/AS.2022.0102.12

Virlet, N., Sabermanesh, K., Sadeghi-Tehran, P., and Hawkesford, M. J. (2017). Field Scanalyzer: An automated robotic field phenotyping platform for detailed crop monitoring. *Funct. Plant Biol.* 44 (1), 143–153. doi: 10.1071/FP16163

Wahab, I. (2020). In-season plot area loss and implications for yield estimation in smallholder rainfed farming systems at the village level in Sub-Saharan Africa. *GeoJournal* 85 (6), 1553–1572. doi: 10.1007/s10708-019-10039-9

Wang, J., Wu, B., Kohnen, M. V., Lin, D., Yang, C., Wang, X., et al. (2021). Classification of rice yield using UAV-based hyperspectral imagery and lodging feature. *Plant Phenomics* 2021, 9765952. doi: 10.34133/2021/9765952

Watt, M., Fiorani, F., Usadel, B., Rascher, U., Muller, O., and Schurr, U. (2020). Phenotyping: New windows into the plant for breeders. *Annu. Rev. Plant Biol.* 71, 689–712. doi: 10.1146/annurev-arplant-042916-041124

Wellington, M. (2023). *Remote Sensing of Croplands, Crop Productivity, and Water use Efficiency with a Focus on Smallholder Systems in Southern Africa (Doctoral dissertation* (Australia: The Australian National University).

Whalen, K., and Yuhas, C. (2019). *Low-cost drone and sensor for agricultural applications on small farms in Tanzania*. Available at: https://globalwater.osu.edu/files/Drone-Final-Design-Report-1.pdf (Accessed 4 January 2023).

White, J. W., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., et al. (2012). Field-based phenomics for plant genetics research. *Field Crops Res.* 133, 101–112. doi: 10.1016/j.fcr.2012.04.003

White, J. W., and Conley, M. M. (2013). A flexible, low-cost cart for proximal sensing. *Crop Sci.* 53 (4), 1646–1649. doi: 10.2135/cropsci2013.01.0054

Woldeyohannes, A. B., Accotto, C., Desta, E. A., Kidane, Y. G., Fadda, C., Pè, M. E., et al. (2020). Current and projected eco-geographic adaptation and phenotypic diversity of Ethiopian teff (Eragrostis teff) across its cultivation range. *Agriculture Ecosyst. Environ.* 300, 107020. doi: 10.1016/j.agee.2020.107020

World Bank (2010). "Global strategy to improve agricultural and rural statistics," in *Economic sector work no. 56719-GLB* (Washington DC: The International Bank for Reconstruction and Development / The World Bank).

Xiao, Q., Bai, X., Zhang, C., and He, Y. (2022). Advanced high-throughput plant phenotyping techniques for genome-wide association studies: A review. *J. advanced Res.* 35, 215–230. doi: 10.1016/j.jare.2021.05.002

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 (2), 187–214. doi: 10.1016/j.molp.2020.01.008

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., et al. (2017a). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01111

Yang, Z., Shao, Y., Li, K., Liu, Q., Liu, L., and Brisco, B. (2017b). An improved scheme for rice phenology estimation based on time-series multispectral HJ-1A/B and polarimetric RADARSAT-2 data. *Remote Sens. Environ.* 195. doi: 10.1016/j.rse.2017.04.016

Yang, H., Yang, J., Lv, Y., and He, J. (2014). SPAD values and nitrogen nutrition index for the evaluation of rice nitrogen status. *Plant Production Sci.* 17 (1), 81–92. doi: 10.1626/pps.17.81

Zaman-Allah, M., Vergara, O., Araus, J. L., Tarekegne, A., Magorokosho, C., Zarco-Tejada, P. J., et al. (2015). Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. *Plant Methods* 11 (1), 1–10. doi: 10.1186/s13007-015-0078-2

Zendonadi dos Santos, N., Piepho, H. P., Condorelli, G. E., Licieri Groli, E., Newcomb, M., Ward, R., et al. (2021). High-throughput field phenotyping reveals genetic variation in photosynthetic traits in durum wheat under drought. *Plant Cell Environ.* 44 (9), 2858–2878. doi: 10.1111/pce.14136

Zhang, J., Yang, C., Song, H., Hoffmann, W. C., Zhang, D., and Zhang, G. (2016). Evaluation of an airborne remote sensing platform consisting of two consumer-grade cameras for crop identification. *Remote Sens.* 8 (3), 257. doi: 10.3390/rs8030257

Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu, S., et al. (2019). Crop phenomics: Current status and perspectives. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.00714

Zhao, Y., Zheng, B., Chapman, S. C., Laws, K., George-Jaeggli, B., Hammer, G. L., et al. (2021). Detecting sorghum plant and head features from multispectral UAV imagery. *Plant Phenomics* 2021. doi: 10.34133/2021/9874650

Zhu, Y., Yao, X., Tian, Y. C., Liu, X. J., and Cao, W. X. (2008). Analysis of common canopy vegetation indices for indicating leaf nitrogen accumulations in wheat and rice. *Int. J. Appl. Earth Observation Geoinformation* 10 (1), 1–10. doi: 10.1016/j.jag.2007.02.006

Zimbabwe Centre For High Performance Computing. (2021). *Zimbabwe centre for high performance computing*. Available at: https://zchpc.ac.zw/history/.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 November 2023
DOI 10.3389/fpls.2023.1260772

Check for updates

OPEN ACCESS

EDITED BY
Jiangang Liu,
Chinese Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Thulani Tshabalala,
University of KwaZulu-Natal, South Africa
Huaguo Huang,
Beijing Forestry University, China

*CORRESPONDENCE
Jianghua Zheng
zheng.jianghua@xju.edu.cn

RECEIVED 18 July 2023
ACCEPTED 27 October 2023
PUBLISHED 14 November 2023

CITATION
Tuerxun N, Zheng J, Wang R, Wang L and Liu L (2023) Hyperspectral estimation of chlorophyll content in jujube leaves: integration of derivative processing techniques and dimensionality reduction algorithms.
Front. Plant Sci. 14:1260772.
doi: 10.3389/fpls.2023.1260772

COPYRIGHT
© 2023 Tuerxun, Zheng, Wang, Wang and Liu. 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 content in jujube leaves: integration of derivative processing techniques and dimensionality reduction algorithms

Nigela Tuerxun[1,2], Jianghua Zheng[1,2]*, Renjun Wang[1,2], Lei Wang[3] and Liang Liu[1,2]

[1]College of Geography and Remote Sensing Sciences, Xinjiang University, Urumqi, China, [2]Xinjiang Key Laboratory of Oasis Ecology, Xinjiang University, Urumqi, China, [3]Institute of Modern Forestry, Xinjiang Academy of Forestry Sciences, Urumqi, China

The leaf chlorophyll content (LCC) of vegetation is closely related to photosynthetic efficiency and biological activity. Jujube (*Ziziphus jujuba* Mill.) is a traditional economic forest tree species. Non-destructive monitoring of LCC of jujube is of great significance for guiding agroforestry production and promoting ecological environment protection in arid and semi-arid lands. Hyperspectral data is an important data source for LCC detection. However, hyperspectral data consists of a multitude of bands and contains extensive information. As a result, certain bands may exhibit high correlation, leading to redundant spectral information. This redundancy can distort LCC prediction results and reduce accuracy. Therefore, it is crucial to select appropriate preprocessing methods and employ effective data mining techniques when analyzing hyperspectral data. This study aims to evaluate the performance of hyperspectral data for estimating LCC of jujube trees by integrating different derivative processing techniques with different dimensionality reduction algorithms. Hyperspectral reflectance data were obtained through simulations using an invertible forest reflectance model (INFORM) and measurements from jujube tree canopies. The least absolute shrinkage and selection operator (LASSO) and elastic net (EN) were employed to identify the important bands in the original spectra (OS), first derivative spectra (FD), and second derivative spectra (SD). Support vector regression (SVR) was used to establish the estimation model. The results show that compared with full-spectrum modeling, LASSO and EN algorithms are effective methods for preventing overfitting in LCC machine learning estimation models for different spectral derivatives. The LASSO/EN-based estimation models constructed using FD and SD exhibited superior R^2 compared to the OS. The important band of SD can best reveal the relevant information of jujube LCC, and SD-EN-SVR is the

most ideal model in both the simulated dataset (R^2 = 0.99, RMSE=0.61) and measured dataset (R^2 = 0.89, RMSE=0.91). Our results provided a reference for rapid and non-destructive estimation of the LCC of agroforestry vegetation using canopy hyperspectral data.

KEYWORDS

hyperspectral data, elastic net, LASSO, support vector regression, invertible forest reflectance model, derivative processing

1 Introduction

The leaf chlorophyll content (LCC) of vegetation is closely related to the absorption of carbon dioxide in the atmosphere and the process of photosynthesis, which is an indicator of the photosynthetic efficiency and biological activity of vegetation (Darvishzadeh et al., 2008). Conducting quantitative and real-time monitoring of chlorophyll content variations in forest could provide crucial information to understand the responses of ecosystems to changes in environmental, meteorological, and ecological factors (Zhen et al., 2021).

Jujube (*Ziziphus jujuba* Mill.) is a traditional economic forest tree species. Its fruit is sweet and juicy and is suitable for fresh food and dry processing. Jujube trees are known for their strong resistance to drought and wind, as well as their high yield and stability (Liu et al., 2020). As a result, they are becoming increasingly important in arid and semi-arid lands (Liu et al., 2020). Xinjiang produces half of the jujubes in China, and they are of higher quality compared to those planted in other regions (Bai et al., 2019). Non-destructive monitoring of LCC of jujube is of great significance for guiding agroforestry production and promoting ecological environment protection in arid and semi-arid lands.

Traditional laboratory LCC determinations are destructive and time consuming (Li et al., 2020). The development of remote sensing technology enables the acquisition of physical and chemical information of vegetation in a non-contact manner. Hyperspectral remote sensing technology is very effective for monitoring LCC due to its rapid and non-destructive capabilities (Shi et al., 2022). Compared with multispectral sensors, hyperspectral data, with its narrower bandwidth, provides distinct advantages for monitoring vegetation health by capturing different physical and chemical reactions in vegetation at various wavelengths (Jingguo et al., 2015; Ali and Imran, 2020). However, hyperspectral data contain much information and many bands; hence, some bands are highly correlated, which increases the redundancy of spectral information, leading to a distortion in prediction results and a reduction in prediction accuracy (Cheng et al., 2022). To address these issues, Sun et al. (2021) selected the first order derivative (FD) spectral data and using the correlation coefficient method to predict the LCC of maize. Lu and Peng (2015) calculated the correlation coefficient between vegetation index and chlorophyll concentration, finding that D715/D705 (D: first derivative), EBFR (simple ratios of the amplitude between the red and blue regions), D705/D722, and BND (normalized difference derivative at 722 and 700 nm) had a better estimation effect on chlorophyll concentration at the cherry leaf scale. Although the above methods have achieved better results, some limitations still exist. For instance, the method of setting the threshold based on the correlation between the band and chlorophyll content may ignore the collinearity of adjacent spectral data (Sun et al., 2021). The vegetation index method utilizes only one to four bands of information, which fails to fully capture the important information present in hyperspectral data (Lu and Peng, 2015). Studies have shown that the selection of important bands for modeling through dimensionality reduction algorithms typically yields equal or superior model prediction performance compared to full-spectrum models (Wang et al., 2022a; Zhu et al., 2022). The least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996) and elastic net (EN) (Zou and Hastie, 2005) are regularization methods that effectively reduce high-dimensional data by adjusting model parameters. These algorithms have demonstrated successful outcomes in various applications, including crop yield estimation (Cao et al., 2021b), leaf nitrogen estimation (Cao et al., 2021a) and forest biomass estimation (Takayama and Iwasaki, 2016). However, the potential of these two algorithms in estimating the LCC of agroforestry vegetation using hyperspectral data, such as jujube trees, remains unclear. FD and/or the second derivatives (SD) are commonly used spectral data preprocessing techniques (Wang et al., 2018b; Wang et al., 2021; Wang et al., 2022b). They are widely employed to mitigate noise, baseline effects, overlap problems, enhance spectral features, capture subtle details of spectral curves, and improve the accuracy of land surface parameter extractions (Li et al., 1993; Cui et al., 2022; Jin and Wang, 2022). However, to the best of our knowledge, there has been no research combining these two derivative processing techniques with LASSO and EN dimensionality reduction algorithms for predicting hyperspectral vegetation LCC.

Support vector regression (SVR), introduced by Cortes and Vapnik (Cortes and Vapnik, 1995) in 1995, is a versatile machine learning regression model. It has proven to enhance the efficiency of modeling vegetation physiological parameters while demonstrating improved stability in parameter estimation compared to other methods (Navarro et al., 2019; Liu et al., 2022). Therefore, this study aims to achieve the following objectives: (1) Propose a method for estimating LCC of agroforestry vegetation by integrating derivative processing techniques and dimensionality reduction algorithms, specifically utilizing FD and SD derivative processing along with LASSO and EN algorithms. (2) Compare the prediction performance of LCC using important spectral bands of different derivative orders (original spectra (OS), FD, SD) selected by LASSO and EN algorithms with the prediction performance of LCC based on

modeling using full-spectrum data by establish SVR. (3) Evaluate the effectiveness of the proposed method using measured data and a substantial dataset of canopy reflectance data generated by the Invertible Forest Reflectance Model (INFORM), which accurately represents the annual growth stages of jujube trees. This analysis will help assess the practicality and applicability of the proposed approach.

2 Materials and methods

2.1 Study area

Ruoqiang County is located in southeastern Xinjiang, the southeastern margin of the Taklimakan Desert, and the eastern Tarim Basin. It is located between 86°45'-93°45' E and 36°05'-41°23' N, with an altitude of 846-4500 m and a total area of 202,300 square kilometers. Ruoqiang County is the largest county in China. It has a warm temperate continental desert arid climate, providing unique natural conditions for the development of the agroforestry vegetation (Cui, 2019). Among them, the 'Huizao' variety of *Z. jujuba* Mill. is a well-known product in Xinjiang.

A total of 69 samples were collected in the study area. The minimum interval of sample points is more than 15m.The location of study area is shown in Figure 1.

2.2 Data collection and preprocessing

2.2.1 Measurement of hyperspectral reflectance

A portable spectroradiometer, the PSR-3500 manufactured by Spectral Evolution, USA, was used to measure the canopy spectra of 69 sample points (trees). The spectral reflectance data were obtained from May 19th to June 1st, 2020, during the flowering stage of jujube trees, between 11:00 and 17:00 Beijing time, under clear, windless, and cloudless conditions (Cui et al., 2022). The spectral range covered 350-2500 nm with a 1 nm interval, resulting in 2150 wavebands. The spectrometer probe was vertically downward during each measurement. This measurement process was repeated 5 times, and the average of the spectral curve from these repetitions was taken as the reflectance of the sample. A total of 345 spectral data were measured. To account for any variations in the ambient radiation, we employed the white reference panel to calibrate the instrument prior to each reflectance measurement. To reduce noise interference caused by atmospheric scattering and other reasons (Badola et al., 2022), only the 350-1050 nm band range is retained, and savitzky-golay smoothing with a second-order polynomial and window size of 5 (Cai et al., 2019) was applied to smooth and denoise the spectral data.

2.2.2 Measurement of leaf relative chlorophyll content

The soil–plant analyzer development (SPAD)-502 Plus portable chlorophyll meter(Konica Minolta, Japan) measures the transmittance in the red region (650 nm) and infrared region (940 nm) through the leaf, providing a correspondence value of chlorophyll content in three significant digit values (leaf relative chlorophyll content), thereby characterizing the chlorophyll content in leaves (Yang et al., 2021). The measurement of SPAD values was carried out concurrently with canopy spectral measurements, two leaves were taken from each jujube tree in the east, west, south, and north, as well as vertical ground directions,

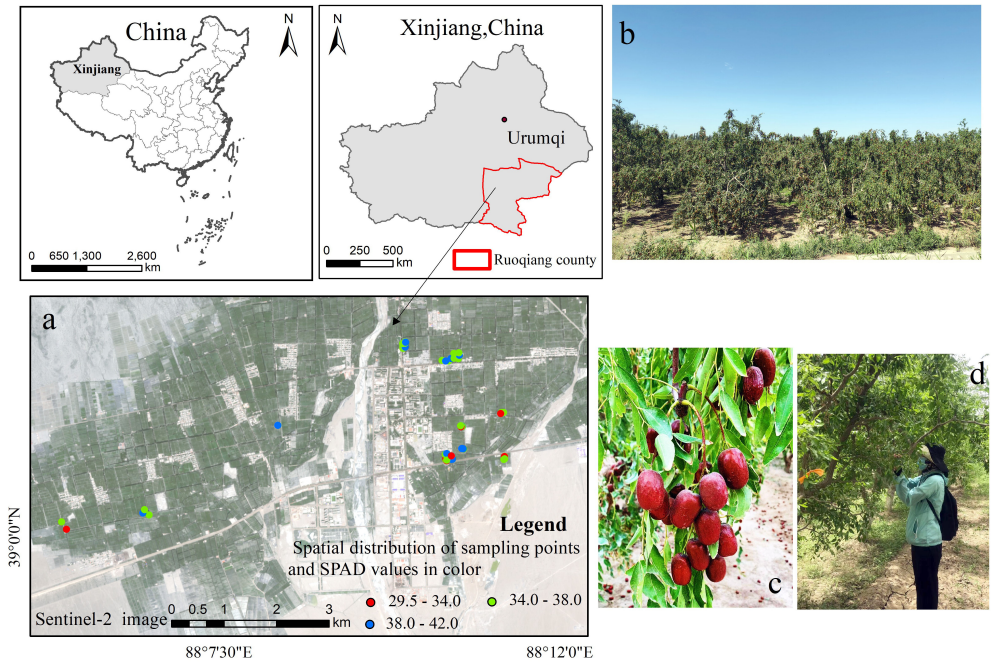


FIGURE 1
Study area locations: **(A)** sampling plots (SPAD value refers to leaf relative chlorophyll content); **(B)** jujube orchard; **(C)** mature state of the 'Huizao' variety of *Z. jujuba* Mill; **(D)** research team collecting jujube leaves.

resulting in a total of 10 leaves per tree. During the measurements, the leaf veins were avoided, and the SPAD value was recorded five times at different positions along the leaf, from the base to the tip. The average of the SPAD measurements for the 10 leaves was considered as the SPAD value for the sampled tree. The measurement processes of canopy spectral and SPAD value are illustrated in the data acquisition section of Figure 2.

Existing research (Zhang et al., 2022) has established and validated the formula (LCC(μg/cm^2))= 0.709 * SPAD – 1.576) for a robust conversion (R^2 = 0.52) of SPAD values to LCC. In this paper, SPAD values are utilized to characterize the LCC in the measurement dataset.

2.2.3 Simulation of hyperspectral reflectance in jujube tree canopies

A simulated hyperspectral dataset of the jujube canopy was generated using the Invertible Forest Reflectance Model (INFORM) (Atzberger, 2000), which combines the Forest Light Interaction Model (FLIM), Scattering by Arbitrary Inclined Leaves (SAILH), and PROSPECT model. The dataset was simulated using a range of input parameters listed in Table 1. Based on previous studies (Lu et al., 2022; Wu et al., 2023) on SPAD estimation of jujube trees at different growth stages and the existing conversion relationship between LCC and SPAD values, the Cab value was set to be 25-50 ug/cm^2 to represent the total growth stage of jujube. N represents the leaf structure parameter, and N is the number of compact layers specifying the average number of air/cell walls interfaces within the mesophyll (Jacquemoud et al., 2009). Usually, the N values of trees fall within the range of 0.63-3, and most studies use a fixed N value (Hernandez-Clemente et al., 2014; Yuan et al., 2015; Brown et al., 2019; Shi et al., 2022). As trees age, N values tend to increase. To ensure diversity in leaf sample types and statuses, this study has set the range for N values between 1 and 1.5. Scale factor for soil reflectance is the spectral reflectance of the underlying (uncovered) soil (Atzberger, 2000). Based on the empirical knowledge from our research team's field investigations, we have observed that the soil conditions in jujube orchards are quite complex. Typically, the soil in jujube orchards is not entirely covered by trees and other vegetation. On poorly managed land, there may even be situations where the soil is completely exposed. Drawing from previous studies (Hernandez-Clemente et al., 2014; Darvishzadeh et al., 2019; Zarco-Tejada et al., 2019) on other tree species, we have decided to set the scale factor for soil reflectance range between 0.6 and 1. This range represents variations in soil parameters during the growth of jujube trees. This parameter helps us more accurately simulate and describe the soil conditions in jujube orchards, enhancing the effectiveness of our simulation data. Other parameters (Single trees LAI (m^2·m^{-2}) and Average leaf angle of tree canopy) were set within a reasonable range according to the field investigation of the research team. Ultimately, a total of 2100 spectra were generated, covering all possible combinations of the input parameters.

2.3 Modeling process

In this study, estimation models were established by integrating OS, FD, and SD spectra, dimensionality reduction algorithms (LASSO, EN) and SVR. The modeling process was as follows: first, the SVR was used to model the OS, FD, and SD spectra for the full-spectrum prediction of LCC. Second, the LASSO and EN were employed to reduce the dimensions of the OS, FD and SD spectral data and SVR was performed on the important bands after dimension reduction. A total of nine estimation models were established. The hyperspectral simulated dataset obtained from the radiative transfer model INFORM and the measured dataset were used to test the validity of the proposed integration method. Modeling process of this research is shown in Figure 2.

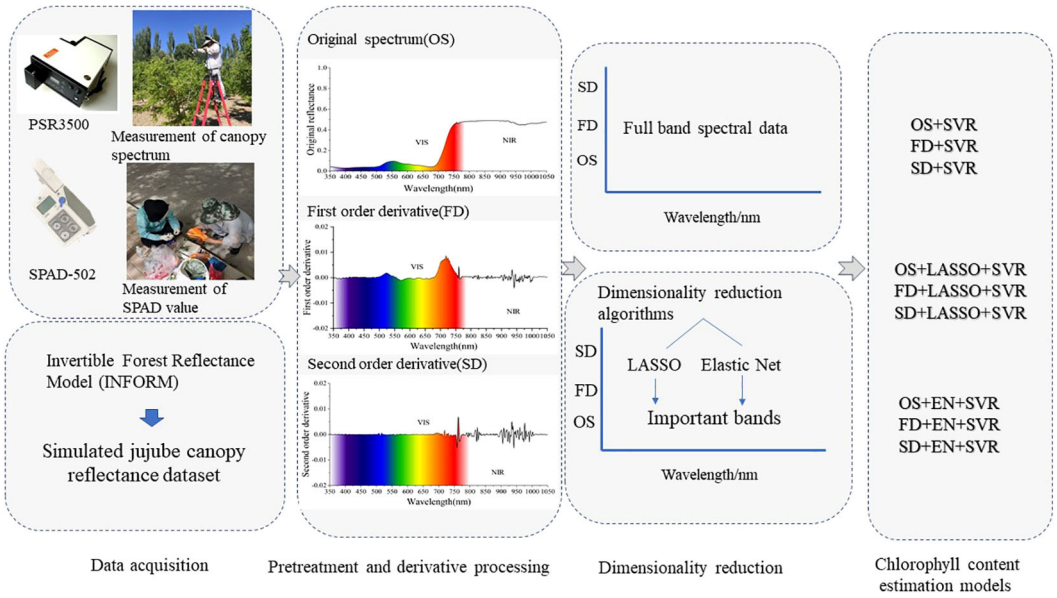


FIGURE 2
The diagram of the study.

TABLE 1 INFORM model parameters that were used in this study to simulate the canopy reflectance spectra.

Parameter	Minimum	Maximum	Step
Cab(ug/cm^2)	25	50	5
The leaf structure parameter(N)	1	1.5	0.5
Single trees LAI (m^2·m^{-2})	1	7	1
Scale factor for soil reflectance	0.6	1	0.1
Average leaf angle of tree canopy (degree)	0	60	15

(Other ten fixed parameters are: solar zenith angle = 30°; Observer zenith Angle = 0°; azimuth angle = 0°; fraction of diffuse incident radiation is 0.1; Equivalent water thickness is 0.03 g·cm^{-2}; Dry matter content is 0.012 g.cm-2; Stem density is 500 ha^{-1}; Tree height is 2.5m; Crown diameter is 2m; LAI understory is 0.1 m^2·m^{-2}).

2.4 Derivative processing

The utilization of derivative techniques has become prevalent in mitigating noise, baseline effects, overlap problems, enhancing spectral features, capturing subtle details of spectral curves, and improving the accuracy of land surface parameter extractions(Wang et al., 2018b). Not only is the derivative technique a potent tool for spectral analysis, it can also tackle collinearity concerns (Wang et al., 2021). In the present study, FD and SD were employed to process the hyperspectral data with the aid of Origin 2021 (OriginLab).

2.5 Dimensionality reduction algorithms

2.5.1 LASSO

LASSO, proposed by Robert Tibshirani in 1996 (Tibshirani, 1996), is a biased estimation method for compressing model coefficients and variable selection. LASSO adds the L1 norm penalty term on the basis of the least squares method to compress the estimated parameters. When the sum of the absolute values of the regression coefficients is less than a constant, the sum of squared residuals is minimized to obtain regression coefficients equal to 0; thus, the effects of independent variables with little or no influence are compressed to zero. The multiple linear model can be expressed as follows:

$$y = X\beta + \epsilon \tag{1}$$

where β is the linear variable, X is the independent variable (that is, the hyperspectral data), y is the dependent variable (that is, the jujube SPAD value), and ϵ is the error.

From Equation (1), the estimation of parameter β can be expressed as follows:

$$J(\beta) = \sum \ (y - X\beta)^2 \tag{2}$$

Compared to linear regression, in LASSO, an L1 penalty term is added as follows:

$$J(\beta) = \sum \ (y - X\beta)^2 + \lambda \parallel \beta \parallel_1 = \sum \ (y - X\beta)^2 + \sum \ \lambda|\beta| \tag{3}$$

where λ is a regularization parameter and $\lambda \geq 0$. The penalty strength of the model is related to the regulation parameter λ. Variable screening can be achieved by controlling the adjustment parameter λ

2.5.2 Elastic net

Zou and Hastie proposed the EN technique (Zou and Hastie, 2005), which integrates the characteristics of ridge and LASSO, and the penalty term has both an L1 norm term and an L2 norm term. The EN includes a mixture parameter α, which is selected based on the criterion of minimizing the MSE of the training samples and the MSE of the prediction bias. α is a number between 0 and 1 that acts in conjunction with λ to regulate the size of the penalty term. The estimation of parameter β can be expressed as follows:

$$\begin{aligned} J(\beta) &= \sum (y - X\beta)^2 + \alpha\lambda \parallel \beta \parallel_1 + \tfrac{1-\alpha}{2}\lambda \parallel \beta \parallel_2^2 \\ &= \sum (y - X\beta)^2 + \lambda \sum (\alpha|\beta| + (1-\alpha)\beta^2) \end{aligned} \tag{4}$$

It has been proven that the LASSO and EN models are more interpretable when using the value of λ with the minimum MSE than when using the value of λ with the minimum SE (Cao et al., 2021a). Therefore, the value of λ with the minimum MSE is chosen in this paper.

2.6 SVR

The support vector machine (SVM) theory proposed by Vapnik was initially used for supervised classification processes(Cortes and Vapnik, 1995). SVR is the regression method of SVM, the idea of SVR has been described by Smola and Schölkopf (Smola and Schölkopf, 2004). In SVR, the mapping of input data in higher-order feature space is accomplished by several types of kernel functions (Li et al., 2021), such as linear, nonlinear, sigmoid, polynomial, and radial basis functions (RBFs). Among various kernel functions, the RBF kernel can achieve good results. Therefore, we adopted the RBF kernel of SVR, where the hyperparameters (C and gamma) were used for cross-validated grid search, parameter tuning and model training in R software.

2.7 Model evaluation method

The determination coefficient (R^2), root mean square error (RMSE) are compared to evaluate and optimize the model accuracy. The higher the R^2 value is, the stronger the prediction ability of the model. The smaller the calculated values of RMSE, the higher the prediction accuracy of the model. The calculation formula is as follows:

$$R^2 = \left[\frac{\sum_{i=1}^{n}(y_i - \bar{y})(f_i - \bar{f})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 (f_i - \bar{f})^2}} \right]^2 \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2} \tag{6}$$

where n is the number of samples, $y_i(i = 1, \ldots, n)$ is the measured value, $f_i(i = 1, \ldots, n)$ represents the predicted value, $\bar{y}$ represents the average observed value, and $\bar{f}$ represents the average predicted value.

3 Results

3.1 Statistical description of measured and simulated dataset

The statistical analysis of measured dataset and simulated dataset are presented in Table 2. A total of 69 samples were collected in measured dataset; the mean SPAD value was 36.00, the range was 29.50 to 42, the median was 35.70, the interquartile range (Q3-Q1) was 3.1, the standard deviation was 2.69, and the coefficient of variation was 0.07. The 69 samples were randomly split into a training set and a validation set at a ratio of 75%:25%. The training set consisted of 52 samples, while the validation set comprised 17 samples.

A total of 2100 samples were collected in simulated dataset; the mean LCC was 37.5 ug/cm^2, the range was 25 ug/cm^{2}to 50 ug/cm^2, the median was 37.5 ug/cm^2, the interquartile range (Q3-Q1) was 15 ug/cm^2, the standard deviation was 8.54 ug/cm^2, and the coefficient of variation was 22.5 ug/cm^2. The coefficient of variation was calculated to be 0.07. The simulated dataset of 2100 hyperspectral data were randomly divided into validation set and training set in the same proportion.

3.2 Results of spectral dimensionality reduction

Figures 3A-C show the important band distribution of the OS, FD and SD spectra after dimension reduction by LASSO and EN. For the measurement data: There were 10 important bands selected in OS-LASSO, 7 important bands were selected in FD-LASSO, and 13 important bands were selected in SD-LASSO. The optimization function $J(\beta)$ of the EN contains coefficients α (0<α<1). In this paper, the range of α (0-1) is divided into 100 parts: the larger α is, the fewer variables are selected, and the smaller α is, the more variables are selected. The α values calculated for the OS, FD and SD spectra were 0.21, 0.94 and 0.52, respectively. Thus, 80, 11, and 30 bands were selected when using EN. The results of the simulated dataset show that 97 important bands are selected by the OS-LASSO method, 161 important bands are selected by FD-LASSO, and 31 important bands are selected by SD-LASSO. The α values calculated for the OS, FD and SD spectra were 0.50, 0.20 and 0.91, respectively. Thus, 172, 208, and 40 bands were selected when using EN.

3.3 Model building and evaluation

Table 3 outlines the detailed outcomes of measured dataset and simulated dataset generated by the SVR. The nine SVR results indicated that the SVR model based on SD-EN-SVR achieved the best quantitative prediction accuracy in both the measured (R^2 = 0.89, RMSE=0.91) and simulated datasets (R^2 = 0.99, RMSE=0.61). In addition, the model fitting effect based on SD and FD was better than the model fitting effect based on OS, indicating that the FD and SD processing is more effective than directly modeling with the OS. The scatter plot (Figure 4) of the measured versus predicted SPAD values demonstrates that SD-EN-SVR achieved the best estimation results with a fitting line close to the 1:1 line; therefore, this model has strong stability and high predictive ability. In contrast, the validation SPAD estimation deviated from the 1:1 line in the SVR model based on full spectra of OS, FD, and SD. It can be concluded that the SVR model was not suitable for processing high-dimensional data, and the result was overfitting. The combination of LASSO and EN with SVR significantly improved the overfitting phenomenon, and the R^2 of the validation set was no longer less than 0.1.

4 Discussion

4.1 Effect of important bands on SPAD values estimation

Hyperspectral data provide ground object reflection information in thousands of bands. Directly using machine learning method such as SVR to process the full-spectrum information may lead to overfitting of the model (Table 2; Figure 4). Improving the accuracy of prediction models while maintaining effective spectral information is a challenging problem. In this paper, LASSO, and EN were adopted to reduce the complexity of the model and prevent overfitting. The LASSO compresses the bands to a greater extent compared to EN, and the

TABLE 2 Statistical analysis of measured dataset and simulated dataset.

Dataset	N	Min	Q1	Median	Q3	Max	Mean	SD	CV
Measured dataset (SPAD)	69	29.50	34.60	35.70	37.70	42.00	36.00	2.69	0.07
Measured training dataset	52	29.50	34.10	35.30	37.30	42.00	35.60	2.80	0.08
Measured validation dataset	17	34.70	35.6	37.00	38.40	41.20	37.30	1.89	0.05
Simulated dataset (LCC)	2100	25.00	30.00	37.50	45.00	50.00	37.50	8.54	22.8
Simulated training dataset	1569	25.00	30.00	40.00	45.00	50.00	37.50	8.52	22.7
Simulated validation dataset	531	25	30.00	35.00	45.00	50.00	37.49	8.62	23.0

N represents the size of the dataset; SD indicates the standard deviation; CV stands for the coefficient of variation.

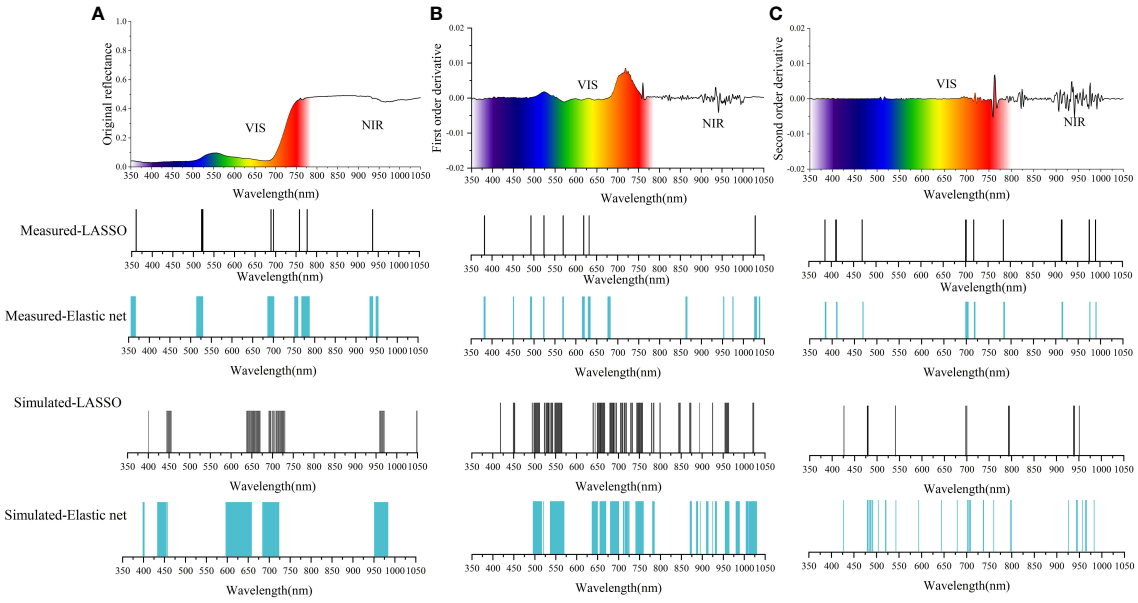


FIGURE 3
Distribution of important bands screened by the LASSO and EN: **(A)** selection based on OS; **(B)** selection based on FD; **(C)** selection based on SD.

EN is more moderate than the LASSO, and the selected bands are more uniform (Figure 3). It is worth noting that the model established by the bands selected by the EN is more stable and accurate than LASSO in LCC estimation. This may be because the EN combines the characteristics of ridge regression and LASSO, and the penalty term has both L1-norm and L2-norm terms. Furthermore, important bands selected based on the EN method were concentrated in the red-edge region (670-760 nm), defined as the boundary between chlorophyll absorption in the red and the onset of leaf scattering in the near-infrared light (Curran et al., 1990). Studies have proved that the red-edge bands are closely related with vegetation LCC and is more sensitive for detecting slight changes in LCC than that in the green region (500-560 nm) of the spectrum (Ju et al., 2010; Delegido et al., 2011; Li et al., 2016).

Figure 3 illustrates that the important bands obtained through LASSO and EN from the measurement dataset closely resemble those from the simulation dataset. Bands around 700 nm are selected in both datasets in the original spectrum. In the First Derivative (FD) spectrum, both datasets primarily select bands in the 500-550 nm and 600-700 nm ranges. In the FD spectrum, both datasets select most of the bands in the 500-550 nm and 600-700 nm ranges. In the SD spectrum, the spectra obtained from the measurement dataset and the simulation dataset are mainly distributed at 425 nm, 700 nm, and 900-975 nm. The method proposed in this study is relatively stable in the important bands selection of different datasets and the results of these important bands can provide a reference for wavelength selection in developing LCC detection equipment in the future.

TABLE 3 The evaluation of nine models (T=training set, V=validation set).

Model	Measured dataset				Simulated dataset			
	R^2_T	R^2_V	$RMSE_T$	$RMSE_v$	R^2_T	R^2_V	$RMSE_T$	$RMSE_v$
OS-SVR	0.83	0.00	1.84	2.53	0.77	0.02	8.67	8.76
FD-SVR	0.69	0.07	2.72	2.65	0.77	0.06	8.67	8.77
SD-SVR	0.74	0.06	2.28	2.58	0.98	0.01	2.39	8.61
OS-LASSO-SVR	0.48	0.41	2.03	1.95	0.79	0.78	4.94	5.14
FD-LASSO-SVR	0.70	0.64	1.52	1.65	0.94	0.93	2.22	2.47
SD-LASSO-SVR	0.72	0.66	1.56	1.56	0.95	0.94	1.90	2.20
OS-EN-SVR	0.74	0.45	1.44	2.10	0.88	0.87	3.26	3.47
FD-EN-SVR	0.72	0.70	1.47	1.60	0.97	0.97	1.43	1.60
SD-EN-SVR	0.89	0.71	0.91	1.53	0.99	0.97	0.61	1.07

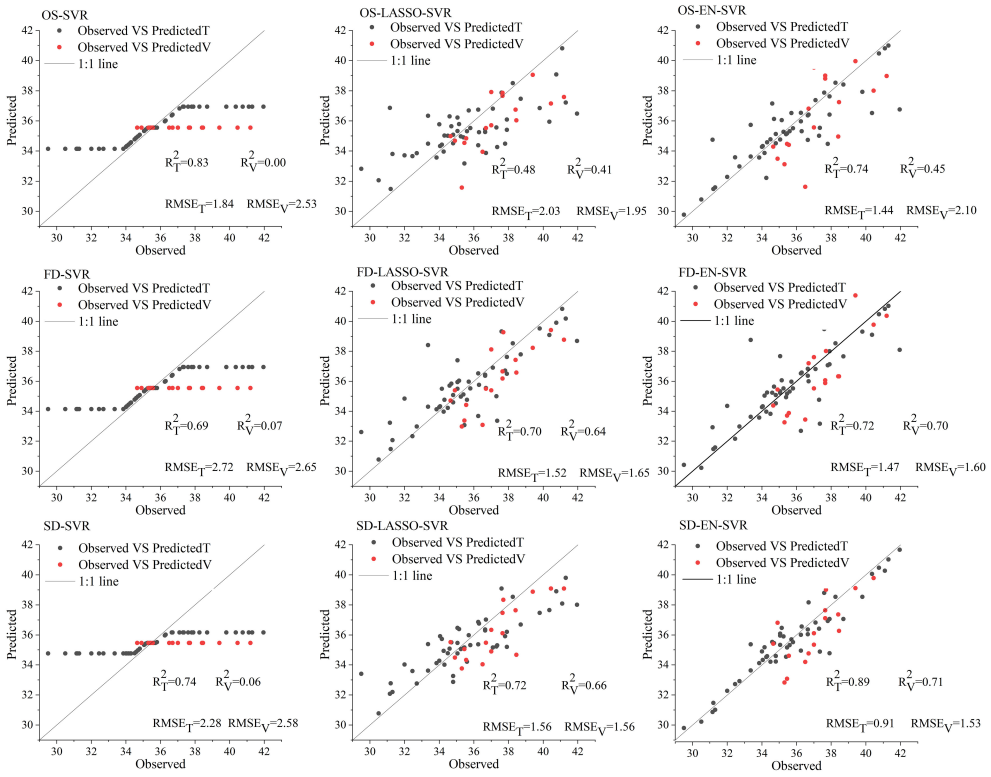


FIGURE 4
Scatterplots of the measured SPAD values vs. predicted SPAD values using SVR (T=training set, V=validation set).

4.2 Effect of derivative treatment

In previous studies, various methods, such as enhancement transform, curve smoothing, continuous curve removal, wavelet-based noise removal, have been commonly used to optimize hyperspectral data to improve the estimation accuracy of vegetation parameters (Wang et al., 2018a). Among them, derivative transformation was used as a robust mathematical analysis tool for processing data such as hyperspectral and remote sensing images because of its advantages of reducing noise and enhancing the details of data (Jin and Wang, 2016; Qu and Liu, 2017). This study proved that the combination of derivative spectral processing and dimensionality reduction algorithms effectively improved the estimation performance of LCC compared with the original spectral data modeling. In this study, FD and SD spectra were more robust than OS during modeling. Compared with OS modeling using the same combination of "EN/LASSO-SVR", the modeling accuracy after derivative processing is greatly improved (Figure 4). However, the FD spectra did not perform well compared to the SD spectra. This pattern occurs because the SD spectra enhance the peaks and valleys in the OS, distinguishing them from noise and enabling more accurate isolation and quantification of the vegetation signal than with the FD (Xiao-chen et al., 2008; Zhang et al., 2018). Additionally, the SD removes more of the soil background effects than the FD (Thorp et al., 2004), which can further improve the accuracy of LCC value estimation.

In Section 3.3, we presented the results of model building and evaluation, where the combination of LASSO and EN with SVR demonstrated relatively favorable performance. However, one notable issue that emerged was the consistent underestimation of accuracy on the validation set. From the model fitting results of the measured dataset presented in Figure 4, it can be observed that the LASSO/EN+SVR models built on FD and SD spectra exhibit a relatively better performance in mitigating underestimation of validation set compared to those built on OS spectra. This phenomenon may be attributed to the lower data quality of the OS spectral dataset, whereas FD and SD spectra are more reliable. This highlights the significance of employing FD and SD spectral preprocessing when utilizing LASSO/EN+SVR models. It can be seen from Table 3 that the problem of low accuracy in the validation set is greatly improved on the simulated data set. For instance, the difference in R^2_T and R^2_V between OS-LASSO-SVR, FD-LASSO-SVR, SD-LASSO-SVR, and OS-EN-SVR is only 0.1, while FD-EN-SVR achieves R^2_T and R^2_V values of 0.97. These findings suggest that the canopy information from the measured data is more intricate, potentially contain more information on soil and environmental factors. Consequently, the accuracy of the validation set may need to be moderately sacrificed to counteract the influence of complex environmental factors. This phenomenon emphasizes the importance of FD and SD derivative preprocessing in future research based on measured data.

4.3 Model uncertainty analysis

This paper presents an integration of derivative processing and dimensionality reduction algorithms method for estimating chlorophyll content in jujube leaves based on hyperspectral data and achieves good results in both measured and simulated datasets. The measured dataset in this study was collected during the flowering stage of jujube trees, during which the canopy reflectance of jujube trees was greatly affected by soil background and canopy structure (Yu et al., 2014). Despite these influences, the proposed combination method of derivative processing and dimensionality reduction algorithms in this paper still achieved favorable results during the flowering stage of jujube trees.

The chlorophyll content, leaf area index and other parameters (scale factor for soil reflectance and average leaf angle of tree canopy) of the simulated data have a wide range, which could represent the growth state of jujube during the whole growth stage. Therefore, the simulated dataset proves the validity of the integration method proposed in the study and generalizes the obtained results. This paper has contributed to the establishment of a prediction model of chlorophyll content in jujube leaves, but there are still the following limitations: (1) This study was based on hyperspectral data collected on the ground and simulated using the INFORM model, not combined with image data. Therefore, the effect of the results on the UAV and satellite scale needs to be verified. (2) The derivative processing method used in this study can be further optimized. The fractional derivative spectral data processing method has achieved good results in hyperspectral estimation of soil salinization (Wang et al., 2018b) and soil total nitrogen content (Yang et al., 2022), However, the effect of improving the estimation accuracy of chlorophyll content in combination with dimensionality reduction algorithms needs to be further explored in the future. (3) In this study, we did not conduct year-round destructive experiments to directly establish the conversion relationship between LCC and SPAD values. However, future research efforts will focus on conducting such experiments at various growth stages to improve the precision of LCC estimation. (4) In future research, we plan to collect data throughout the entire growth stage of jujube trees, taking into account variations in soil background and canopy structure at different stages.

5 Conclusion

In this study, we combined the derivative processing techniques and dimensionality reduction algorithms to improve the hyperspectral estimation of jujube LCC. The main results were as follows: (1) LASSO and EN algorithms are effective methods for preventing overfitting in LCC machine learning estimation models for different spectral derivatives. (2) The LASSO/EN-based estimation models constructed using FD and SD exhibited superior R^2 compared to the OS. (3) The important bands of the SD can effectively reveal the relevant information of LCC. In both the dataset simulated by the radiative transfer model INFORM, which represents the canopy reflectance of jujube trees throughout the entire growth stage (R^2 = 0.99, RMSE=0.61), and the measured dataset collected during the flowering stage of jujube with the interference from soil background and canopy structure (R^2 = 0.89, RMSE=0.91), the SD-EN-SVR model demonstrates the highest performance and is considered the most optimal model. This study provides a convenient method to estimate agroforestry vegetation parameters from canopy hyperspectral data and can provide a scale conversion reference for the LCC estimation of UAV and satellite remote sensing.

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

NT: Writing – original draft, Writing – review & editing. JZ: Conceptualization, Supervision, Writing – review & editing. RW: Formal Analysis, Investigation, Software, Writing – review & editing. LW: Software, Writing – review & editing. LL: Methodology, Software, 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 remote sensing monitoring project of typical tree species distribution in Xinjiang characteristic forestry and fruit industry 201904381000. The 'Tianshan cedar' project in Xinjiang (2020XSO4), and Biomass estimation of Typical Economic Forest in Xinjiang Based on Multi-source Remote Sensing Data (XJU2022BS055).

Acknowledgments

The authors would like to thank the local government of Xinjiang Bazhou Ruoqiang County and Mr. Hu Liangliang for the specific help with the field work investigation and support of the project work. Thanks very much to Jianglong Shen, Risi Ba, Yuan Fan, Fang Dong, Chuqiao Han for their support in the field work.

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

Ali, A., and Imran, M. (2020). Scientia Horticulturae Evaluating the potential of red edge position (REP) of hyperspectral remote sensing data for real time estimation of LAI & chlorophyll content of kinnow mandarin (Citrus reticulata) fruit orchards. *Sci. Hortic. (Amsterdam).* 267, 109326. doi: 10.1016/j.scienta.2020.109326

Atzberger, C. (2000). Development of an invertible forest reflectance model. The INFORM-Model. *A Decad. trans-European Remote Sens. Coop. Proc. 20th EARSeL Symp.* 14, 39–44.

Badola, A., Panda, S. K., Roberts, D. A., Waigl, C. F., Jandt, R. R., and Bhatt, U. S. (2022). International Journal of Applied Earth Observations and Geoinformation A novel method to simulate AVIRIS-NG hyperspectral image from Sentinel-2 image for improved vegetation / wildfire fuel mapping , boreal Alaska. *Int. J. Appl. Earth Obs. Geoinf.* 112, 102891. doi: 10.1016/j.jag.2022.102891

Bai, T., Zhang, N., Mercatoris, B., and Chen, Y. (2019). Jujube yield prediction method combining Landsat 8 Vegetation Index and the phenological length. *Comput. Electron. Agric.* 162, 1011–1027. doi: 10.1016/j.compag.2019.05.035

Brown, L. A., Ogutu, B. O., and Dash, J. (2019). Estimating forest leaf area index and canopy chlorophyll content with Sentinel-2: An evaluation of two hybrid retrieval algorithms. *Remote Sens.* 11, 1752. doi: 10.3390/rs11151752

Cai, Y., Lin, H., and Zhang, M. (2019). Mapping paddy rice by the object-based random forest method using time series Sentinel-1/Sentinel-2 data. *Adv. Sp. Res.* 64, 2233–2244. doi: 10.1016/j.asr.2019.08.042

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

Cheng, J., Sun, J., Yao, K., Xu, M., and Cao, Y. (2022). A variable selection method based on mutual information and variance inflation factor. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 268, 120652. doi: 10.1016/j.saa.2021.120652

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *1 Introduction. Mach. Learn.* 297, 273–297. doi: 10.1007/BF00994018

Cui, H. (2019). Research on the Brand Crisis of "Ruoqiang Jujube" Based on Brand Location Theory. in *Proceedings of the 3rd International Seminar on Education Innovation and Economic Management (SEIEM 2018)* (Atlantis Press) 286, 413–416. doi: 10.2991/seiem-18.2019.106

Cui, S., Zhou, K., Ding, R., Cheng, Y., and Jiang, G. (2022). Estimation of soil copper content based on fractional-order derivative spectroscopy and spectral characteristic band selection. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 275, 121190. doi: 10.1016/j.saa.2022.121190

Curran, P. J., Dungan, J. L., and Gholz, H. L. (1990). Exploring the relationship between reflectance red edge and chlorophyll content in slash pine. *Tree Physiol.* 7, 33–48. doi: 10.1093/treephys/7.1-2-3-4.33

Darvishzadeh, R., Skidmore, A., Abdullah, H., Cherenet, E., Ali, A., Wang, T., et al. (2019). Mapping leaf chlorophyll content from Sentinel-2 and RapidEye data in spruce stands using the invertible forest reflectance model. *Int. J. Appl. Earth Obs. Geoinf.* 79, 58–70. doi: 10.1016/j.jag.2019.03.003

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F., and Cho, M. (2008). LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS J. Photogramm. Remote Sens.* 63, 409–426. doi: 10.1016/j.isprsjprs.2008.01.001

Delegido, J., Verrelst, J., Alonso, L., and Moreno, J. (2011). Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors* 11, 7063–7081. doi: 10.3390/s110707063

Hernandez-Clemente, R., Navarro-Cerrillo, R. M., and Zarco-Tejada, P. J. (2014). Deriving predictive relationships of carotenoid content at the canopy level in a conifer forest using hyperspectral imagery and model simulation. *IEEE Trans. Geosci. Remote Sens.* 52, 5206–5217. doi: 10.1109/TGRS.2013.2287304

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., et al. (2009). PROSPECT + SAIL models: A review of use for vegetation characterization. *Remote Sens. Environ.* 113, S56–S66. doi: 10.1016/j.rse.2008.01.026

Jin, J., and Wang, Q. (2016). Hyperspectral indices based on first derivative spectra closely trace canopy transpiration in a desert plant. *Ecol. Inform.* 35, 1–8. doi: 10.1016/j.ecoinf.2016.06.004

Jin, J., and Wang, Q. (2022). Hyperspectral indices developed from the low order fractional derivative spectra can capture leaf dry matter content across a variety of species better. *Agric. For. Meteorol.* 322, 109007. doi: 10.1016/j.agrformet.2022.109007

Jingguo, T., Shudong, W., Lifu, Z., Taixia, W., Xiaojun, S., and Hailing, J. (2015). Evaluating different vegetation index for estimating lai of winter wheat using hyperspectral remote sensing data. *Work. Hyperspectral Image Signal Process. Evol. Remote Sens* 2015-June, 1–4. doi: 10.1109/WHISPERS.2015.8075437

Ju, C. H., Tian, Y. C., Yao, X., Cao, W. X., Zhu, Y., and Hannaway, D. (2010). Estimating leaf chlorophyll content using red edge parameters. *Pedosphere* 20, 633–644. doi: 10.1016/S1002-0160(10)60053-7

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

Li, L., Ren, T., Ma, Y., Wei, Q., Wang, S., Li, X., et al. (2016). Evaluating chlorophyll density in winter oilseed rape (Brassica napus L.) using canopy hyperspectral red-edge parameters. *Comput. Electron. Agric.* 126, 21–31. doi: 10.1016/j.compag.2016.05.008

Li, Y., Demetriades-Shah, T. H., Kanemasu, E. T., Shultis, J. K., and Kirkham, M. B. (1993). Use of second derivatives of canopy reflectance for monitoring prairie vegetation over different soil backgrounds. *Remote Sens. Environ.* 44, 81–87. doi: 10.1016/0034-4257(93)90104-6

Li, Y., Yang, K., Gao, W., Han, Q., and Zhang, J. (2021). A spectral characteristic analysis method for distinguishing heavy metal pollution in crops: VMD-PCA-SVM. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 255, 119649. doi: 10.1016/j.saa.2021.119649

Liu, Y., Wang, S., Wang, X., Chen, B., Chen, J., Wang, J., et al. (2022). Exploring the superiority of solar-induced chlorophyll fluorescence data in predicting wheat yield using machine learning and deep learning methods. *Comput. Electron. Agric.* 192, 106612. doi: 10.1016/j.compag.2021.106612

Liu, M., Wang, J., Wang, L., Liu, P., Zhao, J., Zhao, Z., et al. (2020). The historical and current research progress on jujube–a superfruit for the future. *Hortic. Res.* 7, 119. doi: 10.1038/s41438-020-00346-5

Lu, X., and Peng, H. (2015). Predicting cherry leaf chlorophyll concentrations based on foliar reflectance spectra variables. *J. Indian Soc Remote Sens.* 43, 109–120. doi: 10.1007/s12524-014-0397-1

Lu, J., Qiu, H., Zhang, Q., Lan, Y., Wang, P., Wu, Y., et al. (2022). Inversion of chlorophyll content under the stress of leaf mite for jujube based on model PSO-ELM method. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1009630

Navarro, J. A., Algeet, N., Fernández-Landa, A., Esteban, J., Rodríguez-Noriega, P., and Guillén-Climent, M. L. (2019). Integration of UAV, Sentinel-1, and Sentinel-2 data for mangrove plantation aboveground biomass monitoring in Senegal. *Remote Sens.* 11, 1–23. doi: 10.3390/rs11010077

Qu, Y., and Liu, Z. (2017). Dimensionality reduction and derivative spectral feature optimization for hyperspectral target recognition. *Optik (Stuttg).* 130, 1349–1357. doi: 10.1016/j.ijleo.2016.11.143

Shi, S., Xu, L., Gong, W., Chen, B., Chen, B., Qu, F., et al. (2022). A convolution neural network for forest leaf chlorophyll and carotenoid estimation using hyperspectral reflectance. *Int. J. Appl. Earth Obs. Geoinf.* 108. doi: 10.1016/j.jag.2022.102719

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

Sun, J., Yang, L., Yang, X., Wei, J., Li, L., Guo, E., et al. (2021). Using spectral reflectance to estimate the leaf chlorophyll content of maize inoculated with arbuscular mycorrhizal fungi under water stress. *Front. Plant Sci.* 12, 1–12. doi: 10.3389/fpls.2021.646173

Takayama, T., and Iwasaki, A. (2016). Optimal wavelength selection on hyperspectral data with fused lasso for biomass estimation of tropical rain forest. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* III–8, 101–108. doi: 10.5194/isprsannals-iii-8-101-2016

Thorp, K. R., Tian, L., Yao, H., and Tang, L. (2004). N -b d -b v i h d. 47, 291–299.

Tibshirani, R. (1996). Regression shrinkage and selection *via* the lasso. *J. R. Stat. Soc Ser. B* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

Wang, T., Gao, M., Cao, C., You, J., Zhang, X., and Shen, L. (2022a). 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

Wang, F., Gao, J., and Zha, Y. (2018a). Hyperspectral sensing of heavy metals in soil and vegetation: Feasibility and challenges. *ISPRS J. Photogramm. Remote Sens.* 136, 73–84. doi: 10.1016/j.isprsjprs.2017.12.003

Wang, T., Li, G., and Dai, C. (2022b). Soluble Solids Content prediction for Korla fragrant pears using hyperspectral imaging and GsMIA. *Infrared Phys. Technol.* 123, 104119. doi: 10.1016/j.infrared.2022.104119

Wang, X., Zhang, F., Kung, H., and Johnson, V. C. (2018b). New methods for improving the remote sensing estimation of soil organic matter content (SOMC) in the Ebinur Lake Wetland National Nature Reserve (ELWNNR) in northwest China. *Remote Sens. Environ.* 218, 104–118. doi: 10.1016/j.rse.2018.09.020

Wang, Z., Zhang, F., Zhang, X., Chan, N. W., Kung, H., Ariken, M., et al. (2021). Regional suitability prediction of soil salinization based on remote-sensing derivatives and optimal spectral index. *Sci. Total Environ.* 775, 145807. doi: 10.1016/j.scitotenv.2021.145807

Wu, Y., Zhao, Q., Yin, X., Wang, Y., and Tian, W. (2023). Multi-parameter health assessment of jujube trees based on unmanned aerial vehicle hyperspectral remote sensing. *Agriculture* 13, 1679. doi: 10.3390/agriculture13091679

Xiao-chen, L., Wen-jie, F., Qing-jiu, T., and Xi-ru, X. (2008). The lai inversion based on directional second derivative using hyperspectral data. *Int. Geosci. Remote Sens. Symp.* 3, 740–743. doi: 10.1109/IGARSS.2008.4779454

Yang, C., Feng, M., Song, L., Jing, B., Xie, Y., Wang, C., et al. (2022). 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, X., Yang, R., Ye, Y., Yuan, Z., Wang, D., and Hua, K. (2021). International Journal of Applied Earth Observations and Geoinformation 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

Yu, K., Lenz-wiedemann, V., Chen, X., and Bareth, G. (2014). ISPRS Journal of Photogrammetry and Remote Sensing Estimating leaf chlorophyll of barley at different growth stages using spectral indices to reduce soil background and canopy structure effects. *ISPRS J. Photogramm. Remote Sens.* 97, 58–77. doi: 10.1016/j.isprsjprs.2014.08.005

Yuan, H., Ma, R., Atzberger, C., Li, F., Loiselle, S. A., and Luo, J. (2015). Estimating forest fAPAR from multispectral Landsat-8 data using the Invertible Forest Reflectance Model INFORM. *Remote Sens.* 7, 7425–7446. doi: 10.3390/rs70607425

Zarco-Tejada, P. J., Hornero, A., Beck, P. S. A., Kattenborn, T., Kempeneers, P., and Hernández-Clemente, R. (2019). Chlorophyll content estimation in an open-canopy conifer forest with Sentinel-2A and hyperspectral imagery in the context of forest decline. *Remote Sens. Environ.* 223, 320–335. doi: 10.1016/j.rse.2019.01.031

Zhang, C., Liu, F., and He, Y. (2018). Identification of coffee bean varieties using hyperspectral imaging: Influence of preprocessing methods and pixel-wise spectra analysis. *Sci. Rep.* 8, 1–11. doi: 10.1038/s41598-018-20270-y

Zhang, R., Yang, P., Liu, S., Wang, C., and Liu, J. (2022). Evaluation of the methods for estimating leaf chlorophyll content with SPAD chlorophyll meters. *Remote Sens.* 14, 5144. doi: 10.3390/rs14205144

Zhen, J., Jiang, X., Xu, Y., Miao, J., Zhao, D., Wang, J., et al. (2021). Mapping leaf chlorophyll content of mangrove forests with Sentinel-2 images of four periods. *Int. J. Appl. Earth Obs. Geoinf.* 102, 102387. doi: 10.1016/j.jag.2021.102387

Zhu, C., Ding, J., Zhang, Z., and Wang, Z. (2022). Exploring the potential of UAV hyperspectral image for estimating soil salinity: Effects of optimal band combination algorithm and random forest. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* 279, 121416. doi: 10.1016/j.saa.2022.121416

Zou, H., and Hastie, T. (2005). Regularization and variable selection *via* the elastic net. *J. R. Stat. Soc Ser. B Stat. Methodol.* 67, 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 19 December 2023
DOI 10.3389/fpls.2023.1284235

Check for updates

OPEN ACCESS

EDITED BY
Zhenjiang Zhou,
Zhejiang University, China

REVIEWED BY
Hengbiao Zheng,
Nanjing Agricultural University, China
Haikuan Feng,
Beijing Research Center for Information Technology in Agriculture, China

*CORRESPONDENCE
Xinwei Li
lixw@ahstu.edu.cn
Wenhui Wang
1172139@lfnu.edu.cn

†These authors have contributed equally to this work

RECEIVED 28 August 2023
ACCEPTED 04 December 2023
PUBLISHED 19 December 2023

CITATION
Liu J, Zhu Y, Song L, Su X, Li J, Zheng J, Zhu X, Ren L, Wang W and Li X (2023) Optimizing window size and directional parameters of GLCM texture features for estimating rice AGB based on UAVs multispectral imagery.
Front. Plant Sci. 14:1284235.
doi: 10.3389/fpls.2023.1284235

COPYRIGHT
© 2023 Liu, Zhu, Song, Su, Li, Zheng, Zhu, Ren, Wang 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.

Optimizing window size and directional parameters of GLCM texture features for estimating rice AGB based on UAVs multispectral imagery

Jikai Liu[1,2†], Yongji Zhu[1†], Lijuan Song[3,4], Xiangxiang Su[1], Jun Li[1], Jing Zheng[5], Xueqing Zhu[1], Lantian Ren[2,6], Wenhui Wang[5*] and Xinwei Li[1,2*]

[1]College of Resource and Environment, Anhui Science and Technology University, Chuzhou, Anhui, China, [2]Anhui Province Crop Intelligent Planting and Processing Technology Engineering Research Center, Anhui Science and Technology University, Chuzhou, Anhui, China, [3]Institute of Agricultural Remote Sensing and Information, Heilongjiang Academy of Agricultural Sciences, Harbin, Heilongjiang, China, [4]School of Management, Heilongjiang University of Science and Technology, Harbin, Heilongjiang, China, [5]College of Life Science, Langfang Normal University, Langfang, Hebei, China, [6]College of Agriculture, Anhui Science and Technology University, Chuzhou, Anhui, China

Aboveground biomass (AGB) is a crucial physiological parameter for monitoring crop growth, assessing nutrient status, and predicting yield. Texture features (TFs) derived from remote sensing images have been proven to be crucial for estimating crops AGB, which can effectively address the issue of low accuracy in AGB estimation solely based on spectral information. TFs exhibit sensitivity to the size of the moving window and directional parameters, resulting in a substantial impact on AGB estimation. However, few studies systematically assessed the effects of moving window and directional parameters for TFs extraction on rice AGB estimation. To this end, this study used Unmanned aerial vehicles (UAVs) to acquire multispectral imagery during crucial growth stages of rice and evaluated the performance of TFs derived with different grey level co-occurrence matrix (GLCM) parameters by random forest (RF) regression model. Meanwhile, we analyzed the importance of TFs under the optimal parameter settings. The results indicated that: (1) the appropriate window size for extracting TFs varies with the growth stages of rice plant, wherein a small-scale window demonstrates advantages during the early growth stages, while the opposite holds during the later growth stages; (2) TFs derived from 45° direction represent the optimal choice for estimating rice AGB. During the four crucial growth stages, this selection improved performance in AGB estimation with R^2 = 0.76 to 0.83 and rRMSE = 13.62% to 21.33%. Furthermore, the estimation accuracy for the entire growth season is R^2 =0.84 and rRMSE =21.07%. However, there is no consensus regarding the selection of the worst TFs computation direction; (3) Correlation (Cor), Mean, and Homogeneity (Hom) from the first principal component image reflecting internal information of rice plant and Contrast (Con), Dissimilarity (Dis), and Second Moment (SM) from the second principal component image expressing edge texture are more important to estimate rice AGB among the whole growth stages; and (4) Considering the optimal parameters, the accuracy of texture-based AGB estimation slightly

outperforms the estimation accuracy based on spectral reflectance alone. In summary, the present study can help researchers confident use of GLCM-based TFs to enhance the estimation accuracy of physiological and biochemical parameters of crops.

KEYWORDS

unmanned aerial vehicles (UAVs), aboveground biomass (AGB), multispectral imagery, texture features (TFs), grey level co-occurrence matrix (GLCM), rice

1 Introduction

Rice (Oryza sativa) serves as a vital staple crop, nourishing almost half of the global population (Seck et al., 2012). In China, as the largest rice producer, consumer, and importer globally, rice constitutes a staple food for approximately 65% of the population (Peng et al., 2009). Accurate and timely prediction of rice yield holds significance in stabilizing rice prices, enhancing global food security, and enabling decision-makers to formulate timely strategies for rice import and export (Spiertz and Ewert, 2009; Cao et al., 2021). Aboveground biomass (AGB), the total dry matter per unit area of land during a specific period, plays a crucial role in understanding crop growth and development. AGB is an essential agronomic parameter to describe crop growth and nutritional status, making it instrumental in predicting crop yield (Li et al., 2020; Li et al., 2022). Thus, rapid and non-destructive assessment of the spatiotemporal dynamics of crop AGB is necessary for formulation and implementation of decision management and yield prediction in the field.

Although the field destructive sampling method is highly accurate, the inherent limitations in terms of time consumption and inefficiency render them inadequate for the timely, rapid, and large-scale requirements of modern agricultural monitoring (Fu et al., 2021). In contrast, the rapid development of remote sensing has presented new opportunities for non-destructive monitoring of AGB information in crop fields (Lu et al., 2019). Multispectral or hyperspectral images acquired from satellites or near-ground platforms offer a rapid means of monitoring AGB in dynamic spatiotemporal contexts. These platforms have been widely used to monitor various crop information, including wheat (Bao et al., 2009; Fu et al., 2014; Teng et al., 2015), rice (Gnyp et al., 2014; Cheng et al., 2017; Alebele et al., 2020), corn (Liao et al., 2019; Naidoo et al., 2021), and rapeseed (Mercier et al., 2020a; Mercier et al., 2020b). However, the continuous and prolonged monitoring of AGB has limitations due to the unavailability of satellite data during critical growth stages caused by weather variability and longer data return cycles (Gan et al., 2023). Additionally, the relatively low spatial resolution of satellite data hampers the acquisition of precise details about the crop canopy (Jin et al., 2020). On the contrary, near-ground platforms, which emphasize plant-to-sensor and sensor-to-plant models, provide certain advantages, such as continuous monitoring and payload availability. However, the limitations in terms of monitoring throughput and data acquisition scale impede their ability to capture crop AGB information at large scales (Lakhiar et al., 2018; Qiu et al., 2018; Jangra et al., 2021). The Unmanned aerial vehicles (UAVs) platform is the latest hot topic for research on crop phenotypes. It can be equipped with multi-source sensors to acquire multi-source remote sensing information, making it an effective supplement to both satellite and near-ground remote sensing platforms. UAVs provide high-precision data support and efficient monitoring capabilities, facilitating non-destructive monitoring of crop AGB (Maes and Steppe, 2019; Mukherjee et al., 2019).

Extensive research has been conducted utilizing spectral information from UAVs imagery to monitor crop AGB, and it has achieved acceptable monitoring accuracy (Han et al., 2019; Maimaitijiang et al., 2019; Wang F. et al., 2022; Wang et al., 2023). However, spectral information is susceptible to interference from water-soil (and weed) background noise during the early stages, while the dense canopy coverage at the later stages can result in spectral saturation (Li et al., 2022; Wang F. et al., 2022; Wang Q. et al., 2022; Wang W. et al., 2022; Zhu et al., 2023). These factors impose constraints on further advancements in monitoring accuracy. To address the issue of low accuracy in monitoring AGB based on spectral information, an increasing number of researchers are inclined toward utilizing texture features (TFs) to improve the precision of crop AGB monitoring (Liu Y. et al., 2019; Yue et al., 2019; Wang F. et al., 2022; Xu T. et al., 2022). TFs describe the frequency of variations in attribute values among adjacent pixel pairs within a specific window (Liu Y. et al., 2019; Yue et al., 2019;

Abbreviations: AGB, aboveground biomass; UAVs, unmanned aerial vehicles; TFs, texture features; GLCM, gray-level co-occurrence matrix; RF, random forest; R^2, coefficient of determination; RMSE, root mean squared error; rRMSE, relative root mean squared error; PCA, principal component analysis; PC1, first principal component of PCA; PC2, second principal component of PCA; D1, 0°; D2, 45°; D3, 90°; D4, 135°; ND, non-directional; P4M, DJI Phantom 4 Multispectral; B, blue; G, green; R, red; RE, red edge; NIR, near-infrared; LT, Late Tillering; B, Booting; HtF, Heading to Flowering; EF, Early Filling; Mean, Mean; Var, Variance; Hom, Homogeneity; Con, Contrast; Dis, Dissimilarity; Ent, Entropy; SM, Second Moment; Cor, Correlation.

Zhang et al., 2021). They can provide complementary information regarding the spatial arrangement and patterns of crop canopy, which helps overcome the limitations of spectral information (Liu Y. et al., 2019; Zhang et al., 2021). The color differences in crop leaves caused by factors such as crop varieties and soil nutrients can be captured by TFs (Zhang et al., 2021; Wang F. et al., 2022). Furthermore, due to crops undergoing growth and development, changes in canopy structure also generate texture variations. For instance, the emergence of rice panicles from leaf sheaths, the flowering of spikelets, the gradual drooping of mature panicles, and variations in the proportion between rice plant and background or shadows all contribute to subtle changes in canopy morphology. These subtle variations in canopy structure inevitably result in alterations in TFs (Yue et al., 2019; Xu L. et al., 2022). Incorporating TFs into the analysis provides a robust means of capturing subtle variations and improving the accuracy of crop AGB monitoring in agricultural remote sensing studies. Many studies utilize the gray-level co-occurrence matrix (GLCM) to extract TFs from UAVs images (Lu, 2005; Dube and Mutanga, 2015; Liu et al., 2018; Liang et al., 2022; Xu L. et al., 2022). By incorporating GLCM-based TFs into AGB monitoring, researchers have improved the accuracy of forest and crop biomass estimation. For example, Kelsey and Neff (Kelsey and Neff, 2014) discovered that forest AGB estimation models incorporating TFs exhibited greater accuracy compared to models relying solely on spectral information. Zheng et al. (2019) demonstrated that the incorporation of TFs can significantly enhance the monitoring accuracy of rice AGB, particularly during the mid to late stages of the rice growth season.

Existing research has demonstrated that window size and direction texture parameters are highly sensitive to texture metrics when extracting TFs using the GLCM (Zheng et al., 2020; Liang et al., 2022; Zhou M. et al., 2023). Throughout the entire growth season, the spatiotemporal fluctuations in the rice canopy coverage determine the requisite selection of appropriate window size for quantifying texture disparities. Furthermore, as rice is a row-cropped crop with an apparent spatial direction, the directional selection may impinge upon the monitoring performance of TFs in assessing AGB. However, previous studies have primarily relied on default texture parameters setting (such as a 3x3 window size and diagonal direction at 45°) (Zheng et al., 2018; Li et al., 2019; Zheng et al., 2019; Xu L. et al., 2022; Zhang D. et al., 2022; Zhang Y. et al., 2022;) or a directionless approach (by averaging multiple directional TFs to eliminate the directional effect) (Wang F. et al., 2021; Liu et al., 2022a; Xu T. et al., 2022) when extracting GLCM-based TFs. A quantitative analysis of the impact of TFs derived from different window sizes and direction parameters for crop AGB estimation has been omitted from these studies.

Fortunately, a limited amount of research has recently emerged that focuses on the impact of texture parameters on the accuracy of AGB monitoring. For example, Fu et al. (2021) and Yue et al. (2019) had demonstrated that the impact of window size and directional parameters on winter wheat AGB estimation was relatively minor. Zheng et al. (2020) discovered a conspicuous directional effect in rice texture information, with the direction parallel to the planting rows (i.e., the 0° direction of GLCM) being the optimal direction for monitoring rice nitrogen content. For monitoring potato AGB, contrasting conclusions had been drawn by Liu et al. (2022a) and Luo et al. (2022). Liu et al. considered directional parameters could be negligible for estimating AGB, whereas Luo et al. considered the 45° direction optimal for texture extraction. Although some studies have confirmed the sensitivity of crop AGB to TFs related to window size and directional parameters, further research in this area is still lacking in comprehensiveness. Given the wide use of GLCM-based TFs in crop growth monitoring, there is an urgent need to develop a deeper understanding of how texture parameters impact rice AGB monitoring. To address this, our study utilizes Random Forest (RF) regression model to investigate the influence of GLCM texture parameters on the monitoring accuracy of rice AGB during critical growth stages. The specific objectives of our study are as follows: 1) to evaluate how texture window size affects the accuracy of rice AGB monitoring and determine the suitable window size to estimate AGB; 2) to assess the impact of the directional parameter on the accuracy of rice AGB monitoring and identify the optimal direction for texture extraction; and 3) to explore the significance of TFs for rice AGB estimation and interaction mechanistic.

2 Materials and methods

2.1 Experimental design

A field experiment was conducted at the modern agricultural research institute, Anhui Science and Technology University located in Xiaogang Village, Fengyang County, Anhui Province, China (117°42' E, 32°16' N) (Figure 1A), where the climate belongs to a transitional pattern from the northern subtropical to the temperate zone. The average annual rainfall in 2020 was 1179.2 mm, and the average annual temperature was 15.5°C (Station No. 58222, Fengyang County Meteorological Station). The experiment design involved four N fertilizer treatments (N0: 0 kg/ha; N1: 100 kg/ha; N2: 200 kg/ha, and N3: 300 kg/ha) and three rice varieties (V1: RunzhuXiangzhan, V2: RunzhuYinzhao, and V3: Hongxiangnuo). A randomized complete block design was adopted for the experiment field, with N fertilizer treatments as the main plot and rice varieties as the split plot. There were three replications, leading to a total of 36 subplots, measuring 2 m × 8 m each, and the double-layer impermeable plastic film was used to isolate different nitrogen fertilizer treatments plot (Figures 1B, C). The experiment started in May 2020 with land preparation and delineation of plots, followed by basal fertilizer application and irrigation. Seeding took place on May 24th, and transplanting was conducted on June 23rd. The rice seedlings were manually inserted at a spacing of 30 cm between rows and 15 cm between plants, with one seedling per hill (including 1-2 tillers). In mid-July, an additional application of topdressing fertilizer was administered during the tillering stage. Subsequently, in early August, the plants entered the heading and panicles initiation stage, and on August 23rd, another additional fertilization was applied. The early grain-

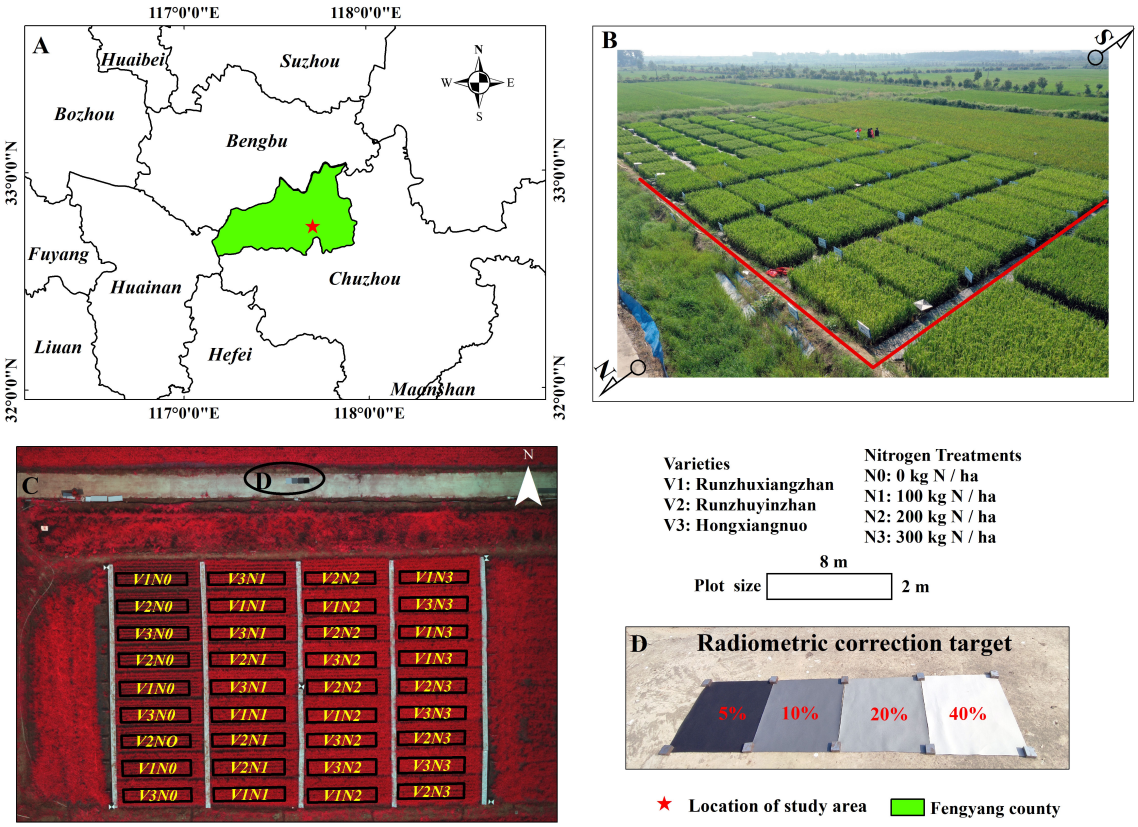


FIGURE 1
Location of study area **(A)**, field experimental design **(B, C)**, and radiometric correction target **(D)**.

filling stage began in early September, and harvest was conducted on October 2nd. Phosphorus fertilizer ($CaP_2H_4O_8$, with available P_2O_5 content of 12% and a pure phosphorus equivalent of 90 kg/ha) and potassium fertilizer (KCl, with available K_2O content of 60% and a pure potassium equivalent of 135 kg/ha) were applied as basal fertilizers. Nitrogen fertilizer (urea/CH_4N_2O, with an available nitrogen content of 46%) was applied in three stages, with a ratio of 4:3:3 for basal, tillering, and panicle fertilizer. Field management followed local high-yield cultivation techniques, including the application of herbicides and pesticides as general practices in this area.

2.2 Main workflow

Figure 2 shows the flowchart describing the procedures for estimating rice AGB from UAVs-based multispectral imagery. This process consisted of five main aspects: (1) acquiring UAVs remote sensing images and ground-truth AGB measurements; (2) preprocessing the acquired data in step I; (3) extracting TFs using various window sizes and directional parameters combination; (4) Random Forest (RF) regression model was built based on TFs calculated from step III for rice AGB estimation, and analyzed the significance of the TFs; (5) Estimating results were used to map AGB during the whole growth season.

2.3 Data acquisition and preprocessing

2.3.1 UAVs data acquisition and preprocessing

The DJI Phantom 4 Multispectral RTK (P4M) UAVs (DJI, Shenzhen, Guangdong, China) was used to acquire multispectral images at four growth stages, including the late tillering stage (LT: 25/07/2020), booting stage (B: 23/08/2023), heading to flowering stage (HtF: 31/08/2023), and early filling stage (EF: 09/09/2023) (Table 1). The P4M equipped with five monochrome sensors are used for multispectral images (blue (450nm±16nm, B), green (560nm±16nm, G), red (650nm±16nm, R), red edge (730nm±16nm, RE), and near-infrared (840nm±26nm, NIR) image). Among them, the RE and the NIR are the important selection of crop parameter inversion (Cui and Kerekes, 2018a; Liu J. et al., 2022).The multispectral sensors record each monochrome for 2.12-megapixel with a 40-mm focal length and 62.7-degree field of view. The multispectral image was geotagged automatically by its built-in multi-frequency high-precision RTK GNSS positioning system which provides about 8 mm and 15 mm accuracy in the vertical and horizontal directions, respectively. Supplementary Table 1 describes the technical specifications of the P4M.

Flight campaigns were planned to utilize DJI GS Pro software, which allows autonomous path points to be defined by the user in all flight campaigns. All campaigns were conducted on cloud-free days from 10:00–12:00 A.M. (local time) and were planned at a flight altitude of 30 m and flight speed of 3 m/s with a forward

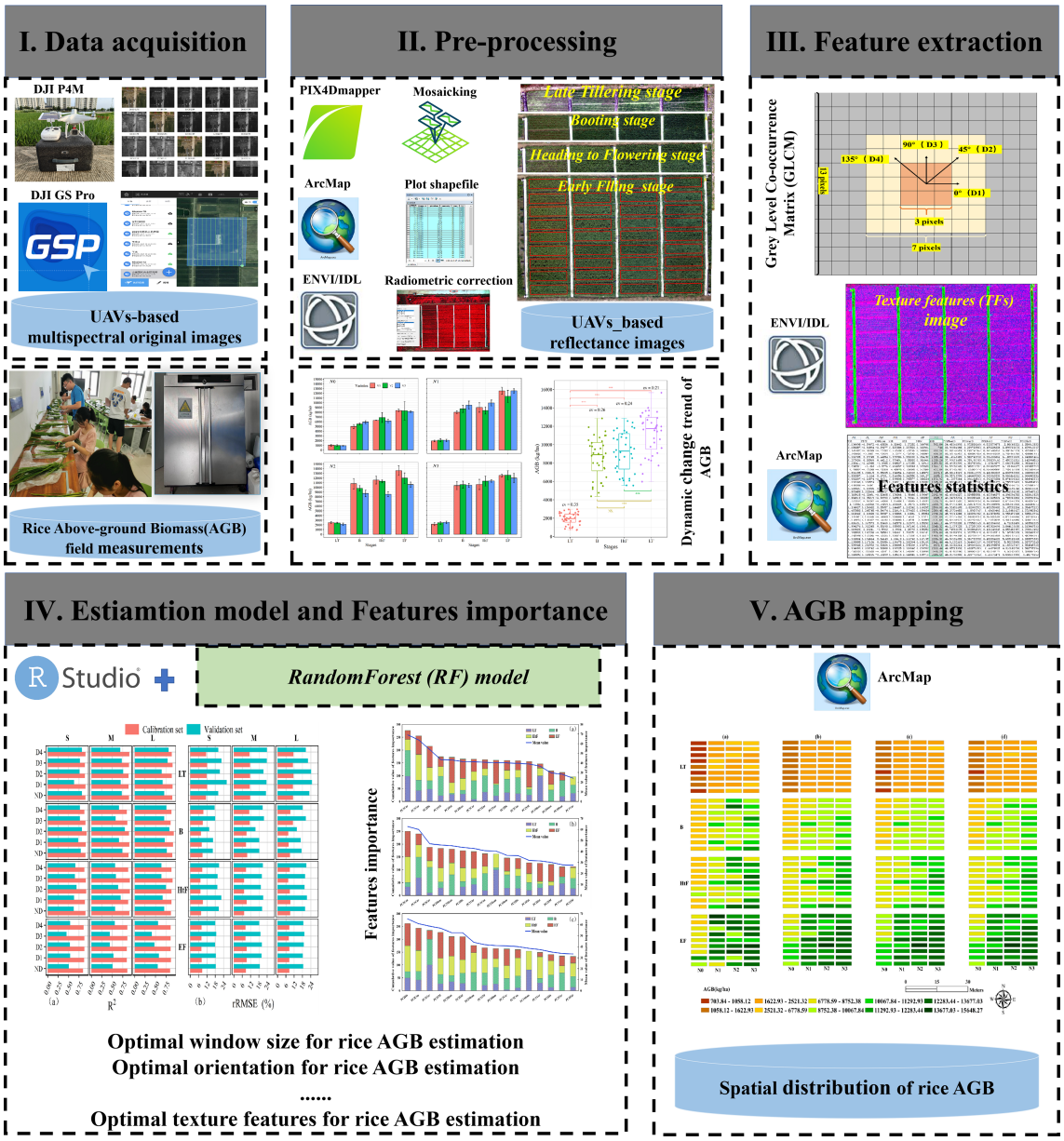


FIGURE 2
Flowchart of this study.

overlap of 90% and a side overlap of 85%. The flight campaigns were kept consistent throughout the growth season to ensure consistency in data collection.

PIX4Dmapper software (Pix4D SA, Lausanne, Switzerland, https://www.pix4d.com/) was used to generate orthophoto images from four critical growth stages (Wang F. et al., 2022). Images were aligned based on key points, then mosaicked and generated dense point clouds. To build high-density point clouds, "Half image size" for the image scale option and "Optimal" for the point density option were selected. Subsequently, a textured mesh was generated based on the constructed point cloud, resulting in the production of the Digital Surface Model (DSM) and Digital Orthophoto Image (DOM).

Due to the absence of spectral response function (SRF) of camera, it is not possible to quantify the influence of SRF (Cui and Kerekes, 2018b). Hence, we employ empirical linear model (ELM) for radiometric correction purposes (Di Gennaro et al., 2022; Liu J. et al., 2022). Four standard diffuse panels (Figure 1D) were placed on one side of the study area within the UAVs field of view (Figure 1C). The reflectance of each diffuse panel was measured using the ASD FieldSpec HandHeld2 portable spectrometer (Analytical Spectral Devices, Boulder, Colorado, USA). The digital number (DN) values of the UAVs images were converted into reflectance values using ELM to eliminate radiometric distortions caused by variations in lighting conditions during different stages. The ELM is conducted through the following Equation (1).

TABLE 1 The flight details during the whole growth season.

Dates	Growth stages	Height(m)	Spatial resolution (cm*cm)
07/25/20	Late Tillering (LT)	30	1.70
08/23/20	Booting (B)		
08/31/20	Heading to Flowering (HtF)		
09/09/20	Early Filling (EF)		

$$R_i = DN_i \times Gain_i + Offset_i \, (i = 1, 2, \ldots, 5) \quad (1)$$

where, R_i and DN_i respectively represent the reflectance values and original DN values corresponding to the i th band in the multispectral image. $Gain_i$ and $Offset_i$ respectively represent the conversion coefficient corresponding to the ith band, which are calculated using the ordinary least squares (OLS) method.

In order to conduct analysis and modeling at the plot level, the unique shapefiles, which removed inward a row of rice plant, were created manually using ArcGIS(10.2 version, Environmental Systems Research Institute, Inc., Redlands, CA, USA) (Figure 1C). This approach was implemented to account for the vigorous growth of rice plant at the margin, which can be influenced by incomplete spatial constraints or competition with contiguous rows. Failing to mitigate such boundary effects could lead to an overestimation of spectral reflectance within the plot.

2.3.2 AGB measurement

Following the collection of UAVs data, we randomly sampled 3 hill plants from each plot by cutting the stems approximately 2 cm above the soil surface and the stems, leaves, and panicles would be separated for AGB measurements. The separated samples were then oven-dried at 105°C for 30 min and then at 75°C until weights stabilized. Dry samples were weighed and summed to obtain the aboveground dry biomass. Finally, 144 AGB measurements were collected in total (four growth stages) and converted into a unified value of kg/ha based on the planting row and plant spacing.

2.4 Feature extraction

Texture analysis methods can be categorized into four types: statistical method, structural method, model-based method, and transformation-based method (Haralick et al., 1973; Hall-Beyer, 2007; Hall-Beyer, 2017). The most commonly used method is based on the Gray-Level Co-occurrence Matrix (GLCM), which was first introduced by Haralick in 1973 to reveal the variation properties of the spatial distribution of grayscale values in an image at a certain distance (d) and specific angle (θ) (Pacifici et al., 2009). GLCM ensures non-deformation, rotation-invariant multi-scale features, and low computational complexity. In this study, eight TFs were selected: Mean (Mean), Variance (Var), Homogeneity (Hom), Contrast (Con), Dissimilarity (Dis), Entropy (Ent), Second Moment (SM), and Correlation (Cor) (Table 2).

The GLCM requires users to define the window size, direction, and displacement of the moving window. In this study, the displacement was set to 1 as it is the most commonly used setting (Liu Y. et al., 2019; Zheng et al., 2019; Wang Q. et al., 2022; Xu L. et al., 2022). To determine and select the most optimal texture parameters, TFs for all multispectral images were calculated using three window sizes (3 × 3, 7 × 7, and 13 × 13 pixels, denoted as S, M, and L, respectively) and four directions (0°, 45°, 90°, and 135°, denoted as D1, D2, D3, and D4, respectively). The average TFs values of these four directions were then obtained to achieve rotation invariance, referred to as the non-directional (ND) texture metrics. Figure 3 depicts more details on the selection of window sizes and directions for TFs calculation.

All TFs were calculated from the component images after principal component analysis (PCA). On one hand, this choice was made to reduce redundancy among the multispectral data; on the other hand, to minimize the computational workload during GLCM construction (Liu C. et al., 2019). In this study, the principal components with a cumulative contribution rate exceeding 90% were selected (Supplementary Figure 1), specifically the first two

TABLE 2 Texture features of multispectral imagery (Hall-Beyer, 2007; Pacifici et al., 2009; Hall-Beyer, 2017).

Texture features	Calculation equations	Features description
Mean (Mean)	$Mean = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} i \times P(i,j)$	The mean value in the GLCM window.
Variance (Var)	$Var = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} (i-u)^2 \times P(i,j)$	The variance in the GLCM window.
Homogeneity (Hom)	$Hom = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{1}{1+(i-j)^2} \times P(i,j)$	The homogeneity of grey level in the GLCM window.
Contrast (Con)	$Con = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} (i-j)^2 \times P(i,j)$	The clarity of texture in the GLCM window, as opposed to HOM.
Dissimilarity (Dis)	$Dis = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \lvert i-j \rvert \times P(i,j)$	The similarity of the pixels in the GLCM window, similar to CON.
Entropy (Ent)	$Ent = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P(i,j) \times \log[P(i,j)]$	The diversity of the pixels in the GLCM window, proportional to the complexity of the image texture.
Second Moment (SM)	$SM = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} [P(i,j)]^2$	The uniformity of greyscale in the GLCM window.
Correlation (Cor)	$Cor = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} (i,j) \times P(i,j) - u_i u_j}{\sigma_i \sigma_j}$	The linear dependency of greyscale on those of neighboring pixels in the GLCM window.

P(i,j) represents the probability of each pixel pair (i,j) value and i, j are the gray tones in the windows, which are also the coordinates of the co-occurrence matrix space; Ng represents the number of distinct grey levels in the quantized image, which has a gray value range of the original image; μ and σ represents the mean and standard deviation of P(i,j), respectively.

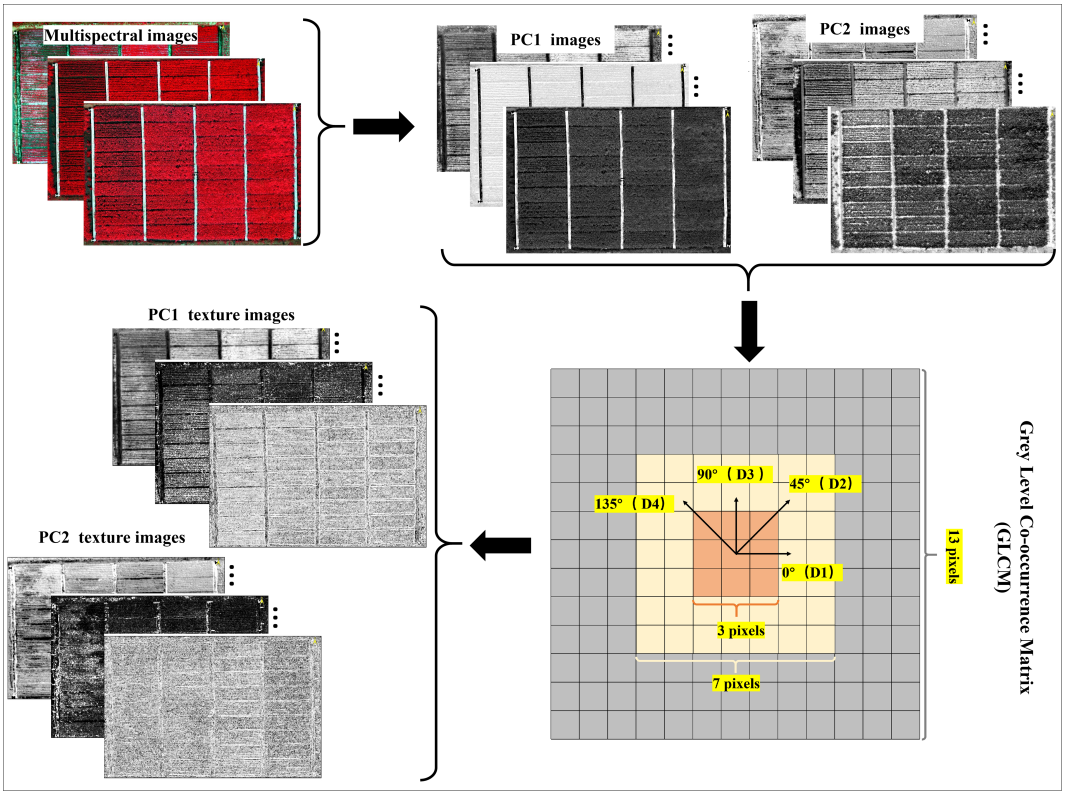


FIGURE 3
Details of GLCM-based texture features calculation.

principal components from the PCA analysis of the multispectral imagery. After calculation, a total of 240 [2 (principal components of PCA) × 8 (texture features) × 3 (window sizes) × 5 (directions) = 240] TFs were generated for each growth stage.

2.5 AGB estimation model

2.5.1 Random Forest model

Random Forest (RF) is an ensemble learning algorithm proposed by Breiman and is based on multiple decision trees and Bagging technology (Breiman, 2001). In the model, decision trees are built in parallel, with each tree trained on a different subset of data. Thus, each decision tree is unique, reducing the model's variance and lowering prediction errors (Yu et al., 2016). For regression models, the main advantages of RF are as follows: 1) lack of sensitivity to collinearity among multiple variables; 2) presence of few parameters that require tuning, with only one hyperparameter in this study; 3) effective reduction of the risk of overfitting; 4) automatic calculation of variable importance scores to assess the contribution of individual predictors to the model (Liu Y. et al., 2019; Burdett and Wellen, 2022; Borrmann et al., 2023).

The RF model comprises two crucial hyperparameters: the number of decision trees (ntree) and the number of input variable features at each node (mtry). When adjusting ntree to a sufficiently large value, it primarily impacts the modeling time rather than the modeling accuracy (Wang et al., 2016; Zhang et al., 2021). Therefore, following its application in other studies, we set ntree to 1,000 (Li et al., 2019; Zhu et al., 2023). On the other hand, the value of mtry significantly affects the modeling accuracy of RF. Thus, it should be adjusted based on the number of input variable sets to optimize the RF model. The determination of the parameter mtry involves a grid search for parameter optimization.

2.5.2 Accuracy assessment

Cross-validation (CV) was employed based on 70% of samples to determine the AGB monitoring model with the highest determination coefficient (R^2) and lowest root mean squared error (RMSE) for improving the model's stability. The RF model was then evaluated on the remaining 30% of samples (Supplementary Table 2). This approach effectively enhances the applicability of the RF algorithm on small datasets.

Richter et al. (2012) recommends a set of statistical test metrics that can comprehensively quantify the performance of models through literature review and experimental calculation. In this study, we adopt three recommended metrics: R^2, RMSE, and relative root mean squared error (rRMSE). Their calculation formulas are presented as Equations (2–4)).

$$R^2 = \frac{\sum_{i=1}^{n}(AGB_{obs}^{i} - \overline{AGB_{obs}}) \times (AGB_{est}^{i} - \overline{AGB_{obs}})}{\sqrt{\sum_{i=1}^{n}(AGB_{obs}^{i} - \overline{AGB_{obs}})^2 \times \sum_{i=1}^{n}(AGB_{est}^{i} - \overline{AGB_{obs}})^2}} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(AGB_{est}^{i} - AGB_{obs}^{i})^2} \quad (3)$$

$$rRMSE = 100 \times \frac{RMSE}{\overline{AGB_{obs}}} \quad (4)$$

where n represents the number of samples in RF model; AGB_{obs}^{i} and AGB_{est}^{i} represent the truth ground AGB measurement and the estimated AGB value of i sample, respectively; and $\overline{AGB_{obs}}$ and $\overline{AGB_{est}}$ represent the average truth ground AGB measurement and the average estimated AGB value of all samples, respectively.

The normalized increase in mean square error (%IncMSE) ranging from 0 to 100 is used to assess the importance of variables in the RF model. %IncMSE is calculated by permuting out-of-bag (OOB) data, where higher percentages indicate greater importance of variables. For a detailed description of %IncMSE, please refer to the literature (Breiman, 2001). All data analyses were conducted using R programming language (https://www.r-project.org) in RStudio software (Version 4.2).

3 Results

3.1 Effects of growth stages and nitrogen levels on rice AGB

Figure 4 shows the changes in rice AGB of all rice plots under different growth stages and treatments. AGB increased rapidly with the development of growth stages, ranging from 1921.42 kg/ha to 11179.76 kg/ha, with a standard deviation of 4,009.28 kg/ha. The coefficient of variation (CV) exhibited a gradual decline during the crop growth stages, diminishing from 0.35 at the LT stage to 0.21 at the EF stage (Figure 4A and Supplementary Table 3). With the increase of N fertilization, the AGB exhibited the same change pattern as the growth stage, yet CV values were higher between different nitrogen levels (CV:0.45 ~ 0.55) (Figure 4B and Supplementary Table 4).

3.2 Effects of the window size parameter on the accuracy of rice AGB estimation

To determine the appropriate window size for extracting TFs from multispectral PCA images (PC1 and PC2 images), we calculated TFs using three different window sizes. The RF estimation model integrated 240 extracted TFs to evaluate the window size that yielded the most optimal model performance (Table 3). Notably, the appropriate window size exhibited dynamic variations across different growth stages. Small window size demonstrated superior performance during the vegetative growth period, while large window size performed better during the reproductive growth period. Interestingly, the monitoring accuracy of AGB was the lowest when TFs derived from a medium-size window were used during the four crucial growth periods. At the LT stage, the highest precision for AGB monitoring was achieved from a small window size, with R^2=0.82, RMSE=378.64 kg/ha, and rRMSE=20.3%. Similarly, the best accuracy of R^2=0.82, RMSE=1173.01 kg/ha, and rRMSE=13.62% was also achieved from a small window size at the B stage. At the HtF and EF stages, the highest estimation accuracy was achieved both from large window size that R^2=0.75, RMSE=1658 kg/ha and rRMSE=19.42% and R^2=0.58, RMSE=1848.66 kg/ha and rRMSE=18.28%, respectively.

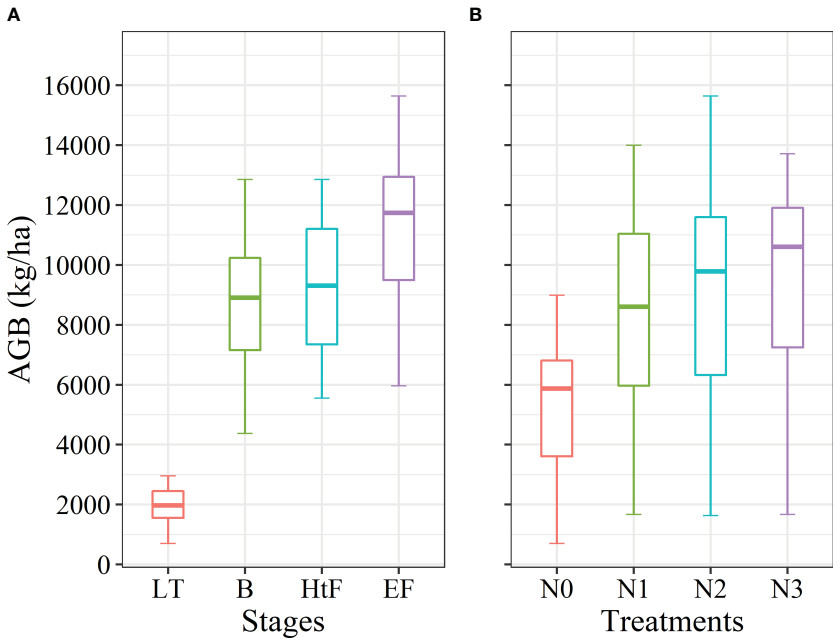


FIGURE 4
Rice AGB variation with **(A)** growth stages and **(B)** N treatments over the experimental plots. LT, Late Tillering stage; B, Booting stage; HtF, Heading to Flowering stage; EF, Early Filling stage.

TABLE 3 Estimation accuracy of rice AGB model during the critical growth stages based on texture features.

Windows	Directions	Late Tillering stage			Booting stage			Heading to Flowering stage			Early Filling stage		
		RMSE (kg/ha)	R^2	rRMSE (%)	RMSE (kg/ha)	R^2	rRMSE (%)	RMSE (kg/ha)	R^2	rRMSE (%)	RMSE (kg/ha)	R^2	rRMSE (%)
Small	ND	376.35	0.82	20.17	1401.00	0.79	16.27	1761.93	0.62	20.64	1869.57	0.50	18.49
	D1	491.46	0.63	26.35	1410.43	0.73	16.38	1918.14	0.63	22.47	1872.06	0.54	18.51
	D2	399.85	0.76	21.43	1173.01	0.82	13.62	1664.21	0.69	19.50	1813.24	0.53	17.93
	D3	421.89	0.77	22.62	1552.01	0.77	18.02	1686.04	0.65	19.75	1908.85	0.45	18.88
	D4	378.64	0.82	20.3	1511.91	0.77	17.56	1779.76	0.66	20.85	1826.19	0.54	18.06
Middle	ND	431.71	0.73	23.14	1577.41	0.74	18.32	1744.06	0.57	20.43	1897.33	0.51	18.76
	D1	436.52	0.7	23.40	1574.8	0.66	18.29	1808.40	0.70	21.19	2019.91	0.53	19.98
	D2	397.88	0.76	21.33	1335.28	0.80	15.51	1658.65	0.64	19.43	2038.48	0.44	20.16
	D3	442.94	0.73	23.74	1866.81	0.58	21.68	1720.56	0.59	20.16	1928.3	0.44	19.07
	D4	443.72	0.69	23.79	1798.19	0.66	20.88	1793.00	0.57	21.00	1848.5	0.56	18.28
Large	ND	426.54	0.71	22.87	1490.65	0.77	17.31	1691.00	0.71	19.81	1870.34	0.55	18.5
	D1	461.52	0.67	24.74	1425.64	0.72	16.56	1802.98	0.67	21.12	1848.66	0.58	18.28
	D2	453.15	0.65	24.29	1198.83	0.83	13.92	1658.00	0.75	19.42	1919.27	0.47	18.98
	D3	410.27	0.79	21.99	1766.66	0.71	20.52	1698.83	0.64	19.90	1949.16	0.41	19.28
	D4	381.10	0.81	20.43	1572.64	0.76	18.26	1774.50	0.64	20.79	1879.00	0.56	18.58

3.3 Effects of the direction parameter on the accuracy of rice AGB estimation

TFs possess inherent directional properties, and the influence of directional parameters on rice AGB estimation was complex (Table 3). Taking into account the 4 growth stages, 3 window sizes, and 5 directions, there was a three-fourth probability (among the 12 optimal models, 8 models achieved the highest accuracy) that AGB estimation models exhibited greater accuracy in the D2 direction (45°), especially during the intermediate periods of rice growth (from B stage to HtF stage in this study). Taking the example of large window size, at the B stage, the highest accuracy for AGB estimation was achieved in the D2 direction, with R^2 = 0.83, RMSE = 1,198.83 kg/ha, and rRMSE = 13.92%. Similarly, at the HtF stage, the highest accuracy for AGB estimation was also observed in the D2 direction, with R^2 = 0.75, RMSE = 1,658.00 kg/ha, and rRMSE = 19.42%. Acceptable estimation accuracy was achieved for rice AGB in the D4 direction (135°, orthogonal to the D2 direction) as well. In contrast, the poorest performing directions did not exhibit a consistent pattern, as it occurred with a probability of 5 out of 12 optimal models in both D1 (0°) and D3 (90°) directions. Taking the large window size as an example, the TFs from D1 direction exhibited the lowest monitoring accuracy for AGB during the LT and HtF stages, with R^2 =0.67 and 0.67, and rRMSE = 24.74% and 21.12%, respectively. Moreover, the TFs from the D3 direction demonstrated the lowest monitoring accuracy for AGB during the B and EF stages, with R^2 = 0.71 and 0.41, and rRMSE = 20.52% and 19.28%, respectively. The accuracy of the AGB estimation model based on ND TFs fell between the best and worst accuracy across the three window sizes.

3.4 Importance analysis of texture features for rice AGB estimation

In this study, we evaluated the importance scores of TFs at different growth stages in the D2 direction based on the RF models(Figure 5). The importance scores of TFs varied with window sizes. For small window size, the important TFs were PC1Cor, PC2Con, and PC2Dis. For medium window size, PC2Con and Cor were found to be important TFs. For large window size, PC2Dis, PC2Con, and PC1Cor were identified as significant TFs. The Hom, Ent, Cor, and Var from the PC2 image demonstrated a moderate level of importance scores.

Further analysis of the TFs extracted from different principal component images revealed an inverse variation in the importance scores of TFs between PC1 and PC2 images (Figure 6). During the vegetative growth period, the significance of TFs in the PC1 image exceeded that of the PC2 image. However, the TFs in the PC2 image were more prominent during the reproductive growth period. For the PC1 image, the TFs with higher importance scores were Mean, Cor, Con, and Hom. For the PC2 image, the TFs that stand out were SM, Ent, Con, Dis, and Hom.

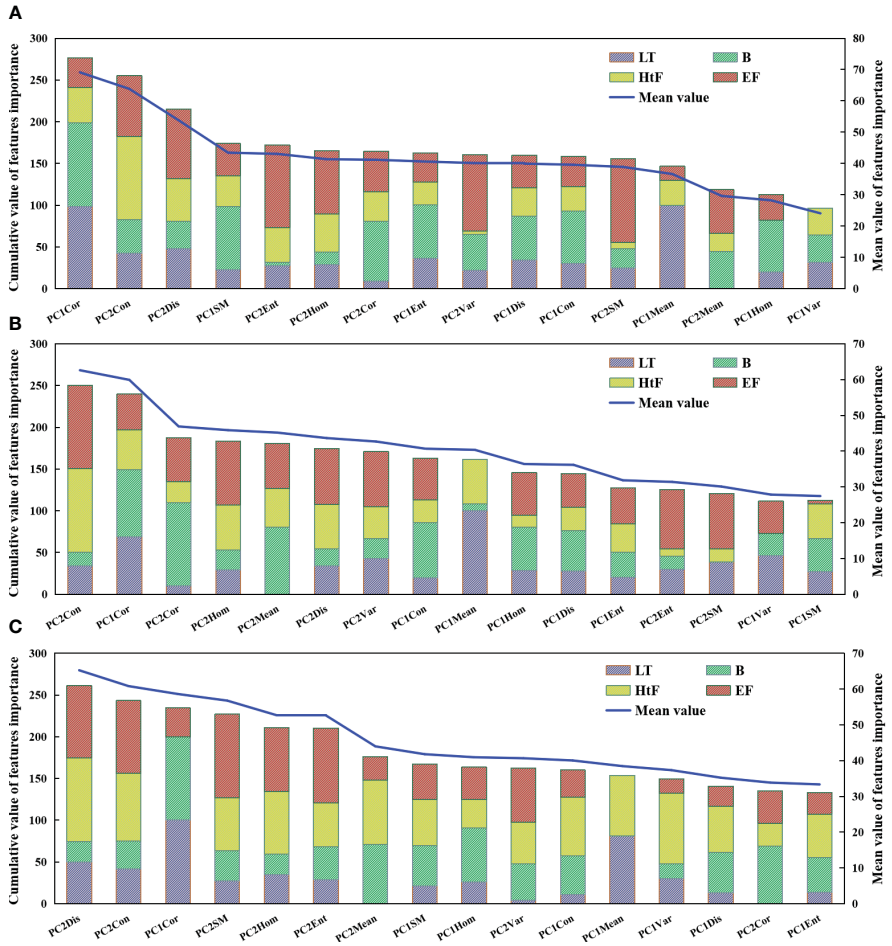


FIGURE 5
Importance analysis of texture features for rice AGB monitoring under small window size **(A)**, middle window size **(B)**, and large window size **(C)**. LT, Late Tillering stage; B, Booting stage; HtF, Heading to Flowering stage; EF, Early Filling stage.

The importance scores of TFs vary with the growth stages (Figures 5, 6). At the LT stage, TFs such as Cor and Mean derived from the PC1 imagery demonstrated higher significance. Almost all TFs extracted from the B stage were found to be important across three different window sizes, and the most important TFs was Cor. At the HtF stage, the importance scores of Con, Dis, and Hom were greater than that of other TFs. At the EF stage, multiple TFs from the PC2 images show importance, such as SM, Ent, Hom, Dis, and Con.

Figure 7 shows the spatial distribution of estimated rice AGB at various growth stages based on TFs derived from the D2 direction. The selection of the D2 direction allowed for a more effective assessment of the estimated accuracy in rice AGB throughout the growth stages, and more details can be found in the Results section (Section 3.3). The results demonstrated a high spatial consistency with the observed AGB throughout the entire growth season. AGB continued to increase, and the differences within subplots intensified over time. AGB showed correlations with nitrogen fertilization levels and rice varieties. As nitrogen application increased, AGB values also increased. Additionally, differences were observed among varieties, with V1 often exhibiting higher AGB values compared to V2 and V3.

3.5 Comparison of spectral-based and texture-based features for AGB estimation

This study also compared the differences in rice AGB estimation between TFs derived from optimal texture parameters and spectral features (SFs) (reflectance of spectral bands) (Table 4). At the LT stage, both SFs and TFs exhibited comparable accuracy for estimating AGB (TFs: R^2=0.82, SFs: R^2=0.82). During the mid-growth stages (e.g. B stage and HtF stage in this study), TFs tend to exhibit higher accuracy in AGB estimation compared to SFs (TFs: R^2=0.83, SFs: R^2=0.78 at the B stage, and TFs: R^2=0.70, SFs: R^2=0.47 at the HtF stage). At the EF stage, the accuracy of estimating AGB using TFs was lower compared to SFs (TFs: R^2=0.58, SFs: R^2=0.63). It is noteworthy that the fusion of SFs and TFs does not yield improved accuracy for rice AGB estimation compared to the use of either SFs or TFs alone (Supplementary Table 5).

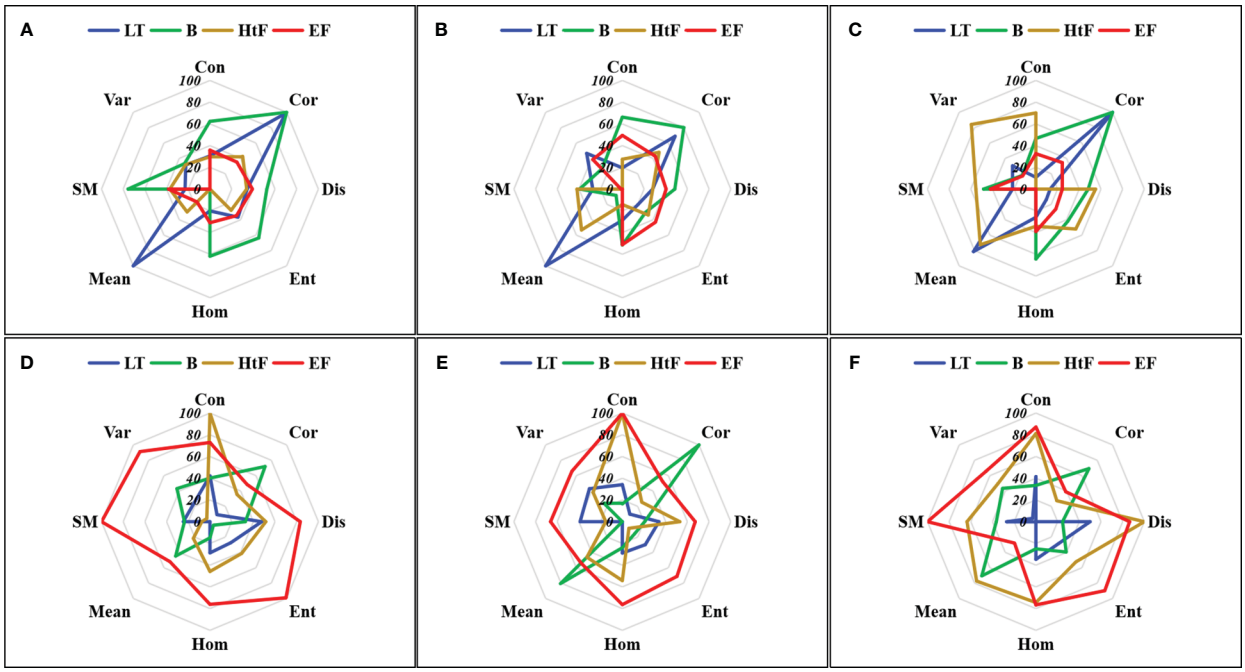


FIGURE 6
Importance analysis of texture features in the first principal component image **(A–C)** and second principal component image **(D–F)**. **(A, D)** Small window size; **(B, E)** Middle window size, and **(C, F)** Large window size. LT, Late Tillering stage; B, Booting stage; HtF, Heading to Flowering stage; EF, Early Filling stage.

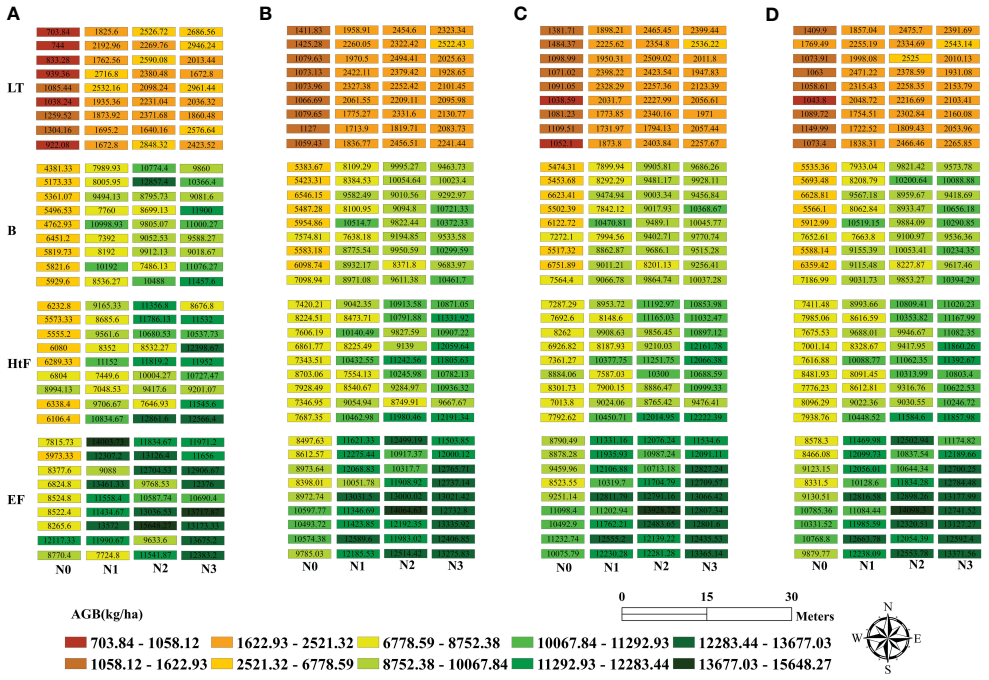


FIGURE 7
Rice AGB mapping based on field measurement **(A)** and estimated AGB values with texture features derived from D2 direction (45°); **(B)** Small window size; **(C)** Middle window size, and **(D)** Large window size. LT, Late Tillering stage; B, Booting stage; HtF, Heading to Flowering stage; EF, Early Filling stage.

TABLE 4 Estimation accuracy of rice AGB model during the critical growth stages based on spectral features.

Stages	Calibration set			Validation set		
	RMSE (kg/ha)	R^2	rRMSE (%)	RMSE (kg/ha)	R^2	rRMSE (%)
LT	237.56	0.89	12.21	343.95	0.82	18.44
B	779.01	0.90	9.09	1339.28	0.78	15.55
HtF	827.80	0.87	8.65	1931.32	0.47	22.62
EF	1021.92	0.87	8.77	1857.75	0.63	18.37
All	755.29	0.97	9.96	1938.27	0.79	23.96

4 Discussion

4.1 The optimal window size for extracting texture features

Window size is a crucial variable in TFs extraction, as different window sizes impact the frequency of pixel value occurrences during the process of texture calculation (Marceau et al., 1990; Liang et al., 2022; Liu et al., 2022a). To capture the object-specific TFs in an image, the window size must be smaller than the object's size but large enough to include the variability of the features of the object (Rodriguez-Galiano et al., 2012; Zhou et al., 2017; Zhang et al., 2020; Liang et al., 2022). This study analyzed the relationship between TFs and AGB estimation accuracy under different window sizes and found that the optimal window size for computing TFs was closely related to the crop growth period (Table 3). During the vegetative growth stage, a small window size is more suitable, while the large window size appears to offer advantages in capturing TFs during the reproductive growth stage. The reason could be explained by the relative relationship between the rice canopy cover of different growth stage and window size. In general, at the early stage of rice growth (e.g., LT stage, Figures 8A–D), the rice plant were relatively small, and the canopy was partly closed. Small window size primarily captured the green information of rice plant, while minimizing the influence of water-soil background noise. This scale provided a more precise representation of the plant's growth status and improved the AGB estimation. However, when using the large window size that exceeded the scope of the rice canopy, significant interference related to water-soil background noise occurred. This led to a decrease in the signal-to-noise ratio of the TFs, reducing the AGB estimation accuracy. At the mid-to-late stage of rice growth (e.g., EF stage, Figures 8E–H), the rice plant experienced vigorous growth, resulting in the formation of a dense canopy. At a small window scale, which mainly comprised panicles and some leaf organs, the TFs were insufficient to capture the macroscopic characteristics of rice plant. However, using a large window size that covers the entire canopy of rice plant, with minimal inclusion of background soil and water information, a more comprehensive depiction of plant structure is achieved. This scale allowed a more comprehensive characterization of the rice canopy, which increased the accuracy of AGB monitoring. This provided a brief explanation of the findings described in Result 3.2, which matched the findings of Yue et al. (2019) and Zhou et al. (2017).

Given the growth of rice plant, there exists a specific period during which the extent of rice plant coverage aligns with a window of moderate size. The medium-sized window was expected to yield the highest accuracy in AGB monitoring during this particular stage. However, we did not observe this outcome during our four growth stages, mainly because the sampling periods did not encompass this critical growth stage. We hypothesized that there might be an offsetting effect between the optimal texture window scales during the early and late stages of rice growth. As a result, the medium-sized window would provide the optimal AGB monitoring accuracy throughout the entire growth season. To further validate this hypothesis, we conducted RF model using all data collected throughout the growth season (Table 5); as anticipated, it has been found that the best accuracy in monitoring AGB can be achieved through the extraction of TFs at the intermediate window scale. Therefore, we suggested that it is imperative to consider using the appropriate window sizes for calculating texture parameters at different growth stages, particularly when utilizing these parameters for rice AGB estimation.

4.2 The optimal direction parameter for extracting texture features

As one of the crucial research questions, the influence of direction on AGB monitoring is multifaceted and warrants thorough investigation (Hall-Beyer, 2007; Zheng et al., 2020; Liang et al., 2022; Liu et al., 2022a). Our findings suggested that D2 was the optimal direction for TFs extraction (Table 3), which matched the conclusions of Fu et al. (2020) and Fu et al. (2021). However, no definitive results have been obtained regarding the worst direction for TFs computations in this study. As we all know, rice plant naturally expands their growth in confined spaces; the planting distance between individual plants was smaller than the spacing between rows. Thus, they tended to close the spaces along the rows at the early growth stage, whereas, in the direction perpendicular to the planting rows, the plants closed the spaces during the middle period of growth. The diagonal directions (D2 and D4), which encompass both lateral and row-wise growth, offered a comprehensive reflection of the rice canopy distribution, making them more widely applicable for monitoring rice AGB. Along the planting rows, rice plant grow close together, reducing the influence of soil-water background noise. Even with the larger window size, canopy closes up in the early growth stage along the rows, and the pixel values become more similar, leading to

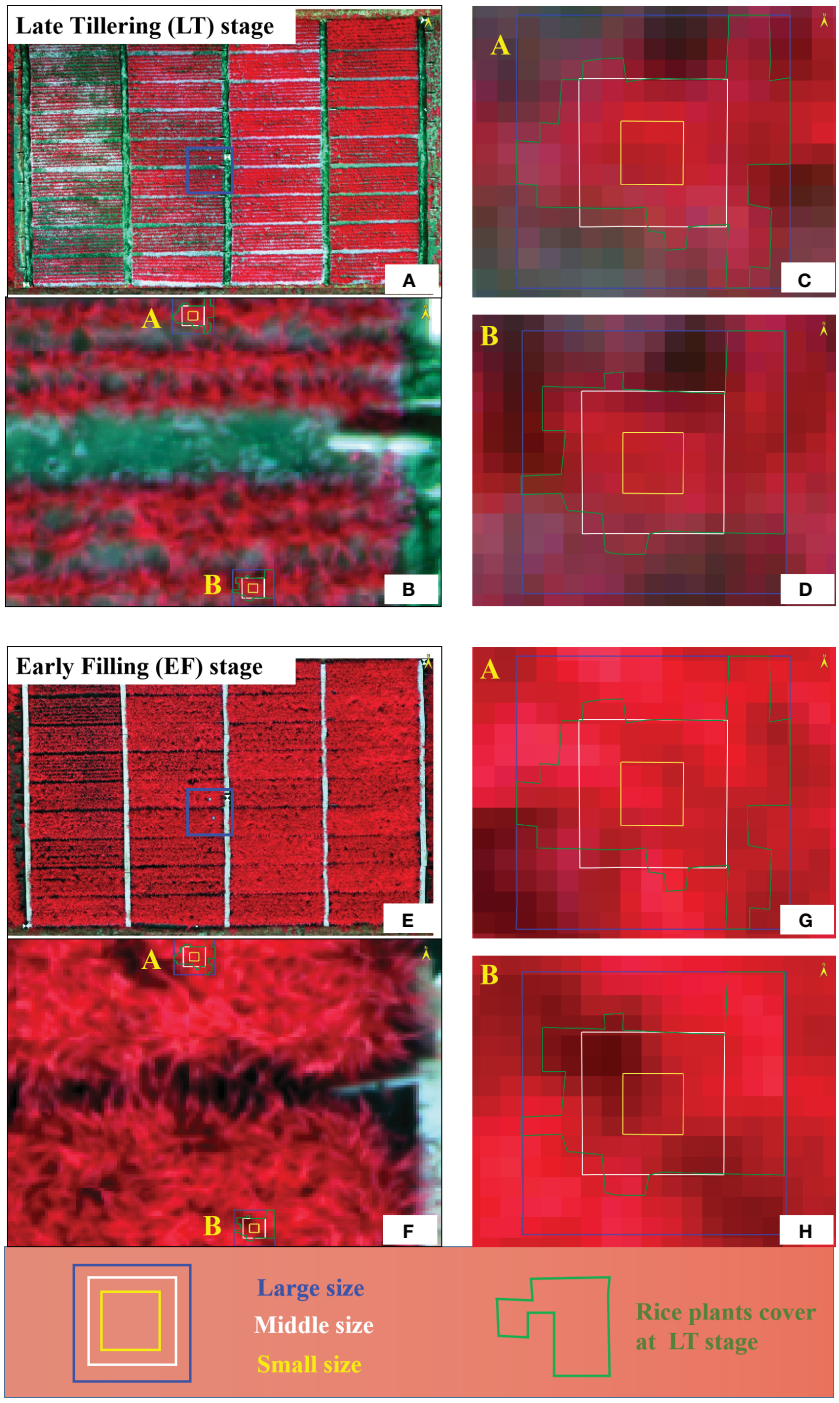


FIGURE 8
Dynamic changes of rice plant coverage with growth stage under different texture window sizes. The uppercase letters **(A, B)** represent two analysis points where the details within the respective window scales are examined at different magnification levels. **(A, E)** represent the full-view window at LT stage and EF stage, respectively; **(B, F)** represent the view window enlarged by 500%; **(C, D, G, H)** represent the view window enlarged by 2000%.

decreased possibility of capturing spatial variations in rice canopy using TFs (Guo et al., 2021; Luo et al., 2022). The direction perpendicular to the planting rows (D3) is susceptible to background noise, significantly reducing the accuracy of AGB monitoring (Zheng et al., 2018; Zheng et al., 2020). The observed differences between the D2 and D4 directions may be attributed to the geometric relationship of the sun, sensor, and rice plant. We conducted all flights before noon (local time) when the sun was positioned in the southeast direction. The flight routes were perpendicular to the rows of rice plant, from northwest to southeast. Therefore, the D2 direction was closer to the backward observation, which led to consistently higher and brighter values

were observed. This finding aligns with the conclusion drawn by Liang et al. (2022), who summarized that the most accurate estimation of rubber plantation AGB using TFs were achieved when conducting flight operations in the afternoon(local time) with 135° direction. We proposed this hypothesis to capture the interest of other researchers and encourage further investigation.

To further demonstrated the advantages of calculating TFs in the D2 direction, we conducted unified modeling throughout the whole growth season (Table 5). The results consistently showed that the D2 direction had the highest accuracy in AGB estimation, irrespective of the window size. This provided evidence supporting the rationale for considering the D2 direction as the optimal direction. Contrary to our finding, Zheng et al. (2020) found that texture information computed along the planting row direction (D1) is more advantageous for estimating rice leaf nitrogen content (LNC) and plant nitrogen content (PNC). This can be attributed to the lower imagery spatial resolution (5.4cm) in Zheng et al.'s study, they extracted TFs from 3 × 3 window size (approximately 16cm × 16cm), which primarily exhibit information pertaining to the row direction. Luo et al. (2022) suggested that TFs computed perpendicular to the ridges exhibited higher accuracy in estimating potato AGB. In contrast, Liu et al. (2022a) demonstrated that the direction selection did not affect the accuracy of TFs-based potato AGB estimation. The differences between these studies can be comprehensively considered from the perspective of image spatial resolution, crop type, planting pattern, and fertilizer application. Particularly, the significant impact of different spatial resolutions on GLCM-based TFs has been confirmed by previous studies (Yue et al., 2019; Liu et al., 2022a).

4.3 Important texture features for estimating rice AGB

Although the tremendous potential of TFs in estimating crop growth parameters has been repeatedly demonstrated, there has been limited research on the physical meanings of these important features (Zheng et al., 2018; Liu C et al., 2019; Yang et al., 2019; Guo et al., 2021; Liu et al., 2022a). In light of this, this study focused on analyzing the mechanistic of TFs that are of significant value for rice AGB estimation (Figure 9).

Compared to the PC1 image, the PC2 image expresses more edge information of internal organs within the rice plant, between rice plant, and between rice and background (Figures 9A, J). It represents high-frequency information in the image and can better capture structural difference in the field. Therefore, TFs derived from the PC2 image have higher importance for rice AGB estimation compared to the TFs derived from the PC1 image (Figure 6). The importance analysis of TFs has identified three texture metrics that are consistently important throughout the entire growth period: PC1Cor, PC2Con, and PC2Dis (Figure 5). From the calculation form, it can be observed that Cor does not assign weights to the difference in digital numbers (DN) between neighboring pixels (Table 2). Unlike other TFs, Cor focuses on internal texture (Hall-Beyer, 2007; Hall-Beyer, 2017), which is consistent with the fact that PC1 image primarily reflects internal information within the plants or the background (Figures 9A, B). Indeed, internal texture primarily represents low-frequency information within a specific window, reflecting gradual changes within the rice plant (Figure 9C). It focuses on capturing information within the plants rather than

TABLE 5 Estimation results of rice AGB during the growth season based on texture features.

Windows	Directions	Calibration set			Validation set		
		RMSE (kg/ha)	R^2	rRMSE (%)	RMSE (kg/ha)	R^2	rRMSE (%)
Small	ND	624.4	0.98	8.23	2077.88	0.76	25.68
	D1	667.89	0.97	8.81	1956.36	0.79	24.18
	D2	560.70	0.99	7.39	1838.56	0.81	22.73
	D3	666.55	0.97	8.79	2162.41	0.74	26.73
	D4	686.44	0.97	9.05	1949.04	0.79	24.09
Middle	ND	565.31	0.98	7.45	1766.42	0.82	21.83
	D1	529.04	0.98	6.98	1746.74	0.83	21.59
	D2	565.59	0.98	7.46	1704.28	0.84	21.07
	D3	596.49	0.98	7.87	1769.39	0.83	21.87
	D4	596.03	0.98	7.86	1836.60	0.81	22.70
Large	ND	572.97	0.98	7.56	1758.00	0.83	21.73
	D1	624.10	0.98	8.23	1825.55	0.81	22.56
	D2	573.87	0.98	7.57	1711.52	0.84	21.16
	D3	611.56	0.98	8.06	1796.42	0.82	22.20
	D4	593.81	0.98	7.83	1823.08	0.81	22.53

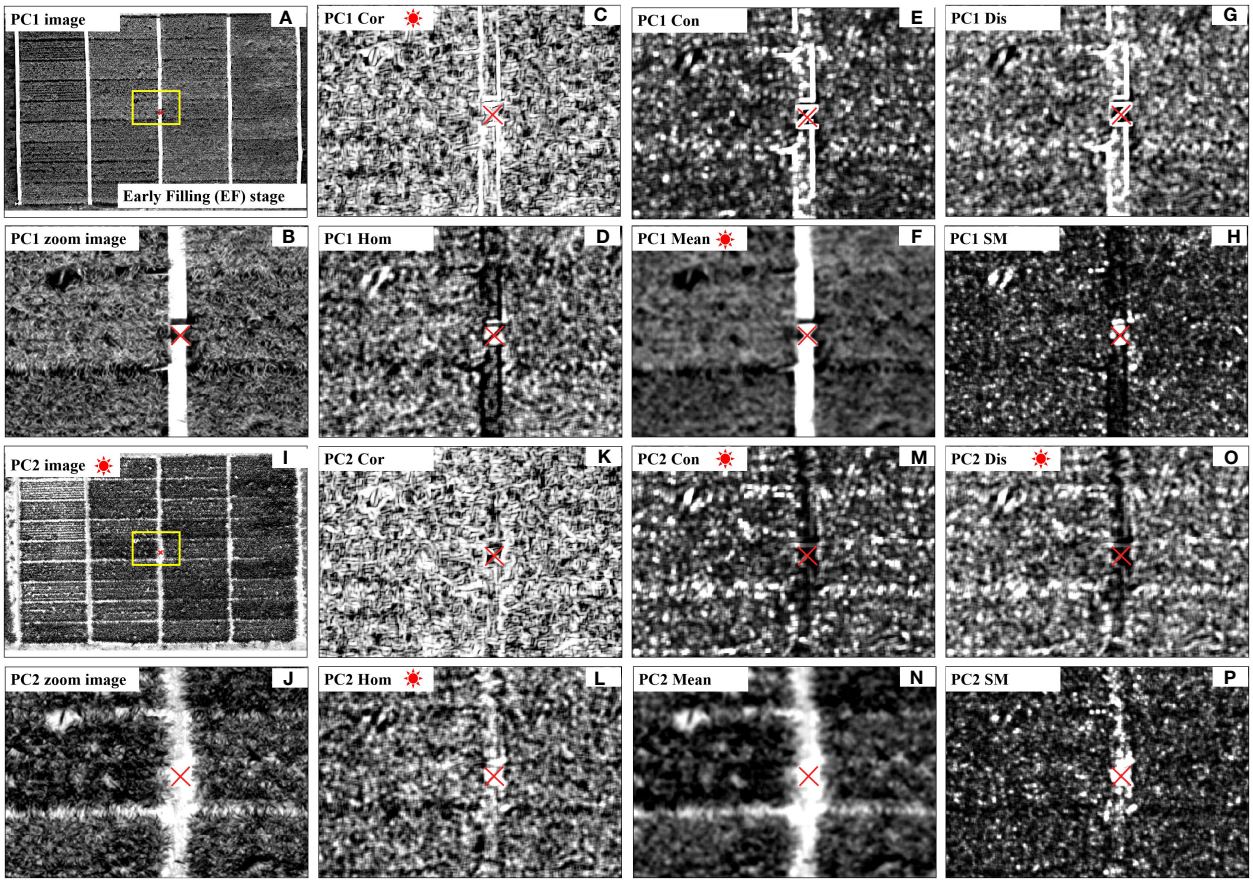


FIGURE 9
Important texture features for rice AGB estimation. The texture metrics Cor **(C, K)**, Hom **(D, L)**, Con **(E, M)**, Mean **(F, N)**, Dis **(G, O)** and SM **(H, P)** are from the principal component images; The PC1 **(A, B)** and PC2 **(I, J)** represent the first principal component image (PC1) and second principal component image (PC2); The red asterisks represent important texture features for rice AGB estimation, and the red multiplication symbol represents an important directional field object.

between plant organs or between plants and background. Cor represents the vigorously growing portions of the rice plant, closely correlated with high AGB values. As for PC1Mean, a metric indicating internal uniformity (Figure 9F), has the same physical meaning as PC1Cor. These findings align consistently with the conclusions from Xu T. et al. (2022); Zhou L. et al. (2023), and Zhu et al. (2022). Unlike Cor and Mean, Con and Dis were utilized to enhance the edge information in the image (Figures 9N, P) (Haralick et al., 1973; Hall-Beyer, 2007; Hall-Beyer, 2017; Guo et al., 2021), indicating high-frequency information within the specific window, that is consistent with the PC2 image representing edge information. Both Con and Dis precisely described the relationship between pixel frequency and the distance from the diagonal, thereby enhancing the effectiveness of AGB estimation. These findings align with the results from Guo et al. (2021), where the performance of Con outperformed other TFs.

Similar to Cor, HOM is a metric that quantifies the homogeneity of the grey level in the GLCM window. The HOM calculated from the PC1 image reflects the uniformity of rice growth. Higher HOM values indicate more uniform rice growth, which is typically associated with higher AGB, consistent with the findings in previous study (Wang F. et al., 2021). The HOM calculated from the PC2 image reflects the relatively smooth portion of high-frequency information, which represents the information on rice plant or the soil background in mixed pixels. Studies by Wang W. et al. (2021) and Liu S. et al. (2023) respectively demonstrate that adjusted vegetation indices considering rice green plants abundance information or soil background in mixed pixel, such as adjusted abundance vegetation index (AAVI), and adjusted vegetation indices considering soil background (VI_{CS}), contribute to improving the estimation accuracy of key crop growth parameters. This could be a potential reason for the importance scores of PC2HOM. However, this conclusion has not been confirmed for AGB estimation and will be one of our future research endeavors. SM shares similar physical implications with HOM, while Ent exhibits contrasting characteristics to HOM. Hence, both SM and Ent attain significant importance scores in estimating rice AGB.

4.4 Difference for rice AGB estimation between SFs and TFs

SFs derived from UAVs-based multispectral images are widely used in crop biophysical parameters estimation (Han et al., 2019; Maimaitijiang et al., 2019; Wang F. et al., 2022; Wang et al., 2023).

This study compared the RF models' performance between SFs and TFs for estimating the rice AGB (Tables 3–5). The result showed that the different performance between SFs and TFs was stage-specific. During the early stages of crop growth, the high leaf-stem density plays a significant role in determining a substantial portion of the AGB (Zhu et al., 2023). The SFs and TFs can reflect the reflectance attribute and spatial variation of rice canopy well, and the correlation between them is high (Bai et al., 2021). As a result, this explains the comparable relationship between these features when estimating AGB. For the mid-growth stages, the complex nature of the canopy during this stage makes it challenging for SFs alone to capture the spatial variability and intricate details, whereas TFs excel in capturing these fine-grained characteristics (Liu Y. et al., 2019; Yue et al., 2019; Zhang et al., 2021). Consequently, the superior performance of TFs in estimating AGB is observed during the mid-growth stages, particularly during the HtF stage. At the EF stage, the dense rice canopy imposes certain limitations on the representation of spatial heterogeneity by TFs, even in the presence of large windows.

However, in contrast to previous findings (Dube and Mutanga, 2015; Liu et al., 2018; Liang et al., 2022; Xu L. et al., 2022), the fused features did not exhibit superiority in estimating AGB, which may be attributed to that this study exclusively utilized band reflectance instead of vegetation indices, information overlap occurred between spectral reflectance and TFs, resulting in limited improvement in the predictive accuracy of rice AGB through feature fusion, which matched the findings of Mao et al. (2021). In addition, variations in the results can occur due to different crop types and regression algorithms used (Liu et al., 2018; Liang et al., 2022; Liu et al., 2022b; Xu L. et al., 2022; Liu Y. et al., 2023).

4.5 Limitations and directions of future work

Our findings provided compelling evidence of the influence of texture window size and direction on the rice AGB estimation using GLCM-based TFs. However, it is crucial to further validate its general applicability in terms of spatial and temporal transferability. In the future, the impact of texture parameters on crop AGB estimation still needs to be tested across diverse ecological conditions, multiple rice genotypes, and various crop types. Compared to high-resolution RGB images, multispectral imagery generally has lower spatial resolution (Lu et al., 2019; Liang et al., 2022; Gan et al., 2023). Therefore, further research is needed to investigate whether the conclusions based on multispectral imagery also apply to high-resolution RGB images.

Although the RF algorithm has strong regression capabilities (Liu Y. et al., 2019; Burdett and Wellen, 2022; Borrmann et al., 2023), this study has highlighted the potential issues of overfitting when dealing with small sample size. In order to improve the reliability of the RF model, it is essential to conduct additional field experiments across various locations, years, and crop varieties for further evaluation. Surprisingly, unlike previous studies, the incorporation of fused features did not notably improve the accuracy of AGB estimation (Dube and Mutanga, 2015; Liu et al., 2018; Liang et al., 2022; Xu L. et al., 2022). In addressing this issue in future research, two potential approaches can be considered. Firstly, before training the estimation models, an analysis can be conducted to assess the correlation among the input predictor variables, utilizing techniques such as Variance Inflation Factor (VIF) or feature selection methods. Secondly, the inclusion of modeling techniques with non-linear structures, such as Support Vector Regression (SVR), Extreme Learning Machines (ELM), and XGBoost, can be explored extensively to evaluate their impact on the estimation results.

5 Conclusions

Accurately assessment of AGB could provide valuable insights into the estimation and management of crop health and productivity. To resolve the saturation issue of spectral information, TFs were introduced to bridge this gap. This study provided a comprehensive evaluation of how GLCM-based TFs with different window size and direction parameters influence the accuracy of rice AGB estimation. The findings revealed that the appropriate window size for extracting TFs varies according to the rice growth stage, highlighting the need to incorporate multi-scale texture to capture the spatial variations of the rice canopy throughout the growing season. Additionally, the diagonal direction at 45° (D2) was identified as the optimal direction for estimating AGB. The important features of rice AGB estimation were Con, Dis, and Cor, which are mainly derived from the PC2 principal component image, which can better capture edge information. TFs were served as a valuable alternative or complement to spectral features, demonstrating estimation accuracy comparable to spectral reflectance for rice AGB estimation. These findings might help identify the best configuration of GLCM parameters to enhance the accuracy of estimating AGB, which can provide valuable insights for efficient monitoring of crop information 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

JLiu: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Writing – original draft, Writing – review & editing. YZ: Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. LS: Funding acquisition, Methodology, Project administration, Software, Supervision, Writing – review & editing. XS: Formal analysis, Investigation, Visualization, Writing – review & editing. JLi: Formal analysis, Investigation, Writing – review & editing. JZ: Formal analysis, Software, Writing – review & editing. XZ: Investigation, Writing – review & editing. LR: Formal analysis, Writing – review & editing. WW: Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Writing –

review & editing. XL: Conceptualization, Funding acquisition, Investigation, Project administration, 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 funded by scientific research projects in higher education institutions of Anhui Province (no. 2022AH051623; 2023AH051855); Provincial Scientific Research Service Expense Project (no. CZKYF2021-2-B010); Anhui Province Crop Intelligent Planting and Processing Technology Engineering Research Center Open Research Project(no. ZHZZKF202306); Natural Science Foundation of Hebei Province (no. C2020408006; C2023408010), and College Students' Innovation and Entrepreneur ship Training Project (no. 202210879043).

Acknowledgments

We are very grateful to Xiaofang Chen, Ying Nian, Yongkang Li, and Weiqiang Wang for their assistance in data collection.

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 collaboration WW 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.2023.1284235/full#supplementary-material

References

Alebele, Y., Zhang, X., Wang, W., Yang, G., Yao, X., Zheng, H., et al. (2020). Estimation of canopy biomass components in paddy rice from combined optical and sar data using multi-target gaussian regressor stacking. *Remote Sensing.* 12, 2564. doi: 10.3390/rs12162564

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. Agriculture.* 190, 106456. doi: 10.1016/j.compag.2021.106456

Bao, Y., Gao, W., and Gao, Z. (2009). Estimation of winter wheat biomass based on remote sensing data at various spatial and spectral resolutions. *Front. Earth Sci. China.* 3, 118–128. doi: 10.1007/s11707-009-0012-x

Borrmann, P., Brandt, P., and Gerighausen, H. (2023). Mispel: a multi-crop spectral library for statistical crop trait retrieval and agricultural monitoring. *Remote Sensing.* 15, 3664. doi: 10.3390/rs15143664

Breiman, L. (2001). Random forests. *Mach. Learning.* 45, 5–32. doi: 10.1023/A:1010933404324

Burdett, H., and Wellen, C. (2022). Statistical and machine learning methods for crop yield prediction in the context of precision agriculture. *Precis. Agriculture.* 23, 1553–1574. doi: 10.1007/s11119-022-09897-0

Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Zhang, J., et al. (2021). Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches. *Agric. For. Meteorology.* 297, 108275. doi: 10.1016/j.agrformet.2020.108275

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 Sensing.* 9, 319. doi: 10.3390/rs9040319

Cui, Z., and Kerekes, J. P. (2018a). Potential of red edge spectral bands in future landsat satellites on agroecosystem canopy green leaf area index retrieval. *Remote Sensing.* 10 (9), 1458. doi: 10.3390/rs10091458

Cui, Z., and Kerekes, J. P. (2018b). Impact of wavelength shift in relative spectral response at high angluaves of incidence in landsat-8 operational land imager and future landsat design concepts. *IEEE Trans. Geosci. Remote Sensing.* 56, 5873–5883. doi: 10.1109/tgrs.2018.2827394

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

Dube, T., and Mutanga, O. (2015). Investigating the robustness of the new landsat-8 operational land imager derived texture metrics in estimating plantation forest aboveground biomass in resource constrained areas. *ISPRS J. Photogrammetry Remote Sensing.* 108, 12–32. doi: 10.1016/j.isprsjprs.2015.06.002

Fu, Y., Yang, G., Li, Z., Song, X., Li, Z., Xu, X., et al (2020). Winter Wheat Nitrogen Status Estimation Using UAV-Based RGB Imagery and Gaussian Processes Regression. *Remote Sens.* 12, 3778. doi: 10.3390/rs12223778

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 Sensing.* 13, 581. doi: 10.3390/rs13040581

Fu, Y., Yang, G., Wang, J., Song, X., and Feng, H. (2014). Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyperspectral measurements. *Comput. Electron. Agriculture.* 100, 51–59. doi: 10.1016/j.compag.2013.10.010

Gan, Y., Wang, Q., Matsuzawa, T., Song, G., and Iio, A. (2023). Multivariate regressions coupling colorimetric and textural features derived from UAV-based rgb images can trace spatiotemporal variations of lai well in a deciduous forest. *Int. J. Remote Sensing.* 44, 4559–4577. doi: 10.1080/01431161.2023.2208709

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

Guo, Y., Fu, Y. H., Chen, S., Robin Bryant, C., Li, X., Senthilnath, J., et al. (2021). Integrating spectral and textural information for identifying the tasseling date of summer maize using UAV based RGB images. *Int. J. Appl. Earth Observation Geoinformation.* 102, 102435. doi: 10.1016/j.jag.2021.102435

Hall-Beyer, M. (2007) *GLCM Texture: A Tutorial v. 1.0 through 2.7*. Available at: http://www.fp.ucalgary.ca/mhallbey/tutorial.htm.

Hall-Beyer, M. (2017). Practical guidelines for choosing glcm textures to use in landscape classification tasks over a range of moderate spatial scales. *Int. J. Remote Sensing.* 38, 1312–1338. doi: 10.1080/01431161.2016.1278314

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

Haralick, R. M., Shanmugam, K., and Dinstein, I. H. (1973). Textural features for image classification. *IEEE Trans. On Systems Man Cybernetics*. 3, 610–621. doi: 10.1109/TSMC.1973.4309314

Jangra, S., Chaudhary, V., Yadav, R. C., and Yadav, N. R. (2021). High-throughput phenotyping: a platform to accelerate crop improvement. *Phenomics*. 1, 31–53. doi: 10.1007/s43657-020-00007-6

Jin, X., Zarco-Tejada, P., Schmidhalter, U., Reynolds, M. P., Hawkesford, M. J., Varshney, R. K., et al. (2020). High-throughput estimation of crop traits: a review of ground and aerial phenotyping platforms. *IEEE Geosci. Remote Sens. Magazine* 9 (1), 200–231. doi: 10.1109/MGRS.2020.2998816

Kelsey, K. C., and Neff, J. C. (2014). Estimates of aboveground biomass from texture analysis of landsat imagery. *Remote Sensing*. 6, 6407–6422. doi: 10.3390/rs6076407

Lakhiar, I. A., Jianmin, G., Syed, T. N., Chandio, F. A., Buttar, N. A., and Qureshi, W. A. (2018). Monitoring and control systems in agriculture using intelligent sensor techniques: a review of the aeroponic system. *J. Sensors*. 2018, 1–18. doi: 10.1155/2018/8672769

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 Sensing*. 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 Sensing*. 11, 1763. doi: 10.3390/rs11151763

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

Liang, Y., Kou, W., Lai, H., Wang, J., Wang, Q., Xu, W., et al. (2022). Improved estimation of aboveground biomass in rubber plantations by fusing spectral and textural information from UAV-based RGB imagery. *Ecol. Indicators*. 142, 109286. doi: 10.1016/j.ecolind.2022.109286

Liao, C., Wang, J., Dong, T., Shang, J., Liu, J., and Song, Y. (2019). Using spatio-temporal fusion of landsat-8 and modis data to derive phenology, biomass and yield estimates for corn and soybean. *Sci. Total Environment*. 650, 1707–1721. doi: 10.1016/j.scitotenv.2018.09.308

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. Agriculture*. 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 Science*. 13. 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. Agriculture*. 198, 107089. doi: 10.1016/j.compag.2022.107089

Liu, S., Jin, X., Bai, Y., Wu, W., Cui, N., Cheng, M., et al. (2023). UAV multispectral images for accurate estimation of the maize LAI considering the effect of soil background. *Int. J. Appl. Earth Observation Geoinformation*. 121, 103383. doi: 10.1016/j.jag.2023.103383

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. Agriculture*. 166, 105026. doi: 10.1016/j.compag.2019.105026

Liu, C., Yang, G., Li, Z., Tang, F., Feng, H., Wang, J., et al. (2019). "Monitoring of winter wheat biomass using UAV hyperspectral texture features," in. Li, D., and Zhao, C. Computer and Computing Technologies in Agriculture XI. CCTA 2017. IFIP Advances in Information and Communication Technology, vol 546. Springer, Cham. doi: 10.1007/978-3-030-06179-1_25

Liu, C., Yang, G., Li, Z., Tang, F., Wang, J., Zhang, C., et al. (2018). Biomass estimation in winter wheat by UAV spectral information and texture information fusion. *Scientia Agricultura Sin*. 51, 3060–3073. doi: 10.3864/j.issn.0578-1752.2018.16.003

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, D. (2005). Aboveground biomass estimation using landsat tm data in the Brazilian amazon. *Int. J. Remote Sensing*. 26, 2509–2525. doi: 10.1080/01431160500142145

Lu, N., Zhou, J., Han, Z., Li, D., Cao, Q., Yao, X., et al. (2019). Improved estimation of aboveground biomass in wheat from rgb imagery and point cloud data acquired with a low-cost unmanned aerial vehicle system. *Plant Methods* 15, 1–16. doi: 10.1186/s13007-019-0402-3

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 Science*. 13. doi: 10.3389/fpls.2022.948249

Maes, W. H., and Steppe, K. (2019). Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. *Trends Plant Science*. 24, 152–164. doi: 10.1016/j.tplants.2018.11.007

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 Sensing*. 151, 27–41. doi: 10.1016/j.isprsjprs.2019.03.003

Mao, P., Qin, L., Hao, M., Zhao, W., Luo, J., Qiu, X., et al. (2021). An improved approach to estimate above-groundvolume and biomass of desert shrub communities based on UAV RGB images. *Ecol. Indicators*. 125, 107494. doi: 10.1016/j.ecolind.2021.107494

Marceau, D. J., Howarth, P. J., Dubois, J. M., and Gratton, D. J. (1990). Evaluation of the grey-level co-occurrence matrix method for land-cover classification using spot imagery. *IEEE Trans. On Geosci. Remote Sensing*. 28, 513–519. doi: 10.1109/TGRS.1990.572937

Mercier, A., Betbeder, J., Baudry, J., Le Roux, V., Spicher, F., Lacoux, J., et al. (2020a). Evaluation of sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages. *ISPRS J. Photogrammetry Remote Sensing*. 163, 231–256. doi: 10.1016/j.isprsjprs.2020.03.009

Mercier, A., Betbeder, J., Rapinel, S., Jegou, N., Baudry, J., and Hubert-Moy, L. (2020b). Evaluation of sentinel-1 and -2 time series for estimating lai and biomass of wheat and rapeseed crop types. *J. Appl. Remote Sensing*. 14, 1. doi: 10.1117/1.JRS.14.024512

Mukherjee, A., Misra, S., and Raghuwanshi, N. S. (2019). A survey of unmanned aerial sensing solutions in precision agriculture. *J. Network Comput. Applications*. 148, 102461. doi: 10.1016/j.jnca.2019.102461

Naidoo, L., Main, R., Cho, M. A., Madonsela, S., and Majozi, N. (2021). "Estimating South African maize biomass using integrated high-resolution UAV and sentinel 1 and 2 datasets," 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 2021,1594-1596. doi: 10.1109/IGARSS47720.2021.9554261

Pacifici, F., Chini, M., and Emery, W. J. (2009). A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sens. Environment*. 113, 1276–1292. doi: 10.1016/j.rse.2009.02.014

Peng, S., Tang, Q., and Zou, Y. (2009). Current status and challenges of rice production in China. *Plant Production Science*. 12, 3–8. doi: 10.1626/pps.12.3

Qiu, R., Wei, S., Zhang, M., Li, H., Sun, H., Liu, G., et al. (2018). Sensors for measuring plant phenotyping: a review. *Int. J. Agric. Biol. Engineering*. 11, 1–17. doi: 10.25165/j.ijabe.20181102.2696

Richter, K., Atzberger, C., Hank, T. B., and Mauser, W. (2012). Derivation of biophysical variables from earth observation data: validation and statistical measures. *J. Appl. Remote Sensing*. 6, 63557. doi: 10.1117/1.JRS.6.063557

Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., and Jeganathan, C. (2012). Random forest classification of mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sens. Environment*. 121, 93–107. doi: 10.1016/j.rse.2011.12.003

Seck, P. A., Diagne, A., Mohanty, S., and Wopereis, M. C. (2012). Crops that feed the world 7: rice. *Food Secur*. 4, 7–24. doi: 10.1007/s12571-012-0168-1

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

Teng, X., Dong, Y., and Meng, L. (2015). "The study of winter wheat biomass estimation model based on hyperspectral remote sensing," in Eds. D. Li and Z. Li *Computer and computing technologies in agriculture IX. CCTA 2015. IFIP advances in information and communication technology*, vol. 479 (Springer, Cham). doi: 10.1007/978-3-319-48354-2_17

Wang, D., Li, R., Zhu, B., Liu, T., Sun, C., and Guo, W. (2023). Estimation of wheat plant height and biomass by combining UAV imagery and elevation data. *Agriculture*. 13, 9. doi: 10.3390/agriculture13010009

Wang, Q., Putri, N. A., Gan, Y., and Song, G. (2022). Combining both spectral and textural indices for alleviating saturation problem in forest lai estimation using sentinel-2 data. *Geocarto Int*. 37, 10511–10531. doi: 10.1080/10106049.2022.2037730

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. Selected Topics Appl. Earth Observations Remote Sensing*. 14, 6716–6728. doi: 10.1109/JSTARS.2021.3086580

Wang, F., Yang, M., Ma, L., Zhang, T., Qin, W., Li, W., et al. (2022). Estimation of above-ground biomass of winter wheat based on consumer-grade multi-spectral UAV. *Remote Sensing*. 14, 1251. doi: 10.3390/rs14051251

Wang, F., Yi, Q., Hu, J., Xie, L., Yao, X., Xu, T., et al. (2021). Combining spectral and textural information in UAV hyperspectral images to estimate rice grain yield. *Int. J. Appl. Earth Observation Geoinformation*. 102, 102397. doi: 10.1016/j.jag.2021.102397

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 Crops Res*. 283, 108543. doi: 10.1016/j.fcr.2022.108543

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., 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 Sensing.* 14, 2534. doi: 10.3390/rs14112534

Xu, L., Zhou, L., Meng, R., Zhao, F., Lv, Z., Xu, B., et al. (2022). An improved approach to estimate ratoon rice aboveground biomass by integrating UAV-based spectral, textural and structural features. *Precis. Agriculture.* 23, 1276–1301. doi: 10.1007/s11119-022-09884-5

Yang, J., Ding, F., Chen, C., Liu, T., Sun, C., Ding, D., et al. (2019). Correlation of wheat biomass and yield with UAV image characteristic parameters. *Trans. Chin. Soc. Agric. Engineering.* 35, 104–110. doi: 10.11975/j.issn.1002-6819.2019.23.013

Yu, N., Li, L., Schmitz, N., Tian, L. F., Greenberg, J. A., and Diers, B. W. (2016). Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. *Remote Sens. Environment.* 187, 91–101. doi: 10.1016/j.rse.2016.10.005

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 Sensing.* 150, 226–244. doi: 10.1016/j.isprsjprs.2019.02.022

Zhang, D. Y., Han, X. X., Lin, F. F., Du, S. Z., Zhang, G., and Hong, Q. (2022). Estimation of winter wheat leaf area index using multi-source UAV image feature fusion. *Trans. Chin. Soc. Agric. Eng.* 38, 171–179. doi: 10.11975/j.issn.1002-6819.2022.09.018

Zhang, C., Huang, C., Li, H., Liu, Q., Li, J., Bridhikitti, A., et al. (2020). Effect of textural features in remote sensed data on rubber plantation extraction at different levels of spatial resolution. *Forests.* 11, 399. doi: 10.3390/f11040399

Zhang, J., Qiu, X., Wu, Y., Zhu, Y., Cao, Q., Liu, X., et al. (2021). Combining texture, color, and vegetation indices from fixed-wing uas imagery to estimate wheat growth parameters using multivariate regression methods. *Comput. Electron. Agriculture.* 185, 106138. doi: 10.1016/j.compag.2021.106138

Zhang, Y., Ta, N., Guo, S., Chen, Q., Zhao, L., Li, F., et al. (2022). Combining spectral and textural information from UAV RGB images for leaf area index monitoring in kiwifruit orchard. *Remote Sensing.* 14, 1063. doi: 10.3390/rs14051063

Zheng, H., Cheng, T., Li, D., Yao, X., Tian, Y., Cao, W., et al. (2018). Combining unmanned aerial vehicle (UAV)-based multispectral imagery and ground-based hyperspectral data for plant nitrogen concentration estimation in rice. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00936

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. Agriculture.* 20, 611–629. doi: 10.1007/s11119-018-9600-7

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 Sensing.* 12, 957. doi: 10.3390/rs12060957

Zhou, J., Yan, G., Sun, M., Di, T., Wang, S., Zhai, J., and Zhao, Z. (2017). The effects of GLCM parameters on LAI estimation using texture values from Quickbird satellite imagery. *Sci. Rep.* 7, 7366. doi: 10.1038/s41598-017-07951-w

Zhou, L., Nie, C., Su, T., Xu, X., Song, Y., Yin, D., et al. (2023). Evaluating the canopy chlorophyll density of maize at the whole growth stage based on multi-scale UAV image feature fusion and machine learning methods. *Agriculture.* 13, 895. doi: 10.3390/agriculture13040895

Zhou, M., Zheng, H., He, C., Liu, P., Awan, G. M., Wang, X., et al. (2023). Wheat phenology detection with the methodology of classification based on the time-series UAV images. *Field Crops Res.* 292, 108798. doi: 10.1016/j.fcr.2022.108798

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 Sensing.* 15, 3332. doi: 10.3390/rs15133332

Zhu, W., Rezaei, E. E., Nouri, H., Sun, Z., Li, J., Yu, D., et al. (2022). UAV-based indicators of crop growth are robust for distinct water and nutrient management but vary between crop development phases. *Field Crops Res.* 284, 108582. doi: 10.1016/j.fcr.2022.108582

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 January 2024
DOI 10.3389/fpls.2023.1272049

Check for updates

OPEN ACCESS

EDITED BY
Bo Li,
Syngenta, United Kingdom

REVIEWED BY
Haikuan Feng,
Beijing Research Center for Information Technology in Agriculture, China
Liujun Li,
University of Idaho, United States
Karansher Singh Sandhu,
Bayer Crop Science, United States
Bolin Fu,
Guilin University of Technology, China

*CORRESPONDENCE
Cunjun Li
licj@nercita.org.cn
Yu Liu
liuyu@nercita.org.cn

RECEIVED 03 August 2023
ACCEPTED 08 December 2023
PUBLISHED 03 January 2024

CITATION
Zou M, Liu Y, Fu M, Li C, Zhou Z, Meng H, Xing E and Ren Y (2024) Combining spectral and texture feature of UAV image with plant height to improve LAI estimation of winter wheat at jointing stage. *Front. Plant Sci.* 14:1272049. doi: 10.3389/fpls.2023.1272049

COPYRIGHT
© 2024 Zou, Liu, Fu, Li, Zhou, Meng, Xing 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.

Combining spectral and texture feature of UAV image with plant height to improve LAI estimation of winter wheat at jointing stage

Mengxi Zou[1,2], Yu Liu[2*], Maodong Fu[3], Cunjun Li[2,4*], Zixiang Zhou[1], Haoran Meng[2], Enguang Xing[2] and Yanmin Ren[2]

[1]College of Geomatics, Xi'an University of Science and Technology, Xi'an, China, [2]Research Center of Information Technology, Beijing Academy of Agriculture and Forestry Science, Beijing, China, [3]Hebei Maodong Xingteng Agricultural Technology Service Co., Ltd, Cangzhou, China, [4]Qingyuan Smart Agriculture and Rural Research Institute, Qingyuan, China

Introduction: Leaf area index (LAI) is a critical physiological and biochemical parameter that profoundly affects vegetation growth. Accurately estimating the LAI for winter wheat during jointing stage is particularly important for monitoring wheat growth status and optimizing variable fertilization decisions. Recently, unmanned aerial vehicle (UAV) data and machine/depth learning methods are widely used in crop growth parameter estimation. In traditional methods, vegetation indices (VI) and texture are usually to estimate LAI. Plant Height (PH) unlike them, contains information about the vertical structure of plants, which should be consider.

Methods: Taking Xixingdian Township, Cangzhou City, Hebei Province, China as the research area in this paper, and four machine learning algorithms, namely, support vector machine(SVM), back propagation neural network (BPNN), random forest (RF), extreme gradient boosting (XGBoost), and two deep learning algorithms, namely, convolutional neural network (CNN) and long short-term memory neural network (LSTM), were applied to estimate LAI of winter wheat at jointing stage by integrating the spectral and texture features as well as the plant height information from UAV multispectral images. Initially, Digital Surface Model (DSM) and Digital Orthophoto Map (DOM) were generated. Subsequently, the PH, VI and texture features were extracted, and the texture indices (TI) was further constructed. The measured LAI on the ground were collected for the same period and calculated its Pearson correlation coefficient with PH, VI and TI to pick the feature variables with high correlation. The VI, TI, PH and fusion were considered as the independent features, and the sample set partitioning based on joint x-y distance (SPXY) method was used to divide the calibration set and validation set of samples.

Results: The ability of different inputs and algorithms to estimate winter wheat LAI were evaluated. The results showed that (1) The addition of PH as a feature variable significantly improved the accuracy of the LAI estimation, indicating that wheat plant height played a vital role as a supplementary parameter for LAI inversion modeling based on traditional indices; (2) The combination of texture

features, including normalized difference texture indices (NDTI), difference texture indices (DTI), and ratio texture indices (RTI), substantially improved the correlation between texture features and LAI; Furthermore, multi-feature combinations of VI, TI, and PH exhibited superior capability in estimating LAI for winter wheat; (3) Six regression algorithms have achieved high accuracy in estimating LAI, among which the XGBoost algorithm estimated winter wheat LAI with the highest overall accuracy and best results, achieving the highest R^2 (R^2 = 0.88), the lowest RMSE (RMSE=0.69), and an RPD greater than 2 (RPD=2.54).

Discussion: This study provided compelling evidence that utilizing XGBoost and integrating spectral, texture, and plant height information extracted from UAV data can accurately monitor LAI during the jointing stage of winter wheat. The research results will provide a new perspective for accurate monitoring of crop parameters through remote sensing.

KEYWORDS

plant height, feature fusion, machine learning, deep learning, UAV, LAI, winter wheat

1 Introduction

Winter wheat is the second-largest grain crop in China in terms of cultivated area and total output (Han, 2011), and it holds significant economic value. Investigating the agronomic parameters of winter wheat is essential to agricultural production management, especially to enhance grain production. Leaf Area Index (LAI) stands as a crucial agronomic parameter for winter wheat, which is defined as the ratio of total plant leaf area per unit of land area to land area. LAI is directly related to crop growth (Casa et al., 2012) and serves as a vital indicator for monitoring crop growth, biomass estimation, and pre-harvest yield prediction during the fertility period (Pinter et al., 2003; Dente et al., 2008). In the context of production management, wheat topdressing during the jointing stage is pivotal to improving yield and quality. China has implemented a zero-growth policy for chemical fertilizers and pesticides (Cui et al., 2021), alongside developing variable rate fertilizer applicators to enable precise fertilizer application based on local conditions. Therefore, the accurate and rapid estimation of LAI for winter wheat at jointing stage is not only conducive to the real-time monitoring of crop growth and development, but also has important significance for the formulation of variable rate fertilization prescription for agricultural machinery, reducing the use of chemical fertilizer and mitigating soil pollution.

The measurement methods of LAI include direct and indirect approaches. The direct method is a traditional and destructive method, mainly through manual field observation, which is time-consuming and laborious (Hu et al., 2018). The indirect method employs optical instruments or remote sensing inversion, offering a convenient and fast approach (Torres-S´anchez et al., 2014). Among these methods, remote sensing technology has gained widespread adoption as an indirect means of monitoring agronomic parameters (Zhang et al., 2008). Remote sensing platforms can be categorized into ground, aerial and space remote sensing platform based on their height above the ground (Lu et al., 2022). Ground remote sensing platforms mainly employ spectrograph for measurement. However, due to the height restrictions, these platforms face the challenge of obtaining Digital Orthophoto Maps (DOM) and monitor large-scale areas efficiently (Behrens and Diepenbrock, 2006; Nie et al., 2016). Space remote sensing platforms mainly rely on satellites to acquire data, enabling the monitoring large areas. Nevertheless, factors such as satellite revisit time and atmospheric conditions often hinder meeting the demands for spatiotemporal resolution (Liu et al., 2012). An alternative approach is the aerial operation mode, which employs unmanned aerial vehicle (UAV) (Li et al., 2019) for remote sensing. Compared to manned aircraft, UAV remote sensing has the advantages of low cost, simple operation and strong flexibility (Hassan et al., 2019). Furthermore, the multispectral sensors carried by UAV provide more bands than digital camera sensors, and the spectral information avoids data redundancy seen in hyperspectral sensors, which can be effectively applied to monitor crop LAI (Sun et al., 2019).

Currently, there are two implementations for monitoring crop LAI using UAV multispectral remote sensing. One is a radiative transfer physical model, and the other is a statistical empirical model. The physical model is based on the reflection and absorption between light and crops, which has a certain mechanism and strong versatility (Fu et al., 2022a). However, the model involves complex formulas and requires many parameters, which makes it difficult to

find the optimal solution (Du et al., 2016). In contrast, the empirical model establishes the relationship between UAV image features and winter wheat growth parameters through statistical methods (Liang et al., 2015; Cao et al., 2020). This method proves to be straightforward and user-friendly, estimating LAI by analyzing the statistical relationship between raw spectra or extracted vegetation indices and ground-based measured LAI data. Nonetheless, apart from spectral features, UAV multispectral images also provide abundant texture information associated with vegetation growth (Kupidura, 2019). Texture information reflect inherent characteristics of the image, making it essential to consider them when extracting vegetation growth parameters from multispectral images. Some scholars have successfully estimated LAI of rice (Cao et al., 2022), potato (Li et al., 2023), sorghum (Potgieter et al., 2017) and other crops, as well as biomass (Dai et al., 2022) and yield (Fu et al., 2020) of winter wheat by integrating spectral and texture features, providing promising results. However, previous studies have mostly directly input a large number of texture features into the model for training, lacking the optimization of texture features. To address this limitation, a more refined and selective approach should be considered to enhance the effectiveness and accuracy of integrating texture information into the model (Zhu et al., 2019). Texture features provide valuable information about the small-scale structures and details in an image. The optimization of extraction methods plays a pivotal role in achieving a more precise capture of structural information within the image, consequently elevating the accuracy of image analysis. Simultaneously, this optimization process contributes to the enhancement of the level of detail present within the image, rendering it richer in information content and enhancing the effectiveness and accuracy of integrating texture information into the model. For instance, the optimized texture indices can simultaneously capture the influence of two distinct texture features on wheat LAI monitoring (Zhang et al., 2022b).

The spectral and texture information extracted from the UAV multispectral image only contains the crop canopy information. During the period of crop growth, the vegetation indices are not sensitive to the changes in canopy information, and the spectral signal may become saturated (Liu et al., 2018), thereby affecting the accuracy of LAI estimation to some extent. Recent studies have revealed a significant correlation between crop canopy height and LAI (Liu et al., 2022). Notably, extracting plant height from UAV based on canopy height model (CHM) can alleviate the issue of spectral saturation. Niu et al. (Niu et al., 2018) have demonstrated improvements in LAI estimation accuracy by fusing plant height data with UAV digital image variables for maize breeding materials, surpassing the performance of using only digital image variables. Gao et al. (Gao et al., 2020) have successfully enhanced the inversion accuracy of LAI by combining crop height parameters with vegetation indices. Therefore, it is necessary to incorporate crop vertical structure characteristics, such as height, into the inversion process of winter wheat LAI to get more accurate and reliable results.

Generally speaking, researchers commonly use empirical or semi empirical models, employing statistical regression analysis of spectral features, to construct LAI estimation models. With the development of crop LAI inversion research, some researchers have begun exploring the utilization of machine learning (ML) techniques to build LAI inversion model and enhance estimation accuracy. Partial Least Square Regression (PLSR) (Hasan et al., 2019), Support Vector Machine (SVM) (Azadbakht et al., 2019), Random Forest (RF) (Zhang et al., 2018), Extreme Gradient Boosting (XGBoost) (Zhang et al., 2021c), and other machine learning algorithms are commonly used for wheat LAI inversion. Furthermore, deep learning (DL), a sophisticated machine learning algorithm, has gained traction in crop yield estimation and prediction, with convolutional neural network (CNN) and recursive neural network (RNN) being widely applied in related studies (Khaki et al., 2019; Koirala et al., 2019). Long short-term memory (LSTM) neural network is an improved RNN with a special recursive structure and gating mechanism, which can adjust the information in and out of the unit, and has high prediction accuracy in the field of wheat yield and biomass estimation (Wang et al., 2022a). These ML-based methods operate on diverse model frameworks, through learning from the training set data to construct the inversion model, thereby establishing the relationship between predictor variables and response variables. Leveraging robust data analysis capabilities and achieving high estimation accuracy, ML approaches effectively circumvent the shortcomings of empirical or semi-empirical models prone to pathological issues (Pearson and Miller, 1972).

The above studies have achieved high estimation accuracy using various methods, which the R^2 of the optimal models was range from 0.74 to 0.78. However, there remains a lack of research in exploring the potential of using ML/DL to improve the accuracy of winter wheat LAI estimation by combining spectral features, optimized texture features, and plant height based on UAV multispectral image. In light of this, the main objectives of the study were (1) to examine the influence of plant height on winter wheat LAI estimation during the jointing stage; (2) to evaluate the disparities in multi-feature estimation of winter wheat LAI with combinations of vegetation indices (VI), texture indices (TI), and plant height (PH); and (3) to conduct a comparative analysis of performances in the inversion modeling of winter wheat LAI with six regression algorithms, including support vector machine (SVM), back propagation neural network (BPNN), random forest (RF), extreme gradient boosting (XGBoost), convolutional neural network (CNN) and long short-term memory neural network (LSTM). The achievement of these objectives will provide a more reliable foundation for winter wheat LAI estimation, offering more precise monitoring and decision support for agricultural production.

2 Materials and methods

2.1 Study area and experimental design

Geographic Location: The study area is located in Xixindian Township, Cangzhou City, Hebei Province, China. The experimental field spans a geographic range from 116°08'38" E to

116°10'42" E longitude, and 37°58'40" N to 37°59'09" N latitude (Figure 1).

Climate: This region exhibits a warm temperate continental monsoon climate, characterized by pleasant temperatures. The annual average temperature is 12.7°C. Abundant sunlight graces the area, with an average annual sunshine duration of approximately 2.78 h. Precipitation predominantly occurs during the summer months, totaling around 543 mm of rainfall per year. The favorable climatic conditions in the region provide conducive circumstances for the growth and development of winter wheat. The above statistics are based on Government of the People&'s Republic of China information published in 2022.

Topography: The topography is flat, providing an ideal environment for the cultivation of diverse crops. The topographical conditions may influence the distribution and drainage of water, thereby impacting the growth conditions of winter wheat. Different topographical features may result in variations in soil moisture across different regions, consequently affecting the estimation of LAI.

Cropping System: Notably, summer maize and winter wheat serve as the primary cash crops within the experimental area, employing a rotational planting system where winter wheat is planted after the maize harvest.

Varieties and Practices: Due to the lack of standardized management practices among winter wheat farmers in the region, there is a significant variation in the selection of winter wheat varieties, irrigation levels, and fertilizer application. Different characteristics among varieties may lead to variations in LAI. Because distinct winter wheat varieties may exhibit differences in growth rates, leaf quantities, and structures, thereby influencing the estimation of their LAI. In this paper, the prominent winter wheat varieties cultivated include Jimai 22, Jimai 518, Luyuan 502, Shandong 20, and Tumai 6, each with distinct seeding rates ranging from 225 kg/ha to 300 kg/ha. The predominant irrigation methods employed are surface irrigation and sprinkler irrigation. In addition, different types and ratios of fertilizers may significantly impact leaf growth, leaf area, and photosynthesis. For example, fertilizers with high nitrogen content may stimulate leaf growth, consequently affecting the estimated values of LAI. The primary fertilizers utilized are Tuboshi (N-P_2O_5-K_2O, 28-6-6) with ≥40% total nutrients, Jindadi (N-P_2O_5-K_2O, 19-15-6) with ≥40% total nutrients, and Xishouliang (N-P_2O_5-K_2O, 17-23-5) with ≥45% total nutrients. Across all varieties, the standard fertilizer application rate during cultivation is 600 kg/ha.

2.2 Data collection and preprocessing

2.2.1 UAV image acquisition and pre-processing

In this experiment, due to its compatibility with the research objectives, the eBee SQ precision agriculture UAV equipped with the Parrot Sequoia multispectral sensor was selected. The Parrot Sequoia sensor was capable of capturing spectral data in four bands simultaneously, including green (550nm), red (660nm), red edge (735nm), and near-infrared (790nm) (Handique et al., 2017). Prior to the experiment, it was crucial for the success of the study to conduct flight planning using specialized software to determine the flight routes and parameters, including a flight altitude of 95.5m, a ground resolution of 9cm and an 80% overlap in both along-track and across-track directions. The UAV was launched using a hand-throw method, and before take off, take photos of the radiation correction plate and ensure the absence of shadows on the calibration board. The data acquisition was conducted on October 21, 2020 (at bare soil stage) and March 31, 2021 (at jointing stage) between 10:00 and 14:00 to reduce the influence of the changes in the solar altitude angle on the experiments. In order to minimize optical distortions and ensure clear image acquisition, the weather conditions were clear skies, without any clouds, and a gentle breeze throughout the entire data collection period.

The acquired UAV multispectral remote sensing data were preprocessed, which primarily included image correction, image mosaic, image clipping and image resampling. Pix4DMapper and ArcGIS 10.4 are widely-used software tools in the fields of Remote Sensing (RS) and UAV data processing. In this study, the UAV multispectral images were geometrically corrected and stitched with Pix4DMapper software to generate Digital Surface Model (DSM)

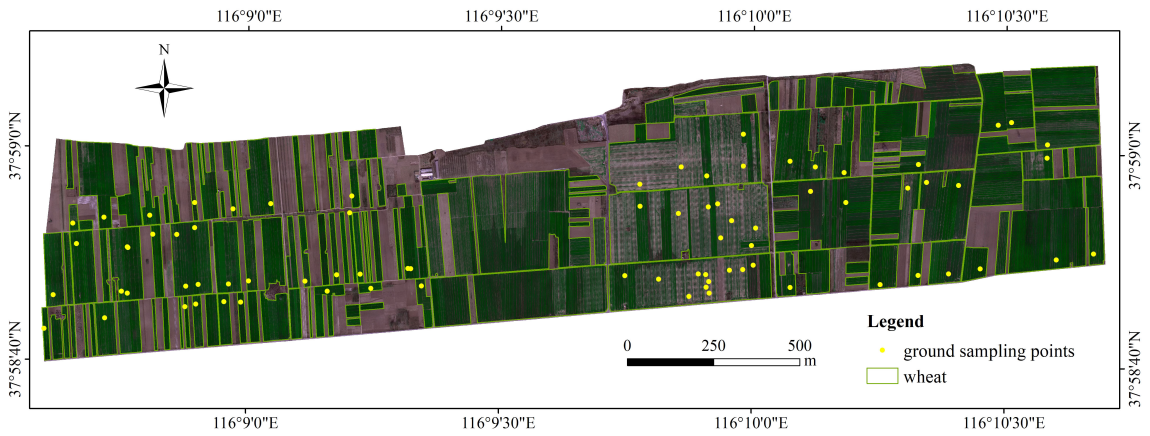


FIGURE 1
Study area of winter wheat LAI estimation experiment using UAV images.

and Digital Orthophoto Map (DOM), so as to obtain the spectral reflectance data and height data of the study area. After clipping based on the vector boundaries of the study area in ArcGIS 10.4 software, the image resolution was resampled to 0.1m using the cubic convolution difference method, aiming to ensure a sufficiently high detail resolution to accurately capture vegetation features.

2.2.2 Ground data collection

The ground data collection was synchronized with the acquisition of UAV multispectral remote sensing data, and mainly included the value of PH and LAI, coordinate data of sample points and control points. The measurements of PH and LAI were primarily conducted for the establishment and validation of estimation models. The sample point coordinates involved recording the positions where PH and LAI were measured. On the other hand, control points were established to provide ground truth coordinates, facilitating the correlation with UAV data. Specifically, data were collected from 79 sampling points evenly distributed in the study area to ensure comprehensive coverage. This uniform distribution helped capture spatial variations in vegetation, enhancing the representativeness and reliability of the data.

Wheat is conventionally sown at approximately 15 cm intervals. This seeding technique plays a pivotal role in ensuring effective soil surface coverage, mitigating soil moisture evaporation, and consequently, fostering water conservation and yield augmentation. Therefore, during the measurement of PH(unit: cm), three representative wheat plants that encapsulate the comprehensive growth status at the sampling point are meticulously selected within a 50 cm radius around the designated point. A tape measure was used to measure the vertical height of each wheat plant, and the average value was taken as the PH of each sample point to ensure the data accuracy.

The measurement of LAI was conducted using the LAI-2200C plant canopy analyzer. The steps were taken to ensure data accuracy as follows. Prior to measurement, the instrument was aligned with the sun to determine the incident light intensity. During measurement, efforts were made to keep the instrument as horizontally aligned as possible. Three measurements were taken at each sampling point, and the average value was considered as the final LAI. LAI in the sample dataset ranged from 0.61 to 8.57, the average was 3.89 and the standard deviation was 1.93. Simultaneously, the HI-TARGET iRTK2 was employed for Real-Time Kinematic (RTK) measurements, allowing the acquisition of coordinates for each sample point and control point.

2.2.3 Winter wheat pixels extraction

The green, red, near-infrared and red edge bands of UAV multispectral images, as well as normalized difference vegetation index (NDVI), enhanced vegetable index 2 (EVI2), red edge optimized soil-adjusted vegetation index (REOSAVI) and optimized soil-adjusted vegetation index (OSAVI) were used to input into the random forest classifier to extract winter wheat pixels (Fu et al., 2022b). NDVI is widely applied to reflect vegetation growth conditions. EVI2 considers atmospheric correction and soil influences, providing more accurate monitoring in areas with high vegetation cover. REOSAVI is an index optimized for the red-edge band that accounts for soil influences. OSAVI is a soil-adjusted vegetation index optimized to reduce the impact of the soil surface. Its reduction of soil effects in high vegetation density environments contributes to a more accurate assessment of vegetation conditions. Random Forest classifier is capable of efficiently handling large-scale datasets, exhibiting high classification accuracy, and is particularly well-suited for pixel-level image classification tasks. This study comprehensively assessed the quality of pixel extraction using overall accuracy and the kappa coefficient. Overall accuracy measures the overall precision of the classification results, while the kappa coefficient provides sensitivity to random errors and omissions in the classification. These metrics reflect the accuracy of the extraction process. The overall accuracy and kappa coefficient were 98.74% and 91.21%, respectively. ENVI 5.3 is software designed for remote sensing data analysis and image processing. Then this process was implemented in ENVI 5.3, with all parameters set to their default values.

2.3 LAI Estimation input features

2.3.1 Vegetation indices

Vegetation indices (VI) are established by using the relationship between spectral data and various physical and chemical parameters of vegetation, which can effectively reflect the growth status of plants, and are widely used in the monitoring of physiological and biochemical parameters of plants. Generally, it is obtained by selecting two or more band reflectance data from spectral data and performing a series of combined operations such as addition, subtraction, multiplication and division. Compared with a single band, the band combination method is not only more sensitive to vegetation characteristics, but also can eliminate environmental background noise to a certain extent (Huete et al., 1985). At present, there are many kinds of vegetation indices, such as the difference vegetation index (DVI) and enhanced vegetable index2 (EVI2), which can control the impact of soil and environmental background. Simultaneously, red edge renormalized difference vegetation Index (RERDVI) primarily focuses on the structure and coverage of vegetation, aiding in understanding the spatial distribution and density of vegetation. Red edge chlorophyll Index (Clre) reflects variations in chlorophyll content within vegetation, serving to assess the growth status and overall health of the vegetation. Based on the previous research results, this paper selected 17 vegetation indices with great effect for retrieving wheat LAI, which were divided into greenness indices, structure indices and pigments indices, according to their main functions, as shown in Table 1. The size of the range of spectra measured at each sampling point was 10×10cm.

2.3.2 Texture indices

Richer texture information related to plant growth can be extracted from UAV multispectral images. Texture features are distinct from image attributes such as grayscale and color, it is represented by the grayscale distribution of pixels and their surrounding spatial neighbors. It is the reflection of the internal characteristics of plant on remote

TABLE 1 VI used in this study.

Vegetation Index		Calculations	References
Greenness Indices	Ratio Vegetation Index	$RVI = \frac{NIR}{R}$	Pearson and Miller, 1972
	Difference Vegetation Index	$DVI = NIR - R$	Becker and Choudhury, 1988
	Normalized Difference Vegetation Index	$NDVI = \frac{NIR - R}{NIR + R}$	Rouse et al., 1974
	Green Normalized Difference Vegetation Index	$GNDVI = \frac{NIR - G}{NIR + G}$	Gitelson et al., 1996
	Enhanced Vegetation Index2	$EVI2 = \frac{2.5 \times (NIR - R)}{NIR + 2.4 \times R + 1}$	Jiang et al., 2008
	Green Chlorophyll index	$CLgreen = \frac{NIR}{G - 1}$	Gitelson et al., 2006
Structure Indices	Modified simple ratio	$MSR = \frac{\frac{NIR}{R} - 1}{\sqrt{\frac{NIR}{R}} + 1}$	Chen and Cihlar, 1996
	Modified Soil Adjusted Vegetation Index	$MSAVI = 0.5 \times \left[\begin{array}{l} (2 \times NIR + 1) - \\ \left(\begin{array}{l} (2 \times NIR + 1)^2 \\ -8 \times (NIR - R) \end{array} \right)^{0.5} \end{array} \right]$	Daughtry et al., 2000
	Green Optimized soil-adjusted vegetation index	$GOSAVI = 1.16 \times \frac{NIR - G}{NIR + G + 0.16}$	Gilabert et al., 2002
	Red Edge Optimized soil-adjusted vegetation index	$REOSAVI = 1.16 \times \frac{NIR - R}{NIR + R + 0.16}$	Kim et al., 1994
	Red Edge Renormalized difference vegetation Index	$RERDVI = \frac{NIR - RE}{NIR + RE}$	Kim et al., 1994
Chlorophyll Content Indices	Red edge Chlorophyll Index	$CLre = \frac{NIR}{RE - 1}$	Gitelson et al., 2006
	Chlorophyll Absorption Ratio Index	$CARI = (RE - R) - 0.2 \times (RE + R)$	Rondeaux et al., 1996
	Normalized Green Red Difference Index	$NGRDI = \frac{G - R}{G + R}$	Potgieter et al., 2017
	Triangular Vegetation Index	$TVI = 60 \times (NIR - G) - 100 \times (R - G)$	Broge and Leblanc, 2001
	Modified triangular vegetation index 2	$MTVI2 = \frac{1.5 \times \left(\begin{array}{l} 1.2 \times (NIR - G) \\ -2.5 \times (R - G) \end{array} \right)}{\sqrt{\begin{array}{l} (2 \times NIR + 1)^2 - \\ (6 \times NIR - 5 \times \sqrt{R}) - 0.5 \end{array}}}$	Haboudane, 2004
	MERIS Terrestrial Chlorophyll Index	$MTCI = \frac{NIR - RE}{RE - R}$	Dash and Curran, 2004

sensing images. It helps to reveal the details of vegetation structure, including the arrangement and density of leaves, which directly affect the estimation of LAI. At present, the most widely used is the gray level co-occurrence matrix (GLCM) of statistical analysis method, which has rotation invariance, multi-scale characteristics and low computational complexity. In this study, GLCM was utilized to extract texture features from the green, red, red edge, and near-infrared bands of the multispectral imagery in ENVI 5.3 software. Specifically, eight texture features were extracted from each band, including mean (MEA), variance (VAR), homogeneity (HOM), contrast (CON), dissimilarity (DIS), entropy (ENT), second moment (SEC), and correlation (COR), resulting in a total of 32 texture features. MEA provides insights into the overall trends of vegetation distribution and brightness, while VAR reveals the degree of dispersion between pixel grayscale levels in GLCM. HOM measures the uniformity of vegetation texture, CON captures brightness variation, DIS assesses dissimilarity in texture, ENT indicates texture complexity, SEC reflects overall texture trends, and COR measures the linear relationship between vegetation structures. The selection of these texture features is based on their sensitivity to vegetation structure and their proven performance in previous studies on texture analysis. In the course of extraction, the inter-pixel offset was established at a distance of one pixel, with preference given to the utilization of 3×3 window. Grayscale gradation was set to 64, while the angle parameter retained its default value.

Similar to the calculation of vegetation indices, based on the 32 texture features obtained above, the texture indices can be calculated by combining two different features (Zheng et al., 2018). The three kinds of texture indices used in this paper are

normalized difference texture indices (NDTI), difference texture indices (DTI) and ratio texture indices (RTI). Matlab r2021a is a kind of software widely used in data processing. It was used to calculate them and 1984 TI were produced.

$$NDTI = \frac{T_1 - T_2}{T_1 + T_2} \quad (1)$$

$$DTI = T_1 - T_2 \quad (2)$$

$$RTI = \frac{T_1}{T_2} \quad (3)$$

where T_1 and T_2 represent the random texture features of any band in the green, red, red edge and near-infrared bands respectively.

2.3.3 Plant height

The PH of winter wheat was extracted by generating DSM using UAV multispectral images during the bare soil stage and jointing stage in the study area. The DSM generated during the bare soil stage was labeled as DSM_0, while the DSM generated during the jointing stage was labeled as DSM_1. The PH was obtained by calculating the difference between DSM_1 and DSM_0, with DSM_0 serving as the reference baseline (Figure 2).

$$PH = DSM_1 - DSM_0 \quad (4)$$

2.4 LAI estimation with UAV images and accuracy verification

The technical route of winter wheat LAI estimation at jointing stage using UAV multispectral images is shown in Figure 3.

2.4.1 LAI Estimation algorithms

The correlation between the above indices and LAI was analyzed. Vegetation indices and texture indices with strong correlations (VI, TI), along with PH, were chosen as independent variables. The measured LAI on the ground was considered as the dependent variable. Different machine learning algorithms (RF, XGBoost, SVM, BPNN) and two deep learning algorithms (CNN and LSTM) were employed to explore the potential of combining spectral and texture features with PH for the winter wheat LAI estimation. The selection of these algorithms is based on their widespread application and success in handling complex nonlinear relationships, feature extraction, and generalization performance. SVM and RF were chosen for their excellence in capturing nonlinear patterns, while BPNN was favored for its ability to model intricate relationships. XGBoost was considered a robust regression algorithm due to its outstanding performance in handling high-dimensional data, mitigating overfitting, and improving overall model performance. As for deep learning algorithms, CNN and LSTM were selected to extract spatial and temporal features from multispectral images captured by UAV, providing a more comprehensive understanding of the dynamic processes of vegetation growth. The above models were constructed by Matlab r2021a.

RF algorithm, employing decision trees as base learners (Beriman, 2001), constructs multiple trees in parallel by randomly selecting attributes. The prediction results of all decision trees are averaged to obtain the final regression modeling result of the entire random forest. Due to its random sampling and feature generation methods in decision trees, RF can improve the prediction accuracy of the model without significantly increasing computational complexity. The key parameters in RF include the number of trees and the number of nodes. After repeated debugging and optimization, the number of trees was determined to be 500, and

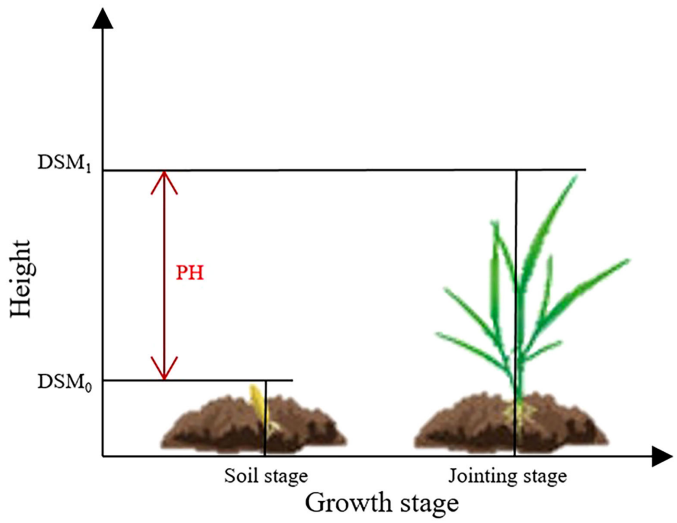


FIGURE 2
Principle of height extraction based on DSM.

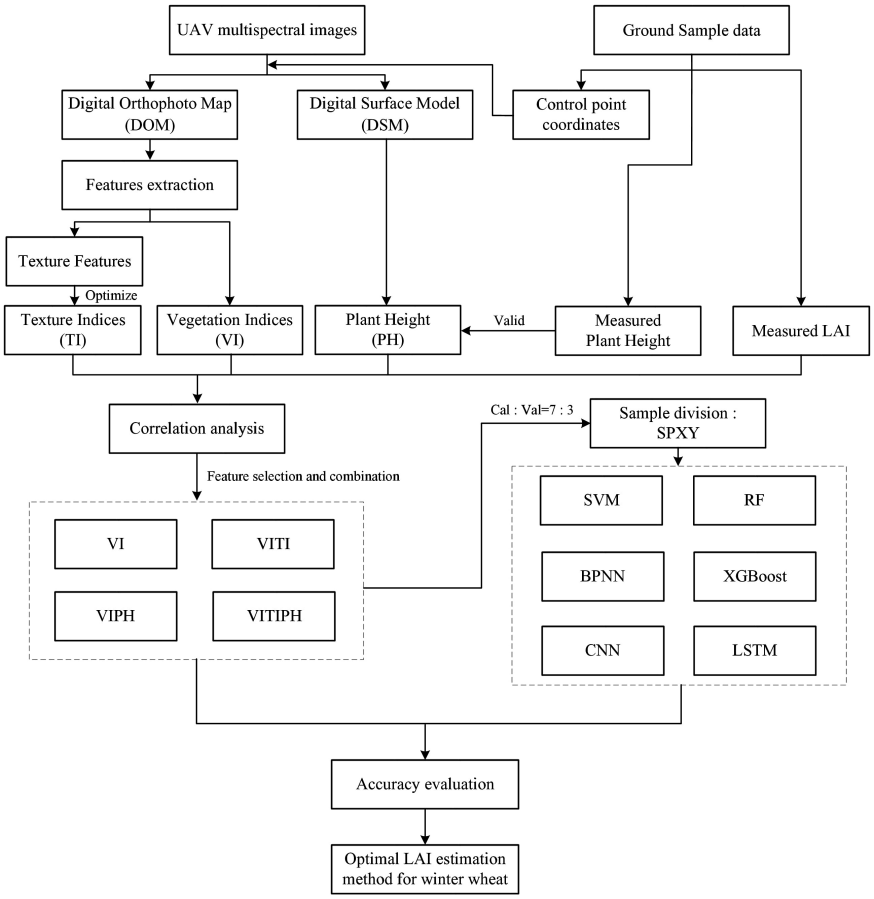


FIGURE 3
The technical route of winter wheat LAI estimation at jointing stage using UAV multispectral images.

the minimum number of samples for leaf nodes was set to 8 in this study.

XGBoost, an enhanced gradient boosting algorithm, combines multiple weak classifiers into a robust classifier (Chen and Guestrin, 2016). By separating the selection of the loss function from the optimization of the modeling algorithm and the selection of parameters, the algorithm can adaptively choose the appropriate loss function based on specific requirements or objectives, thereby enhancing the algorithm's applicability (Dhaliwal et al., 2018). XGBoost intergrates weak classifiers and enabling flexible loss function selection to enhance the modeling capability and overall performance. The primary settings for the kernel parameters were as follows: learning rate was set to 0.5, the maximum depth of tree was set to 1, the gamma was set to 0.01, the regularization parameters alpha and lambda were set to 0.02 and 0.1, respectively. Additionally, the subsampling method, subsample was set to 0.3, and colsample_bytree was set to 0.5.

SVM was chosen for small sample learning (Zhang et al., 2021a), utilizing the Radial Basis Function (RBF) kernel function. Its fundamental idea is to find an optimal hyperplane that minimizes the error between training sample points and the hyperplane. The optimal kernel parameters (g) and regularization parameter (c) were determined through adjustment and optimization. BPNN possesses strong fault tolerance and adaptive learning capabilities. It consists of input layers, hidden layers, and an output layer (Panda et al., 2010). By continuously adjusting the number of neurons in the hidden layers, the data is iteratively trained to obtain an optimal model. In this study, the number of hidden layers and the number of nodes were determined to be 1, respectively.

CNNs are capable of unsupervised feature learning, demonstrating remarkable performance in automated feature acquisition.The model architecture comprises convolutional layers, pooling layers, batch normalization layers, fully connected layers, dropout layers, and a regression layer (Lee et al., 2015). In this study, the size of the convolutional kernel was set to half the number of input variables. The Rectified Linear Unit (ReLU) activation function was used to accelerate the convergence speed of the model. During the training process, the dropout layer was employed with a dropout rate of 20% to enhance the generalization ability of the model and prevent overfitting. The Stochastic Gradient Descent with Momentum (SGDM) algorithm was utilized to optimize the weights of the model, and the initial learning rate was set to 0.01.

LSTM algorithm was built upon the foundation of recurrent neural networks, which introduces a gating mechanism to control the path of information transmission (Hochreiter and Schmidhuber, 1997). It utilizes input gates, forget gates, and output gates to dynamically adjust the weights of self-recurrent. In this way, when

the model parameters are fixed, the integration scale at different times can change dynamically, so as to effectively address the challenges of gradient explosion or disappearance of simple recursive neural network. The main parameter settings were consistent with CNN.

2.4.2 Accuracy evaluation

The samples were divided into calibration and validation sets using the sample set partitioning based on joint x-y distance (SPXY) (Galvao et al., 2005), which is based on the Kennard-Stone algorithm. This method considers both the feature variable (x) and the target variable (y) when selecting the data, aiming to determine a sub-feature space that maximally represents the original data space. It achieves this by calculating the relative Euclidean distance within the data space. By applying SPXY, the selected samples are more uniformly distributed and reasonably divided, providing a more comprehensive representation of vegetation conditions within the study area. The use of this approach may improve the repeatability and generalizability of the study.

$$d_{xy}(p,q) = \frac{d_x(p,q)}{\max\limits_{p,q\in[1,N]} d_x(p,q)} + \frac{d_y(p,q)}{\max\limits_{p,q\in[1,N]} d_y(p,q)}, (\mathrm{p,q} \in [1,N]) \quad (5)$$

$$d_x(p,q) = \sqrt{\sum_{j=1}^{J}\left[x_p(j) - x_q(j)\right]^2}, (\mathrm{p,q} \in [1,N]) \quad (6)$$

$$d_y(p,q) = \sqrt{(\mathrm{y}_p - \mathrm{y}_q)^2}, (\mathrm{p,q} \in [1,N]) \quad (7)$$

where $d_{xy}(p,q)$ represents the Euclidean distance of two spaces considered, $d_x(p,q)$ represents the Euclidean distance of two samples in x space (feature space), $d_y(p,q)$ represents the Euclidean distance of two samples in y space (target space), max $d_x(p,q)$ and max $d_y(p,q)$ denotes the maximum Euclidean distance of p and q in x and y space respectively. N is the total number of samples; J is the number of feature spaces. $x_p(j)$ and $x_q(j)$ respectively represent the values of p and q samples on the j-th feature. where y_p and y_q are the Euclidean distances of the two samples in y space.

The calibration dataset for estimating winter wheat LAI values was constructed using 70% of the sample data, while the remaining 30% served as the validation dataset for evaluation. The accuracy of the model was assessed using various performance metrics, including the Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and Ratio of Performance to Standard Deviate (RPD). For the evaluation indexes mentioned above, a higher R^2 value indicates a better prediction effect, while a smaller RMSE indicates a more accurate model. In terms of RPD, it is generally considered that RPD<1.4 indicates an unreliable model, 1.4< RPD< 2.0 suggests a moderate reliability, and model has a higher level of reliability if RPD > 2.0. These metrics provide quantitative measures to assess the accuracy and reliability of the model in estimating the winter wheat LAI.

$$R^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2(y_i - \bar{y})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}} \quad (9)$$

$$RPD = \frac{SD(x_i)}{RMSE} \quad (10)$$

where x_i and $\bar{x}$ represent the measured value and its mean value, y_i and $\bar{y}$ are the predicted value and its mean value of each sample, respectively. n indicates the number of samples in the calibration set or validation set. SD is the standard deviation of the measured value of the sample.

3 Results

3.1 Correlation between vegetation indices and winter wheat LAI

The correlation between the selected 17 vegetation indices and the measured LAI in the field was analyzed. The Pearson correlation coefficient is shown in Figure 4. The results revealed that all vegetation indices showed a correlation coefficient with LAI above 0.74, indicating a highly significant relationship (p<0.01). Among the vegetation indices, Clre exhibited the strongest correlation with LAI, with a correlation coefficient of 0.83, while CARI exhibited the weakest correlation, with a correlation coefficient of 0.74. Based on these findings, vegetation indices with correlation coefficients greater than 0.80 were selected as independent variables. The selected VI, ranked in descending order of correlation coefficient, were Clre, RERDVI, MTCI, Clgreen, GOSAVI, MSAVI, EVI2 and DVI, with corresponding correlation coefficients of 0.83, 0.83, 0.81, 0.81, 0.81, 0.80, 0.80, and 0.80, respectively.

3.2 Correlation between texture features and winter wheat LAI

Based on the correlation analysis between the selected texture features and LAI, it is evident from Figure 5 that more than half of the texture features exhibited a relatively low correlation with LAI. Only a small subset of texture features demonstrated a high correlation with LAI. Specifically, the MEA in the green band, near-infrared band, and red band exhibited correlation coefficients of 0.72, 0.73, and 0.72, respectively, indicating a strong correlation (P<0.01). However, for other texture features showing a highly significant correlation, the absolute values of the correlation coefficients generally ranged from 0.18 to 0.50. Given the

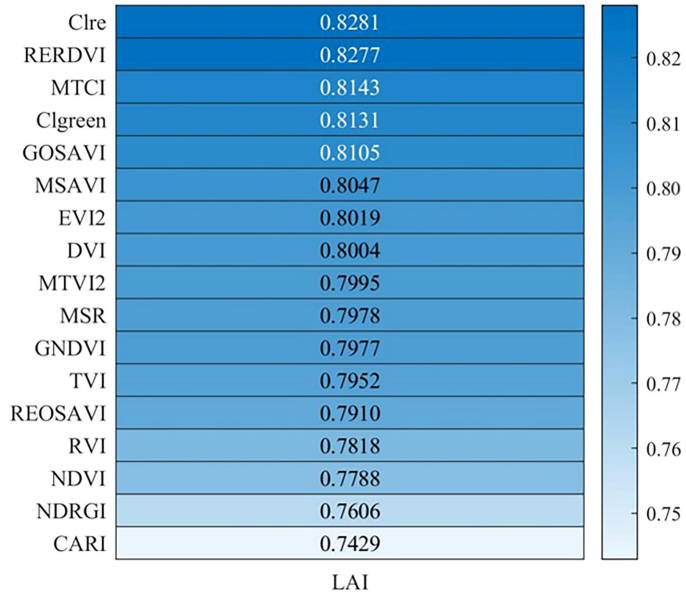


FIGURE 4
Correlation coefficients between winter wheat LAI and vegetation indices (P<0.01).

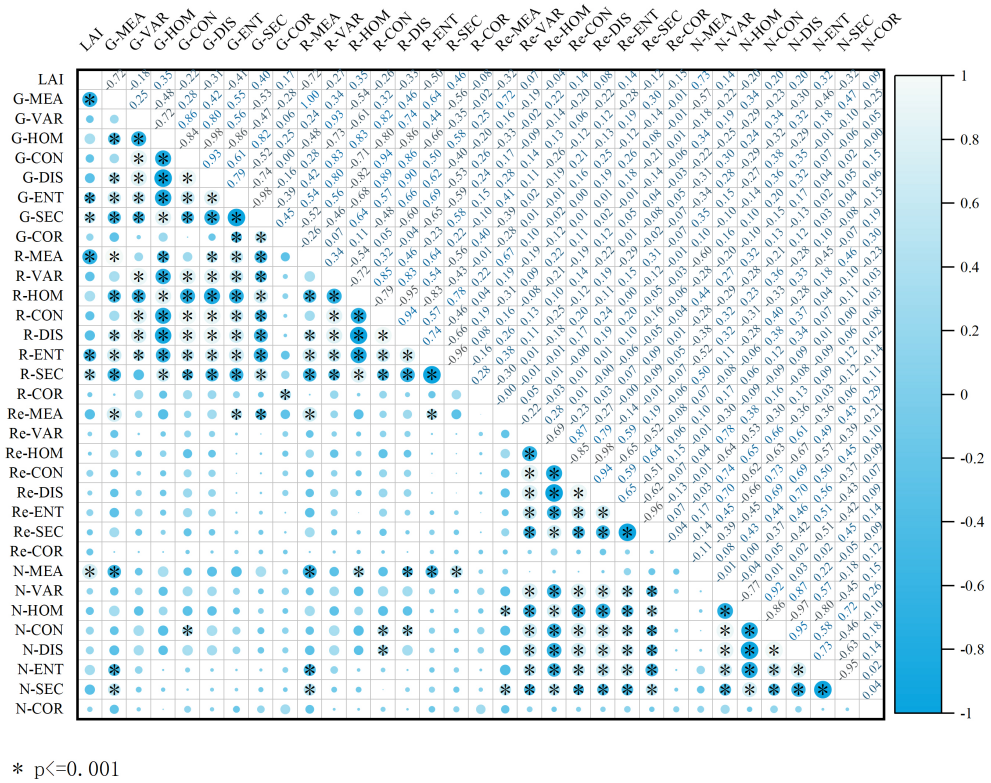


FIGURE 5
Correlation coefficients between winter wheat LAI and texture features (* represents P ≤ 0.001).

relatively weak correlation between texture features and LAI, their utility for accurate LAI prediction was limited. Consequently, this study employed some texture indices composed of texture features from different bands, specifically NDTI, DTI and RTI.

The correlation analysis between the texture indices and LAI showed that by combining the texture features, the overall correlation between the texture features and LAI was significantly improved. Figure 6 shows a high correlation between the combination of MEA for each band and LAI. Similarly, texture indices with a correlation coefficient greater than 0.80 were selected as independent variables. The correlations, listed from highest to lowest, were as follows: the ratio and normalized difference between the MEA of the near-infrared and red edge bands ($RTI_{MEA(N)-MEA(RE)}$, $NDTI_{MEA(N)-MEA(RE)}$), the ratio and difference between the MEA of the red edge and near-infrared bands ($RTI_{MEA(RE)-MEA(N)}$, $DTI_{MEA(N)-MEA(RE)}$), the ratio, difference and normalized value between the MEA of the near-infrared and green bands ($RTI_{MEA(N)-MEA(G)}$, $DTI_{MEA(G)-MEA(N)}$, $NDTI_{MEA(G)-MEA(N)}$), the difference between the MEA of the near-infrared and red bands ($DTI_{MEA(N)-MEA(R)}$), and the ratio between the MEA of the red edge and green bands ($RTI_{MEA(RE)-MEA(G)}$). The correlation coefficients for these TI were 0.84, 0.83, -0.83, 0.83, -0.82, 0.82, -0.81, 0.80, and 0.80, respectively. Figure 7 shows the correlation coefficients between winter wheat LAI and input features, including VI, TI and PH.

3.3 Winter wheat LAI estimation

Using four combinations of VI, VITI, VIPH, and VITIPH as input data, six regression algorithms including RF, XGBoost, SVM,

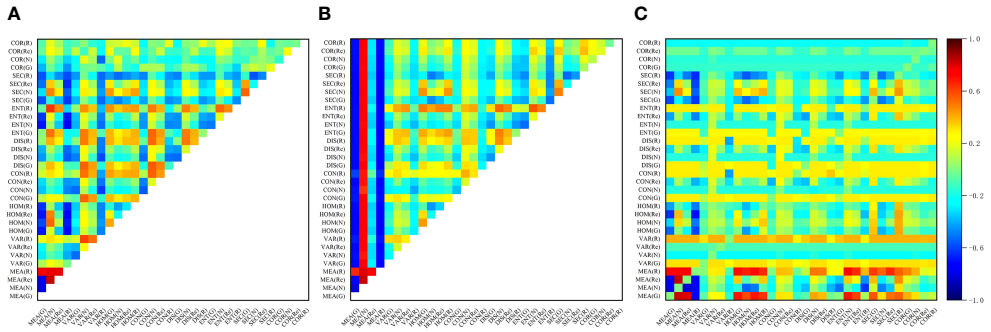


FIGURE 6
Correlation coefficients between winter wheat LAI and texture indices: **(A)** NDTI; **(B)** DTI; and **(C)** RTI. The abscissa and ordinate represent the correlation coefficient between the texture index and LAI of the corresponding two texture features after normalization, difference and ratio operations.

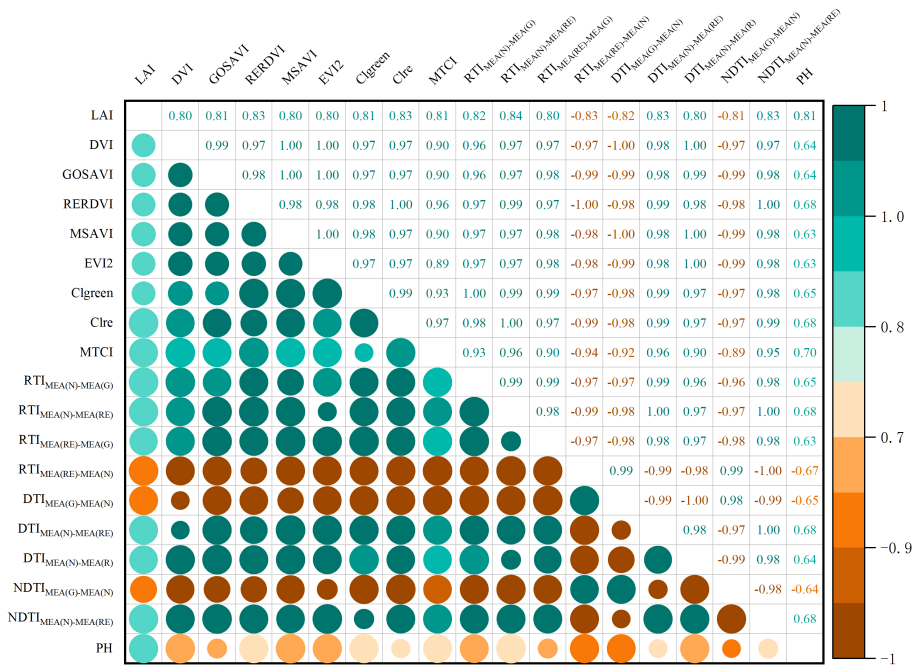


FIGURE 7
Correlation coefficients between winter wheat LAI and input features (VI, TI, PH).

BPNN, CNN, and LSTM were employed to estimate LAI for winter wheat at jointing stage, and their accuracy was evaluated. The results are shown in Figure 8.

3.3.1 LAI estimation based on different features

In the RF model, the accuracy of the LAI estimation model constructed using VITI and VIPH input variables was improved compared to the model constructed using VI (R^2 = 0.74, RMSE=0.99, RPD=1.49). The R^2, RMSE and RPD of the input VITI modeling set were 0.78, 0.93 and 1.60 respectively, and the R^2, RMSE and RPD of the validation set were 0.68, 1.15 and 1.40 respectively. For the VIPH modeling set, R^2, RMSE, and RPD were 0.83, 0.82, and 1.81, respectively, while the validation set yielded R^2 of 0.67, RMSE of 1.01, and RPD of 1.43. And the model constructed using VITIPH demonstrated the highest accuracy (R^2 = 0.85, RMSE=0.78, RPD=1.89). The XGBoost and RF algorithms exhibited similar model performance. Further analysis was conducted using the XGBoost method, which showed high overall accuracy. The calibration set constructed by inputting VITIPH showed the best performance among all regression models, achieving the highest R^2 (R^2 = 0.88), the lowest RMSE (RMSE=0.69), and an RPD greater than 2 (RPD=2.54).

Based on the SVM model, the models constructed using VI (R^2 = 0.69, RMSE=1.09, RPD=1.46) and VITI (R^2 = 0.69, RMSE=1.10, RPD=1.48) as input feature variables showed similar results in terms of model accuracy evaluation metrics on the calibration and validation sets. The model constructed by introducing PH as the input variable (VIPH, VITIPH) showed better effect than VI and VITI. The R^2, RMSE and RPD of calibration set constructed by VIPH were 0.82, 0.85 and 1.94 respectively, and the R^2, RMSE and RPD of validation set were 0.69, 0.93 and 1.40 respectively. The effect of the model constructed by VITIPH was the best. The R^2 and RPD of the modeling set increased to 0.83 and 2.13 respectively, and the RMSE decreased to 0.81.

The models constructed using the BPNN, CNN, and LSTM neural network algorithms also demonstrated positive predictive performance in estimating winter wheat LAI (all RPD > 1.4). The R^2 of the calibration set were all above 0.64, and the R^2 of the validation set were all above 0.51. The RMSE ranged from 0.77 to 1.16. All the three models showed that the model based on the combination of characteristic input variables of VITI and VIPH

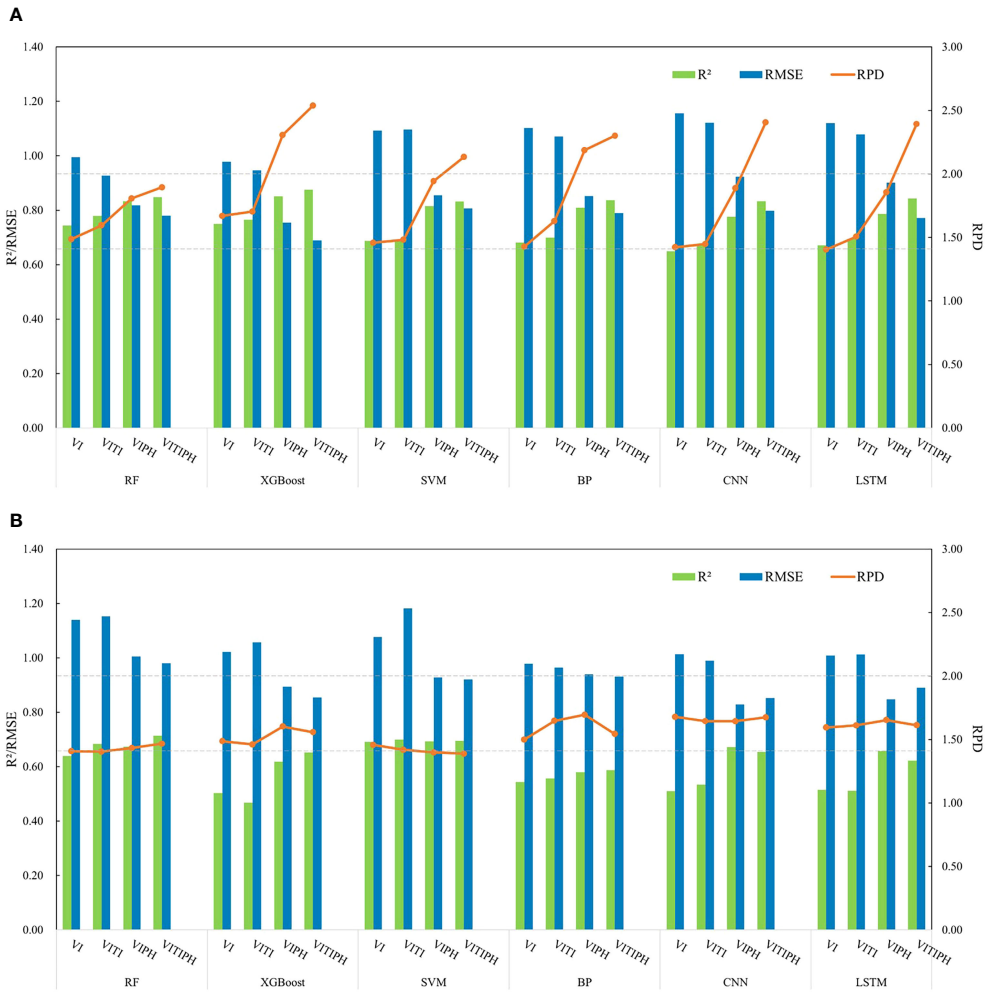


FIGURE 8
Accuracy of winter wheat LAI estimation based on different algorithms and input variables. **(A)** calibration dataset; **(B)** validation dataset.

performed better than the regression model with VI input, especially the inclusion of the PH further enhanced the performance of the model. In the calibration set, the BPNN model showed an increase in R^2 from 0.6818 to 0.8096, a decrease in RMSE from 1.10 to 0.85, and an increase in RPD from 1.43 to 2.19. Similarly, the CNN model exhibited an increase in R^2 from 0.65 to 0.78 and a decrease in RMSE from 1.16 to 0.92 in the calibration set. The modeling dataset of the LSTM model also had an increase in R^2 from 0.67 to 0.79, and in the validation dataset, with an increase in R^2 from 0.52 to 0.66. Meanwhile, the best model performance was achieved when using the VITIPH as input variables. In the BPNN model, the R^2, RMSE and RPD for the modeling dataset were 0.84, 0.79, and 2.30, respectively. In the CNN model, the R^2, RMSE, and RPD for the modeling dataset were 0.83, 0.80 and 2.41, respectively. In the LSTM model, the R^2, RMSE, and RPD for the modeling dataset were 0.84, 0.77 and 2.39, respectively, while for the validation dataset, the R^2, RMSE and RPD were 0.62, 0.89 and 1.61, respectively.

Overall, the models constructed using different combinations of features demonstrated better performance compared to those relying solely on VI input, with the following order of accuracy: VI<VITI<VIPH<VITIPH. This indicated that introducing other feature variables can effectively improve the accuracy of the winter wheat LAI estimation model. Notably, the incorporation of PH as a variable significantly enhanced the model's ability to estimate LAI. As an essential plant parameter, PH exhibited a strong correlation with LAI, making its inclusion in model construction crucial for achieving more accurate LAI estimation.

3.3.2 Combining VI, TI, and PH to estimate LAI using different algorithms

The scatter distribution of the measured and estimated LAI values obtained from the VITIPH input, using six algorithms, is presented in Figure 9. It was evident that the fitted distribution of the winter wheat LAI estimation, constructed based on the XGBoost algorithm, closely approximated a 1:1 relationship, indicating its superior predictive capability. However, some issues remained, such as overestimation of low values and underestimation of high values. Notably, the SVM, RF, and XGBoost models exhibited robust stability when compared to the BPNN, CNN, and LSTM neural network models. This is primarily attributable to the superior adaptability of SVM, RF, and XGBoost to small-sample datasets. Even in situations with limited data volume, they can provide relatively accurate estimates. Furthermore, these algorithms exhibit robustness to outliers and noise within the data. This resilience stems from the fundamental principles of these algorithms; for instance, SVM demonstrates enhanced tolerance to outliers through the concept of support vectors, while RF and XGBoost effectively mitigate the sensitivity of individual decision trees to noise by employing ensemble learning techniques. Moreover, SVM, RF, and XGBoost models possess a relatively parsimonious set of hyperparameters compared to neural network models. These attributes contribute to the stability and reliability of the models during the training process, thereby manifesting superior performance in our research.

Moreover, it can be observed from Figures 8, 9 that a consistent trend occurred when comparing the LAI estimation using various regression algorithms. Combined vegetation indices and texture indices with plant height, the calibration set R^2 of the six algorithms was greater than 0.8, indicating that the accuracy of LAI estimation by machine learning and deep learning regression algorithm was high.

3.4 Winter wheat LAI Inversion map

The XGBoost model, which combined VI, TI, and PH, achieved the most accurate prediction of winter wheat LAI. This model was utilized to generate a spatial distribution map of winter wheat LAI inversion in the study area, LAI values range from 0.96 to 8.86, as depicted in Figure 10. By extrapolating LAI from the pixel scale to the regional scale, remote sensing monitoring of winter wheat LAI at a broader scale was achieved. This approach facilitated a more comprehensive understanding of the growth status and spatial distribution patterns of winter wheat in the region. The findings will provide valuable information for real-time monitoring of wheat growth and development, as well as the formulation of customized fertilization prescriptions and other agricultural production management and decision-making processes.

For example, the plant water use efficiency can be evaluated through the further analysis of LAI inversion map.This aids in the rational development of irrigation plans, ensuring that plants receive adequate water and enhancing the overall water resource utilization efficiency in agricultural fields. Simultaneously, the leaf area of a plant is associated with nutrient absorption. Agricultural practitioners can leverage LAI inversion maps to precisely understand the nutrient requirements of plants, thereby optimizing fertilizer plans and improving fertilizer efficiency. Furthermore, agricultural decision-makers can utilize LAI inversion maps for early detection of diseases and pests, enabling the implementation of preventive or curative measures to mitigate the adverse impact of plant diseases on yield. Additionally, LAI inversion maps can be employed for preliminary estimations of crop yields.

4 Discussion

The main objective of this study was to explore the potential of using ML/DL to improve the accuracy of winter wheat LAI estimation by combining VI, TI and PH based on UAV multispectral image.

The study revealed a robust correlation between vegetation indices and LAI for winter wheat during the jointing period. The selected 17 vegetation indices (RVI, DVI, NDVI, GNDVI, EVI2, Clgreen, MSR, MSAVI, GOSAVI, REOSAVI, RERDVI, Clre, CARI, NGRDI, TVI, MTVI2, MTCI) in this study demonstrated correlations exceeding 0.74 with LAI, with p-values lower than 0.01, signifying a highly significant relationship. Notably, indices

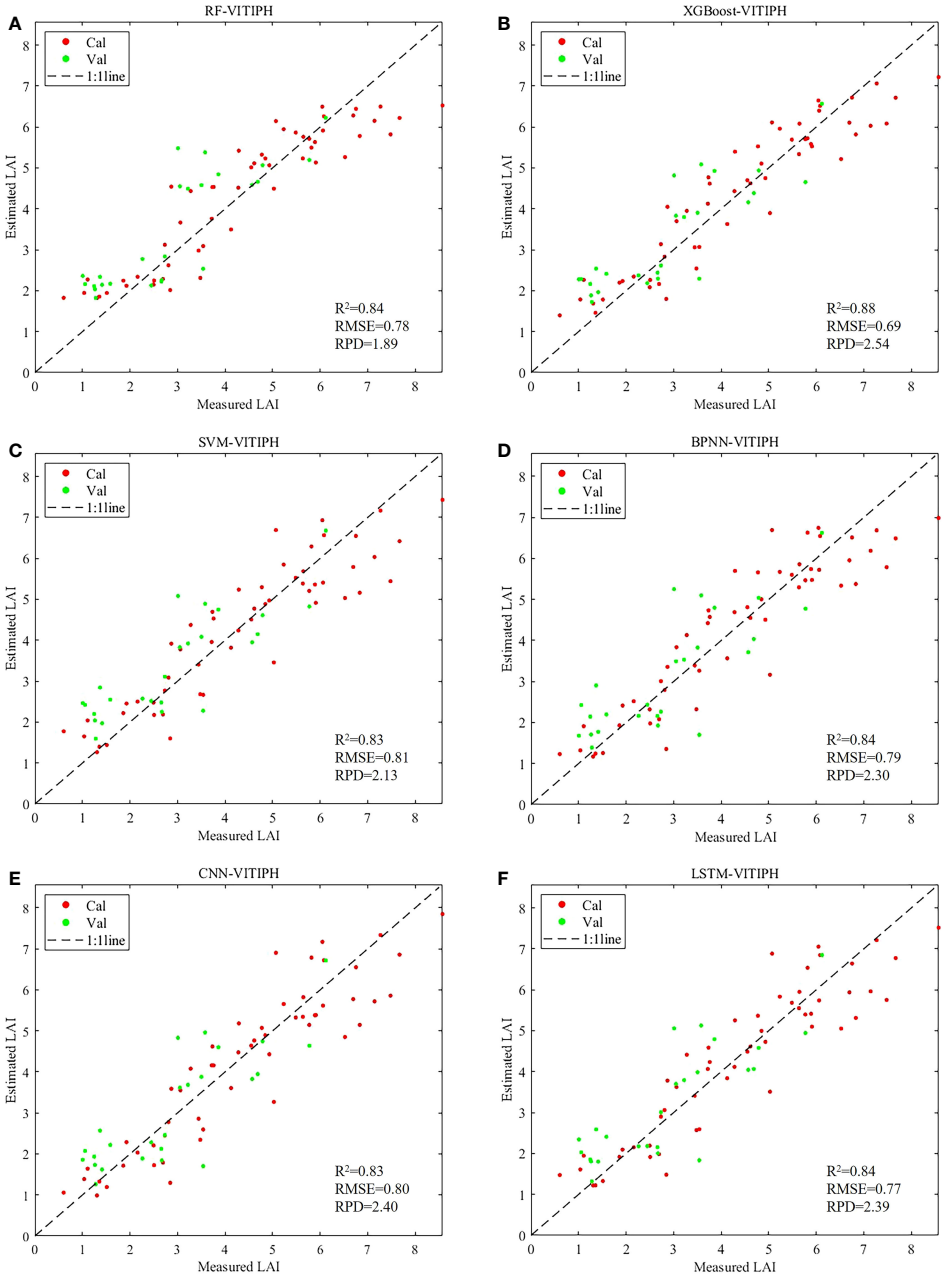


FIGURE 9
Scatter diagram of LAI estimation results of winter wheat with six algorithms using VITIPH as input variables. **(A)** RF-VITIPH; **(B)** XGBoost-VITIPH; **(C)** SVM-VITIPH; **(D)** BPNN-VITIPH; **(E)** CNN –VITIPH; **(F)** LSTM -VITIPH. The calibration set and validation set are represented in red and green point respectively.

like Clre, RERDVI, MTCI, and Clgreen, founded on the ratio of near-infrared to visible light bands, provide insights into chlorophyll content and photosynthetic activity. Given that LAI of winter wheat characterizes the total area of the leaves, and the chlorophyll content in the leaves is related to LAI in some cases. During the jointing stage, winter wheat exhibits higher chlorophyll content and more vigorous photosynthetic activity, fostering in a strong correlation between these indices and LAI. Moreover, certain indices depict alterations in vegetation structure and coverage, such as NDVI, MSAVI, and EVI2. As winter wheat grows, the plant's structure progressively develops, leading to an increase in leaf number and density, alongside expanding vegetation coverage, ultimately resulting in elevated LAI. These indices sensitively capture variations in vegetation structure and coverage, thereby demonstrating a significant positive correlation with LAI. Indices such as EVI2 and DVI have been corrected for soil background and atmospheric effects, effectively reducing interference with vegetation reflectance and providing more precise information regarding wheat LAI. Consequently, these indices exhibit a significant correlation with LAI. The integration of these diverse

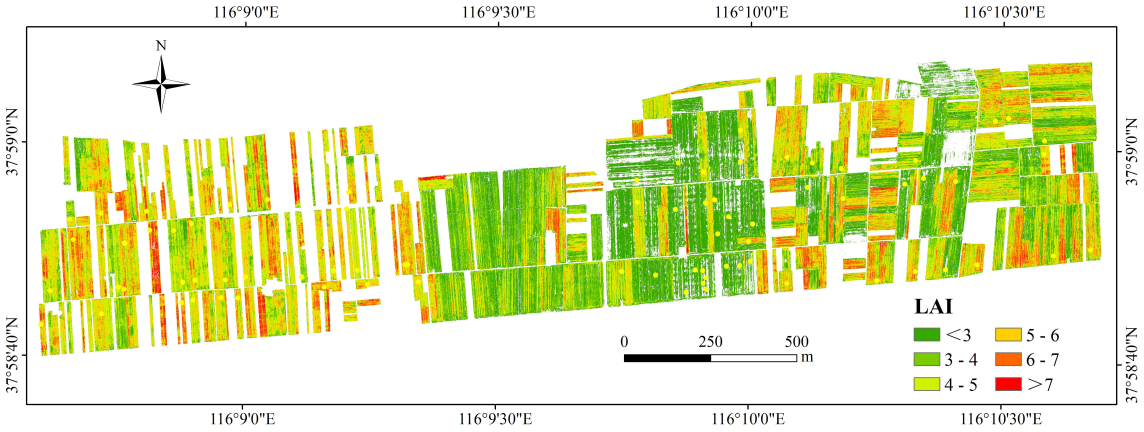


FIGURE 10
Spatial distribution of winter wheat LAI estimation results based on optimal method in the study area.

vegetation indices contributes to a comprehensive understanding of LAI dynamics for winter wheat during the jointing period, enhancing the accuracy of LAI estimation through remote sensing approaches.

Nevertheless, the majority of texture features exhibited a weak correlation with LAI. In order to address this limitation, combining multiple texture features to create a new texture index can integrate diverse and comprehensive texture information. This approach mitigates the influence of soil, terrain, and shadow backgrounds while accentuating pertinent features (Hang et al., 2021). It is worth noting that the extraction of texture features is often susceptible to image noise, variations in illumination, and other interfering factors, which may lead to the instability of the results of a single texture feature. By employing combination operations, random noise within individual features can be eliminated or reduced, while uncertainties stemming from changes in illumination and other factors can be minimized. This contributes to enhancing the stability and reliability of the correlation between texture features and LAI. The NDTI, DTI and RTI formed by the combination of MEA of each band have high correlation with LAI. The mean value of texture measurement includes the average value of the target and background in the moving window, which can smooth the image and minimize background interference (Wang et al., 2022b). Additionally, green vegetation absorbs most visible light in the red edge band, and in the near-infrared band, the diffuse reflection of the canopy structure leads to a higher reflectance in the near-infrared region (Yu et al., 2020). The difference, ratio, and normalized difference between the near-infrared and red edge bands can enhance the difference in light absorption and reflection of vegetation, thus better reflecting the canopy structure

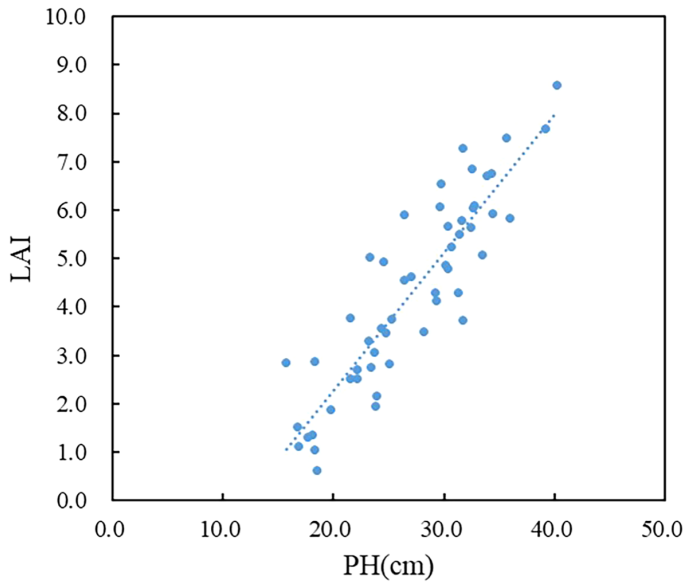


FIGURE 11
Scatter plot between LAI and PH.

of green vegetation. This further reinforces the capability of texture information in representing LAI, aligning with the findings of Zhang (Zhang et al., 2022a). Through the comparison of six regression algorithms, this study identified that integrating VI and TI enhanced the performance of the LAI estimation model.

However, solely combining TI does not effectively address the saturation issue of VI in high-density canopies. On the other hand, incorporating plant structure information, such as PH, proves more beneficial in addressing or improving this concern. Analysis of Figure 11 reveals a clear linear positive correlation between PH and LAI during the jointing stage of winter wheat, indicating that LAI progressively increases with the growth of wheat plant height. This is attributed to the fact that an increase in wheat plant height is often accompanied by a corresponding increase in leaf area. The expanded leaf area enables the plant to absorb and utilize more light energy, which facilitates enhanced photosynthesis, increases organic matter production, and promotes overall plant growth and development. It is also used to form plant organs such as roots, stems and ears, which reacts on the increase of plant height. Yuan et al. (Yuan et al., 2013) have demonstrated a significant positive correlation between vegetation canopy height and LAI. Accordingly, this study incorporated the structural information of PH extracted from UAV images in the construction of the winter wheat LAI estimation model.Furthermore, it is worth noting that this paper achieved a satisfactory level of accuracy in predicting plant height based on UAV-derived data, with an R^2 of 0.86 and an RMSE of 2.07cm between predicted and measured values.

Studies have indicated that when constructing a LAI estimation model with a single type of feature input as a variable, the model constructed by texture feature is not as stable as the VI and is not suitable for estimating crop LAI as an independent feature variable (Zhang et al., 2021b). Therefore, this paper only considered the VI as a single type of feature variable to construct the model. Simultaneously, three types of multi-feature fusion combinations were considered (VITI, VIPH, VITIPH). The results showed that, compared to a single type of variable, the winter wheat LAI estimation model constructed by combining VI and TI with PH exhibited the highest accuracy. Furthermore, the combination of VI and PH in the model construction exhibited significantly higher accuracy compared to models based solely on VI or the combination of VI and TI in estimating winter wheat LAI. This highlighted the feasibility of integrating plant height and spectral information in LAI estimation, leading to an improvement in model accuracy and alleviating the issue of spectral saturation. For future studies, it is recommended to consider the inclusion of other crop vertical structural features during model construction, which may further enhance LAI estimation accuracy and provide comprehensive insights into the growth dynamics of winter wheat.

Analyzing the estimated results, it can be seen that all six algorithms used in this paper achieved strong ability to estimate winter wheat LAI. The RF model constructed using VITIPH obtained the R^2 was 0.85, RMSE was 0.78, and RPD was 1.90. The R^2, RMSE and RPD of calibration set constructed by SVM using VITIPH were 0.83, 0.81 and 2.13 respectively. The models constructed using the BPNN, CNN, and LSTM neural network algorithms also demonstrated positive predictive performance in estimating winter wheat LAI (all RPD > 1.4). It is notable that the model based on the XGBoost algorithm demonstrated the highest performance among all the models tested, achieving the highest R^2 (R^2 = 0.88), the lowest RMSE (RMSE=0.69), and an RPD greater than 2 (RPD=2.54). By comprehensively comparing the R^2, RMSE and RPD of the correction set and the verification set of the six algorithms, it was found that the six evaluation indexes of the model performed well as a whole, so XGBoost was selected as the most optimal model. The advantage of this algorithm is that it is an ensemble learning algorithm, which constructs a strong classifier by combining multiple weak classifiers (decision trees) (Chen and Guestrin, 2016). And ensemble learning enables the combination of predictions from multiple models, reducing individual model biases and variances, and improving the overall generalization ability of the model. This feature allows XGBoost to effectively capture the intricate relationship between LAI and the input features, thereby enhancing the performance of the model. Additionally, XGBoost employs an optimized second-order Taylor expansion, enabling it to accurately estimate complex nonlinear datasets and achieve superior results. Additionally, the SPXY algorithm was used to screen the calibration samples and validation samples before constructing the model in this paper, which considered the distance between different feature variables and the distance between target variables at the same time. By doing so, the algorithm ensures a more even distribution and reduces the differences between the calibration and validation sets, thus improving the accuracy of the model. Researches have proved that machine learning and deep learning algorithms combined with remote sensing features had a good effect on LAI estimation of winter wheat (Li et al., 2021; Zhang et al., 2021b). In this study, the performance of the CNN and LSTM algorithms did not surpass expectations, potentially attributed to the limited sample size. It might have hindered the ability of deep learning to fully leverage its advantages, resulting in a decline in model performance. Increasing the sample size has various positive impacts on the performance of deep learning models. Firstly, a larger sample size enhances the model's generalization ability, allowing it to better adapt to a wide range of data distributions, thus exhibiting more robust performance when faced with new, unseen data. Additionally, the increase in sample size helps mitigate the risk of overfitting, making the model's performance on the test set more reliable.Ultimately, a greater sample size provides more accurate model evaluations, offering a more reliable basis for the objective assessment of model performance. In general, the deep learning has advantages in processing complex, high-dimensional and large-scale data, and can automatically learn and extract features (Nevavuori et al., 2019). However, due to its large demand for data and computing resources, high requirements for interpretability and comprehensibility, complex hyper-parameters selection and optimization, and sensitivity to data quality and noise, it may be limited in these aspects when processing some tasks, which is relatively inferior to machine learning algorithms. Therefore,

selecting an appropriate algorithm needs to comprehensively consider the task requirements, data characteristics and resource constraints.

This paper specifically focused on the estimation of winter wheat LAI during the jointing stage. Whether the accuracy of multiple features combination modeling is higher than that of single type of VI modeling in the whole growth cycle remains to be further studied. Additionally, it is important to examine whether the XGBoost algorithm maintains the highest LAI estimation accuracy across the entire growth period of winter wheat. After that, the data collection of the whole growth cycle of crops should be carried out to more comprehensively grasp the growth state of winter wheat. At the same time, the universality and applicability of the findings should be further studied in different regions.

The flight altitude of UAV is often closely linked to ground resolution. The extraction of texture features at varying ground resolutions has a discernible impact on the monitoring of Wheat LAI. Estimating LAI typically necessitates the consideration of vegetation texture information, encompassing leaf arrangement, size, and morphology. When the flight altitude is set at 9 cm, UAV sensors can capture finer details of ground vegetation, thus providing a richer source of texture information. This, in turn, contributes to a more precise characterization of vegetation structure and distribution. Given that this study was conducted with specific flight altitude and spatial resolution settings tailored for agricultural fertilization decisions, future research may require experiments involving different flight altitudes to validate the methodologies and outcomes delineated in this paper.

The method utilized in this study may also be extended to satellite data, such as WorldView-3 satellite, ZY-3 satellite, and others. High spatial resolution satellite remote sensing offers rich spectral and texture information, while plant height can be extracted through multi-temporal stereo image pairs. However, the critical growth and development period of crops is often relatively short, emphasizing the necessity of obtaining relevant satellite data within a specific time window. Challenges may arise when the satellite's coverage and revisit cycle fail to meet the data requirements during this crucial period. Additionally, the quality of satellite imagery becomes difficult to control due to adverse weather conditions, such as clouds and rain, which can lead to the loss of significant monitoring opportunities. In future research endeavors, satellite data retains its potential for monitoring crop growth parameters in lager areas, possibly surmounting the previously mentioned challenges through techniques like image fusion.

According to this research, it was known that UAV multispectral data played a pivotal role in estimating LAI. However, it is crucial to recognize and address certain limitations associated with the utilization of multispectral data. This constraint arises from the inherently restricted spectral resolution of multispectral data, wherein the UAV-mounted multispectral sensor captures information within a limited set of pre-defined bands. This limitation has the potential to hinder the precise characterization of subtle variations in vegetation properties. In contrast, hyperspectral data, characterized by capturing an extensive range of contiguous narrow bands across the electromagnetic spectrum (Tao et al., 2020), offer increased spectral resolution. The augmentation in resolution contributes to a more accurate characterization of vegetation attributes, potentially enhancing the accuracy of LAI estimation. Additionally, LiDAR data, capable of capturing three-dimensional structural information of vegetation, serves as a complementary dataset to spectral data (Luo et al., 2018). The enhancement of LAI estimation is primarily achieved by considering the vertical structure of vegetation. Therefore, it is worthwhile to consider the utilization of hyperspectral or LiDAR data for vegetation LAI estimation in future research endeavours.

In this study, winter wheat LAI monitoring based on UAV multiple image provides the underlying data for wheat production. The future work will apply this approach to different wheat growth periods under various production management practices, and explore the applicability in different regions and scales.

5 Conclusion

This study utilized UAV multispectral images to extract spectral and texture information, while incorporating plant height information for LAI estimation. The comparative analysis was conducted to assess the efficacy of four machine learning and two deep learning regression algorithms in estimating the LAI of winter wheat. The main conclusions were as follows:

1. At the jointing stage of winter wheat, the PH derived from UAV images played an important role in LAI estimation, which can improve the estimation accuracy of winter LAI.
2. The combination of texture features can significantly improve the correlation between texture features and LAI, especially the combination of the MEA of each band had high correlation with wheat LAI. Comparing VI, VITI, VIPH and VITIPH as input variables, it was found that the ability of combining multiple features to estimate LAI of winter wheat was better than the estimation model constructed by only inputting VI, and the fusion of three kinds of features involved in the construction of the estimation model was the best.
3. The machine learning and deep learning algorithms were shown promising results in accurately estimating winter wheat LAI using UAV remote sensing data. The RF, XGBoost and SVM model constructed using VITIPH obtained the R^2 values of 0.85, 0.88 and 0.83, while the R^2 of BPNN, CNN, and LSTM were 0.81, 0.78 and 0.79, respectively. It is notable that the model based on the XGBoost algorithm demonstrated the highest performance among all the models tested.

The research results demonstrate that UAV data and advanced algorithms will provide technical support for the rapid and nondestructive estimation of winter wheat LAI and help to formulate variable rate fertilization prescription of agricultural machinery. This research provides a valuable framework for optimizing agricultural practices, underscoring their potential of leveraging advanced technologies for precision agricultural and making significant contributions to sustainable farming.

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

MZ: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. YL: Funding acquisition, Resources, Writing – review & editing. MF: Funding acquisition, Resources, Writing – review & editing. CL: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing. ZZ: Supervision, Writing – review & editing. HM: Supervision, Writing – review & editing. EX: Supervision, Writing – review & editing. YR: Funding acquisition, Resources, 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 Key R&D Plan Projects in Hebei Province (22327402D), the Youth Foundation of Beijing Academy of Agriculture and Forestry Sciences (QNJJ202232), National Natural Science Foundation of China (41671435), Science and Technology Planning Project of Guangdong Province of China (2019B090905006).

Acknowledgments

We would like to express our heartfelt gratitude to all the participants who generously contributed their time and effort to this study. Their valuable insights and cooperation have been indispensable in shaping the success of this research. Our sincere thanks go to the funding agencies for their generous financial support, which made this research possible. Additionally, we extend our appreciation to the teachers and reviewers for their valuable feedback and constructive suggestions, which significantly improved the quality of this work.

Conflict of interest

Author MF was employed by the company Hebei Maodong Xingteng Agricultural Technology Service Co., Ltd.

The remaining 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

Azadbakht, M., Ashourloo, D., Aghighi, H., Radiom, S., and Alimohammadi, A. (2019). Wheat leaf rust detection at canopy scale under different LAI levels using machine learning techniques. *Comput. Electron. Agric.* 156, 119–128. doi: 10.1016/j.compag.2018.11.016

Becker, F., and Choudhury, B. J. (1988). Relative sensitivity of normalized difference vegetation Index (NDVI) and microwave polarization difference Index (MPDI) for vegetation and desertification monitoring. *Remote Sens. Environ.* 24 (2), 297–311. doi: 10.1016/0034-4257(88)90031-4

Behrens, T., and Diepenbrock, W. (2006). Using Digital Image Analysis to Describe Canopies of Winter Oilseed Rape (Brassica napus L.) during Vegetative Developmental Stages. *J. Agron. Crop Sci.* 192 (4), 295–302. doi: 10.1111/j.1439-037X.2006.00211.x

Beriman, L. (2001). Random forests. *Mach. Learn.* 45 (1), 5–32. doi: 10.1023/A:1010933404324

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 (2), 156–172. doi: 10.1016/S0034-4257(00)00197-8

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 vegetationindex (WDRVI) derived from UAV multispectral images. *Comput. Electronatic. Agric.* 171, 105331. doi: 10.1016/j.compag.2020.105331

Cao, Z., Li, Y., Huang, J., Ye, C., Sun, B., Shu, S., et al. (2022). Monitoring rice leaf area index based on unmanned aerial vehicle (UAV) digital images. *Chin. J. Rice Sci.* 36 (3), 308–317. doi: 10.16819/j.1001-7216.2022.210712

Casa, R., Varella, H., Buis, S., Guérif, M., De Solan, B., and Baret, F. (2012). Forcing a wheat crop model with LAI data to access agronomic variables: Evaluation of the impact of model and LAI uncertainties and comparison with an empirical approach. *Eur. J. Agron.* 37 (1), 1–10. doi: 10.1016/j.eja.2011.09.004

Chen, J. M., and Cihlar, J. (1996). Retrieving leaf area index of boreal conifer forests using Landsat TM images. *Remote Sens. Environ.* 55 (2), 153–162. doi: 10.1016/0034-4257(95)00195-6

Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *ACM*.doi: 10.1145/2939672.2939785

Cui, Y., Wei, Z., Wang, J., and Xue, Q. (2021). Development status and path of application of chemical fertilizers and pesticides under the background of reduced. *Northern Horticulture* 000 (009), 164–173.

Dai, M., Yao, Z., Liu, T., and Sun, C. (2022). Wheat biomass estimation in different growth stages based on color and texture features of UAV images. *Smart Agric.* 4, 71–83. doi: 10.12133/j.smartag.SA202202004

Dash, J., and Curran, P. J. (2004). MTCI: The meris terrestrial chlorophyll index. *Int. J. Remote Sens.* 25 (549), 151–161. doi: 10.1109/IGARSS.2004.1369009

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., Colstoun, E. B. D., and Iii, M. M. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens. Environ.* 74 (2), 229–239. doi: 10.1016/S0034-4257(00)00113-9

Dente, L., Satalino, G., Mattia, F., and Rinaldi, M. (2008). Assimilation of leaf area index derived from ASAR and MERIS data into CERES-Wheat model to map wheat yield. *Remote Sens. Environ.* 112 (4), 1395–1407. doi: 10.1016/j.rse.2007.05.023

Dhaliwal, S. S., Nahid, A. A., and Abbas, R. (2018). Effective intrusion detection system using XGBoost. *Information* 9 (7), 149. doi: 10.3390/info9070149

Du, Y., Jiang, X., Zhang, Y., Huang, C., Liu, Z., and Liu, L. (2016). Retrieving leaf area index using PROSAIL radiative transfer model based on Landsat 8 image. *Arid Land Geogr.* 39, 1096–1103. doi: 10.13826/j.cnki.cn65-1103/x.2016.05.021

Fu, B., He, X., Yan, H., Liang, Y., Deng, T., He, H., et al. (2022b). Comparison of RFE-DL and stacking ensemble learning algorithms for classifying mangrove species on

UAV multispectral images. *Int. J. Appl. Earth Observation Geoinformation* 112, 102890. doi: 10.1016/j.jag.2022.102890

Fu, Z., Jiang, J., Gao, Y., Krienke, B., and Liu, X. (2020). Wheat growth monitoring and yield estimation based on multi-rotor unmanned aerial vehicle. *Remote Sens.* 12 (3), 508. doi: 10.3390/rs12030508

Fu, B., Sun, J., Wang, Y., Yang, W., He, H., Liu, L., et al. (2022a). 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

Galvao, R. K. H., Araujo, M. C. U., Jose, G. E., Pontes, M. J. C., Silva, E. C., and Saldanha, T. C. B. (2005). A method for calibration and validation subset partitioning. *Talanta: Int. J. Pure Appl. Analytical Chem.* 67 (4), 736–740. doi: 10.1016/j.talanta.2005.03.025

Gao, M., Zhang, J., Pan, Y., Duan, Y., and Zhang, D. (2020). Retrieval of winter wheat leaf area index based on vegetation index and crop height. *Chin. J. Agric. Resour. Regional Plann.* 41 (8), 49–57.

Gilabert, M. A., González-Piqueras, J., Garca-Haro, F. J., and Meliá, J. (2002). A generalized soil-adjusted vegetation index. *Remote Sens. Environ.* 82 (2), 303–310. doi: 10.1016/S0034-4257(02)00048-2

Gitelson, A. A., Kaufman, Y. J., and Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from eos-modis. *Remote Sens. Environ.* 58 (3), 289–298. doi: 10.1016/S0034-4257(96)00072-7

Gitelson, A. A., Keydan, G. P., and Merzlyak, M. N. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Res. Lett.* 33 (11), 431–433. doi: 10.1029/2006GL026457

Haboudane, D. (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 (3), 337–352. doi: 10.1016/j.rse.2003.12.013

Han, Y. J. (2011). Influence of global warming on agriculture and its cause analysis, countermeasures. *J. Anhui Agric. Sci.* 39, 9884–9885+10006. doi: 10.13989/j.cnki.0517–6611.2011.16.023

Handique, B. K., Khan, A. Q., Goswami, C., Prashnani, M., Gupta, C., and Raju, P. L. N. (2017). Crop discrimination using multispectral sensor onboard unmanned aerial vehicle. *Proc. Natl. Acad. Sciences India Section A: Phys. Sci.* 87, 713–719. doi: 10.1007/s40010-017-0443-9

Hang, Y., Su, H., Yu, Z., Liu, H., Guan, H., and Kong, F. (2021). Estimation of rice leaf area index combining UAV spectrum, texture features and vegetation coverage. *Trans. Chin. Soc. Agric. Eng.* 37 (9), 64–71. doi: 10.11975/j.issn.1002-6819.2021.09.008

Hasan, U., Sawut, M., and Chen, S. (2019). Estimating the leaf area index of winter wheat based on unmanned aerial vehicle RGB-image parameters. *Sustainability* 11 (23), 6829. doi: 10.3390/su11236829

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 Science.* 282, 95–103. doi: 10.1016/j.plantsci.2018.10.022

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Computing* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Hu, G., Li, S., and Yang, R. (2018). Comparison of three regression models for remote sensing estimation of leaf area index. *Sci. Surveying Mapp. 43(10)* 46-50, 66. doi: 10.16251/j.cnki.1009-2307.2018.10.007

Huete, A. R., Jackson, R. D., and Post, D. F. (1985). Spectral response of a plant canopy with different soil backgrounds. *Remote Sens. Environ.* 17 (1), 37–53. doi: 10.1016/0034-4257(85)901117

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 (10), 3833–3845. doi: 10.1016/j.rse.2008.06.006

Khaki, S., Wang, L., and Archontoulis, S. V. (2019). A CNN-RNN framework for crop yield prediction. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01750

Kim, M. S., Daughtry, C. S. T., Chappelle, E. W., Mcmurtrey, J. E., and Walthall, C. L. (1994). "The use of high spectral resolution bands for estimating absorbed photosynthetically active radiation (A par)," in *CNES, Proceedings of 6th International Symposium on Physical Measurements and Signatures in Remote Sensing*.

Koirala, A., Walsh, K. B., Wang, Z., and McCarthy, C. (2019). Deep learning – Method overview and review of use for fruit detection and yield estimation. *Comput. Electron. Agric.* 162, 1233. doi: 10.1016/j.compag.2019.04.017

Kupidura, P. (2019). The comparison of different methods of texture analysis for their efficacy for land use classification in satellite imagery. *Remote Sens.* 11 (10). doi: 10.3390/rs11101233

Lee, S. H., Chan, C. S., Wilkin, P., and Remagnino, P. (2015). "Deep-plant: Plant identification with convolutional neural networks," in *Proceedings - International Conference on Image Processing*. doi: 10.1109/ICIP.2015.7350839

Li, J., Jiang, H., Luo, W., Ma, X., and Zhang, Y. (2023). Potato LAI estimation by fusing UAV multi-spectral and texture features. *J. South China Agric. Univ.* 44 (1), 93–101. doi: 10.7671/j.issn.1001-411X.202201002

Li, Y., Liu, H., Ma, J., and Zhang, L. (2021). Estimation of leaf area index for winter wheat at early stages based on convolutional neural networks. *Comput. Electron. Agric.* 190. doi: 10.1016/j.compag.2021.106480

Li, S., Yuan, F., Ata-Ui-Karim, S. T., Zheng, H., and Cao, Q. (2019). Combining color indices and textures of UAV-based digital imagery for rice LAI estimation. *Remote Sens.* 11 (15), 21. doi: 10.3390/rs11151763

Liang, L., Di, L., Zhang, L., Deng, M., Qin, Z., Zhao, S., et al. (2015). Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method. *Remote Sens. Environ.* 165, 123–134. doi: 10.1016/j.rse.2015.04.032

Liu, Y., Feng, H., Yue, J., Jin, X., Li, Z., and Yang, G. (2022). 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. doi: 10.3389/fpls.2022.938216

Liu, J., Pattey, E., and Jégo, G. (2012). Assessment of vegetation indices for regional crop green LAI estimation from Landsat images over multiple growing seasons. *Remote Sens. Environ.* 123, 347–358. doi: 10.1016/j.rse.2012.04.002

Liu, C., Yang, G., Li, Z., Tang, F., and Zhang, L. (2018). Biomass estimation in winter wheat by UAV spectral information and texture information fusion. *Scientia Agricultura Sin.* 51, 3060–3073. doi: 10.3864/j.issn.0578-1752.2018.16.003

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

Luo, S., Chen, J., Wang, C., Gonsamo, A., Xi, X., Lin, Y., et al. (2018). Comparative performances of airborne LiDAR height and intensity data for leaf area index estimation. *IEEE J. selected topics Appl. Earth Observations Remote Sens.* 11 (1), 300–310. doi: 10.1109/JSTARS.2017.2765890

Nevavuori, P., Narra, N., and Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. *Comput. Electronic Agric.* 163, 104859. doi: 10.1016/j.compag.2019

Nie, S., Wang, C., Dong, P., and Xi, X. (2016). Estimating leaf area index of maize using airborne full-waveform lidar data. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 7 (1-3), 111–120. doi: 10.1080/2150704X.2015.1111536

Niu, Q., Feng, H., Yang, G., Li, C., Yang, H., and Xu, B. (2018). Monitoring plant height and leaf area index of maize breeding material based on UAV digital images. *Trans. Chin. Soc. Agric. Eng.* 34 (5), 73–82. doi: 10.11975/j.issn.1002-6819.2018.05.010

Panda, S. S., Ames, D. P., and Panigrahi, S. (2010). Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sens.* 2, 673–696. doi: 10.3390/rs2030673

Pearson, R. L., and Miller, L. D. (1972). Remote mapping of standing crop biomass for estimation of productivity of the shortgrass prairie. *Remote Sens. Environment VIII* 8. doi: 10.1177/002076409904500102

Pinter, P. J., Hatfield, J. L., Schepers, J. S., Barnes, E. M., Moran, M. S., and Daughtry, C. S. T. (2003). Remote sensing for crop management. *Photogrammetric Eng. Remote Sens.* 69 (6), 647–664. doi: 10.14358/PERS.69.6.647

Potgieter, A. B., Barbara, G. J., Chapman, S. C., Kenneth, L., and Suárez Cadavid Luz, A. (2017). Multi-spectral imaging from an unmanned aerial vehicle enables the assessment of seasonal leaf area dynamics of sorghum breeding lines. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01532

Rondeaux, G., Steven, M., and Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* 55 (2), 95–107. doi: 10.1016/0034-4257(95)00186-7

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA Special Publ.* 351, 309–317.

Sun, S., Zhao, Y., Wang, Y., Wang, X., and Zhang, S. (2019). Leaf area index inversion of winter wheat based on multispectral remote sensing of UAV. *J. China Agric. Univ.* 24 (11), 51–58. doi: 10.11841/j.issn.1007-4333.2019.11.06

Tao, H., Feng, H., Xu, L., Miao, M., and Long, H. (2020). Estimation of crop growth parameters using UAV-based hyperspectral remote sensing data. *Sensors* 20 (5), 1296. doi: 10.3390/s20051296

Torres-S´anchez, J., Pe˜na, J. M., de Castro, A. I., and L´opez-Granados, F. (2014). Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. *Compututer Electronaic Agric.* 103, 104–113. doi: 10.1016/j.compag.2014.02.009

Wang, Q., Putri, N. A., Gan, Y., and Song, M. (2022a). Combining both spectral and textural indices for alleviating saturation problem in forest LAI estimation using Sentinel-2 data. *Geocarto Int.* 37 (25), 10511–10531. doi: 10.1080/10106049.2022.2037730

Wang, J., Si, H., Gao, Z., and Shi, L. (2022b). Winter wheat yield prediction using an LSTM model from MODIS LAI products. *Agriculture* 12 (10), 1707. doi: 10.3390/agriculture12101707

Yu, Z., Ustin, S. L., Zhang, Z., Liu, H., Zhang, X., Meng, X., et al. (2020). Estimation of a new canopy structure parameter for rice using smartphone photography. *Sensors* 20, 4011. doi: 10.3390/s20144011

Yuan, Y., Wang, X., Yin, F., and Zhan, J. (2013). Examination of the quantitative relationship between vegetation canopy height and LAI. *Adv. Meteorology* 2013 (3), 1–6. doi: 10.1155/2013/964323

Zhang, Y., Chen, J. M., Miller, J. R., and Noland, T. L. (2008). Leaf chlorophyll content retrieval from airborne hyperspectral remote sensing imagery. *Remote Sens. Environ.* 112 (7), 3234–3247. doi: 10.1016/j.rse.2008.04.005

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 04 January 2024
DOI 10.3389/fpls.2023.1242948

Check for updates

OPEN ACCESS

EDITED BY
Jiangang Liu,
Chinese Academy of Agricultural Sciences (CAAS), China

REVIEWED BY
Magdi A. A. Mousa,
King Abdulaziz University, Saudi Arabia
Karansher Singh Sandhu,
Bayer Crop Science, United States

*CORRESPONDENCE
Minglu Tian
tianminglu@saas.sh.cn

†These authors have contributed equally to this work

‡These authors share first authorship

RECEIVED 20 June 2023
ACCEPTED 11 December 2023
PUBLISHED 04 January 2024

CITATION
Liu Y, Ban S, Wei S, Li L, Tian M, Hu D, Liu W and Yuan T (2024) Estimating the frost damage index in lettuce using UAV-based RGB and multispectral images.
Front. Plant Sci. 14:1242948.
doi: 10.3389/fpls.2023.1242948

COPYRIGHT
© 2024 Liu, Ban, Wei, Li, Tian, Hu, Liu and Yuan. 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.

Estimating the frost damage index in lettuce using UAV-based RGB and multispectral images

Yiwen Liu[1,2,3†‡], Songtao Ban[2,3†‡], Shiwei Wei[4], Linyi Li[1,2,3], Minglu Tian[2,3*], Dong Hu[2,3], Weizhen Liu[5] and Tao Yuan[2,3]

[1]College of Information Technology, Shanghai Ocean University, Shanghai, China, [2]Institute of Agricultural Science and Technology Information, Shanghai Academy of Agricultural Sciences, Shanghai, China, [3]Key Laboratory of Intelligent Agricultural Technology (Yangtze River Delta), Ministry of Agriculture and Rural Affairs, Shanghai, China, [4]Jinshan Experimental Station, Shanghai Agrobiological Gene Center, Shanghai, China, [5]School of Computer and Artificial Intelligence, Wuhan University of Technology, Wuhan, China

Introduction: The cold stress is one of the most important factors for affecting production throughout year, so effectively evaluating frost damage is great significant to the determination of the frost tolerance in lettuce.

Methods: We proposed a high-throughput method to estimate lettuce FDI based on remote sensing. Red-Green-Blue (RGB) and multispectral images of open-field lettuce suffered from frost damage were captured by Unmanned Aerial Vehicle platform. Pearson correlation analysis was employed to select FDI-sensitive features from RGB and multispectral images. Then the models were established for different FDI-sensitive features based on sensor types and different groups according to lettuce colors using multiple linear regression, support vector machine and neural network algorithms, respectively.

Results and discussion: Digital number of blue and red channels, spectral reflectance at blue, red and near-infrared bands as well as six vegetation indexes (VIs) were found to be significantly related to the FDI of all lettuce groups. The high sensitivity of four modified VIs to frost damage of all lettuce groups was confirmed. The average accuracy of models were improved by 3% to 14% through a combination of multisource features. Color of lettuce had a certain impact on the monitoring of frost damage by FDI prediction models, because the accuracy of models based on green lettuce group were generally higher. The MULTISURCE-GREEN-NN model with R^2 of 0.715 and RMSE of 0.014 had the best performance, providing a high-throughput and efficient technical tool for frost damage investigation which will assist the identification of cold-resistant green lettuce germplasm and related breeding.

KEYWORDS

lettuce, frost damage, unmanned aerial vehicle, high-throughput detection, multisource data

1 Introduction

Lettuce (*Lactuca sativa* L.) is one of the most widely consumed leafy vegetables worldwide with high nutritional value(Shi et al., 2022). It is also one of the most economically important vegetable crops in the world(Soldatenko et al., 2018), with China consistently leading in production (Shatilov et al., 2019; Wang et al., 2022). Lettuces prefer cool temperatures between 7 and 24°C with an average of 18°C(Jenni, 2005), and their growth will be retarded or stagnant when temperature goes below 7°C. Frost damage of lettuce is a stress caused by low temperatures and generally occurs below 0° C(Yu and Lee, 2020). When frost damage occurs, ice nuclei are formed outside the cells and ice crystals are gradually developed as long as the low temperature continues. When the ice crystals spread into the cells, irreversible damage will occur, causing the leaves to appear watery, yellow or dark brown and the whole plant to wilt. Fresh lettuce is not storage-resistant, and its supply relies on fresh harvesting. Frost can damage the outer leaves of mature lettuce, leading to decay in handling and storage(Turini et al., 2011). As a result, low temperature is one of the most important factors threatening to supply lettuce in winter (Pirinc and Alas, 2021). Under harsh environmental conditions, crop yields can be lost ranging from 50% to 70%(Francini and Sebastiani, 2019). Due to the low availability and high market demand for lettuce in winter (Han et al., 2014), it is urgent to breed cold-resistant lettuce cultivars to keep the yield of lettuce in winter. Meanwhile, frost damage investigation is an important basement in the breeding programs of cold-resistant lettuce. The traditional method relies on manual surveys plot-by-plot in the field, which is time-consuming, laborious, subjective, and low in efficiency, especially when the number of lettuce cultivars is large. Therefore, it is of great significance to develop high-throughput methods of frost damage investigation to improve breeding efficiency.

Remote sensing based on Unmanned Aerial Vehicle (UAV), a newly developed technique for high-throughput crop growth information acquisition, has been widely used in crop monitoring under growth (Jiang et al., 2022)and various stresses such as pests, diseases, water deficit, salt-stressed(Johansen et al., 2019) and frost (Wójtowicz et al., 2016; Perry et al., 2017a; Chen et al., 2019; Choudhury et al., 2019; Goswami et al., 2019; Jełowicki et al., 2020; Millan et al., 2020; Marin et al., 2021). Although satellite remote sensing technology was also used in frost damage monitoring(Feng et al., 2009; Romanov, 2009; Romani et al., 2011; Rudorff et al., 2012; She et al., 2015; She et al., 2017; Li et al., 2021; Gabbrielli et al., 2022a; Gabbrielli et al., 2022b), the UAV-based remote sensing is more accurate in the breeding field due to its high spatial resolution. UAVs, including DJI, 3D Robotics solo and Ebee, were equipped with spectral cameras to detect frost damage of crops such as wheat (Guo et al., 2014; Wang et al., 2014; Murphy et al., 2020), maize (Choudhury et al., 2019; Goswami et al., 2019; Shu et al., 2022), oat (Macedo-Cruz et al., 2011), oilseed rape(She et al., 2015), and coffee plants(Marin et al., 2021; Marin et al., 2022). In these studies, data analysis techniques such as Pearson correlation analysis and principal component analysis were used to extract stress-related spectral features including the reflectance of different spectral bands and several commonly used vegetation indexes (VIs) such as normalized difference vegetation index (NDVI)(She et al., 2015), green NDVI, photochemical reflectance index (PRI), carotenoid reflectance index(CRI), and anthocyanin reflectance index(ARI) (Choudhury et al., 2019; Marin et al., 2021; Marin et al., 2022). Pixel-based classification with thresholds, random forest, random committee, support vector machine (SVM) and other classification methods were used to predict frost damage degree in some of these studies(Goswami et al., 2019; Jełowicki et al., 2020). Besides, regression methods such as multiple linear regression (MLR) and principal component regression were employed to predict crop yield or other physical parameters and then their changes before and after frost damage were compared to assess stress severity(Guo et al., 2014; Choudhury et al., 2019). Currently, several studies have utilized RGB and multispectral image features for lettuce, such as prediction of lettuce health(Pham et al., 2019), detection of lettuce anthocyanin content(Kim and Van Iersel, 2023) and classification of lettuce seeds(Concepcion et al., 2020). However, as far as we know, few studies have been conducted to apply these techniques to detect the frost damage of lettuce in the field during the growth stage. Unlike other crops, different varieties of lettuce in breeding trial fields exhibit significant differences in both morphology (Iceberg, Batavian, Butterhead, etc.) and color (green, red, etc.). Therefore, it remains to be seen whether these spectral features, which are widely used in crop stress monitoring, are suitable for screening the frost damage of lettuce with different varieties, and whether these non-destructive and high-throughput methods for evaluating frost damage can be applied to field investigation of lettuce breeding materials.

Frost damage to lettuce not only leads to changes of appearance, but also affects its physiological and biochemical indexes. Spectral imagers have an advantage in responding to changes in the physiological and biochemical indexes of crops due to their ability to detect reflectance spectra in the visible and near infrared wavelengths (Tao et al., 2022). This is the reason why spectral imagers were chosen in the most existing researches on frost damage of field crops(Macedo-Cruz et al., 2011; Wang et al., 2014; Yang et al., 2019; Lassalle, 2021). However, the cost of multispectral cameras is high, and the resolution of multispectral images is generally low, with insufficient texture information in the images. On the other side, Red-Green-Blue (RGB) cameras have an advantage in responding to the surface characteristics of crops due to their high spatial resolution. Despite the low cost and high resolution of RGB cameras, they cannot capture spectral information beyond the visible spectrum. The combination of data from the RGB camera and the multispectral imager allows for a comprehensive analysis of the changes in lettuce after frost damage. In early research, various methodologies utilizing data from distinct sensors were contrasted to determine the most advantageous approach (Dammer et al., 2011). Currently, there have been studies of combining multisource image features for crop monitoring (Perry et al., 2017b; Zheng et al., 2020; Li et al., 2021), confirming the improvement of detection effect when multisource image features were integrated. Thus, it is worthwhile to try to use both RGB and multispectral images to evaluate the lettuce frost damage.

Therefore, the objectives of this study were to evaluate the frost damage in lettuce by analyzing UAV-based RGB and multispectral imagery. To accomplish this objective, correlation analysis with

frost damage index (FDI) was employed to find FDI-sensitive features, new FDI-sensitive VIs were constructed by modifying existing well-performing VIs and tested, and multivariate regression models using different algorithms were compared to explore the potential of lettuce FDI estimation in the field.

2 Materials and methods

2.1 Plant material and study area

The experiment was performed in the test site of Shanghai Academy of Agricultural Sciences in Fengxian District, Shanghai City, China (30.891°N, 121.359°E), as shown in Figure 1. Fengxian District is located in the alluvial plain of the Yangtze River Delta. It has a subtropical marine monsoon climate and the average annual temperature is about 15.8°C.

A total of 209 distinct cultivars of lettuce, consisting of 160 green and 49 red varieties, were randomly assigned to 209 plots. Each plot, measuring approximately 4 m^2 (4m x 1m), contained approximately 24 plants of each cultivar, as illustrated in Figure 1. All the lettuces were sown on September 24, 2020, and were transplanted to the field on October 22, 2020. The harvest period commenced around December 29, 2020.

2.2 Data acquisition of frost damage index

Figure 2 shows the minimum and maximum temperatures from Dec 15, 2020 to Jan 8, 2021. The temperatures from Dec 15, 2020 to Dec 28, 2020 and from Jan 2, 2021 to Jan 5, 2021 range from 0 to 15° C. Lettuce leaves remain undamaged at temperatures near freezing, but are susceptible to damage at temperatures below freezing (Turini et al., 2011). The first day when temperature went below 0°C after field-planting was Dec 29, 2020, on which day the temperature dropped to -5°C. The lowest temperature, which was -9°C, appeared on Jan 7, 2021. During these two periods, from Dec 29, 2020 to Jan 1, 2021 and from Jan 6, 2021 to Jan 8, 2021, different cultivars of lettuce suffered from frost damage of different degrees, as shown in Figure 3.

The field investigation of FDI of each lettuce cultivar plot was carried out on Jan 8, 2021. The FDI was defined by referring to the statistical method of other damage indexes, such as chilling injury index(Fernandez-Trujillo et al., 1998; Porat et al., 2000; González-Aguilar et al., 2004; Zhao et al., 2006; Yang et al., 2011; Pan et al., 2016) and leaf disease index(Wang et al., 2015), which involved in a variety of crop stress investigations. In this research, the severity of frost damage was graded according to the characteristics of frost damage, which was shown in Supplementary Table 1. Then the FDI was calculated using the following equation:

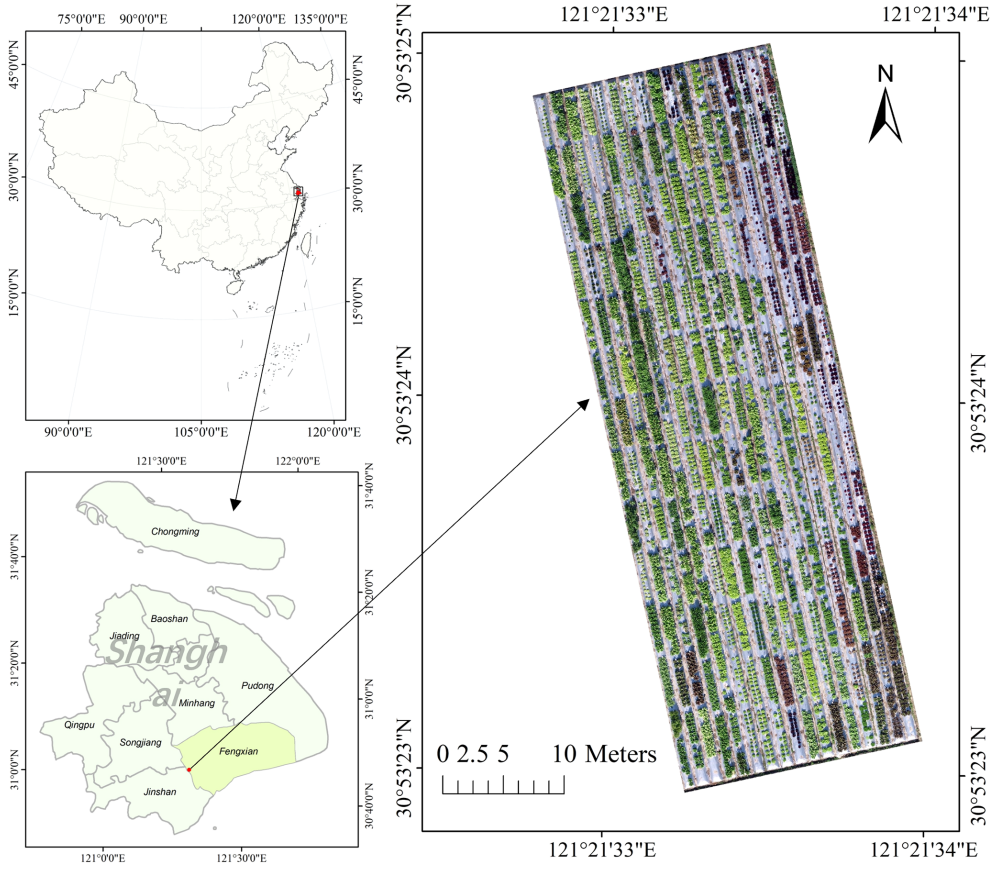


FIGURE 1
Location and RGB image of the experiment field in this study.

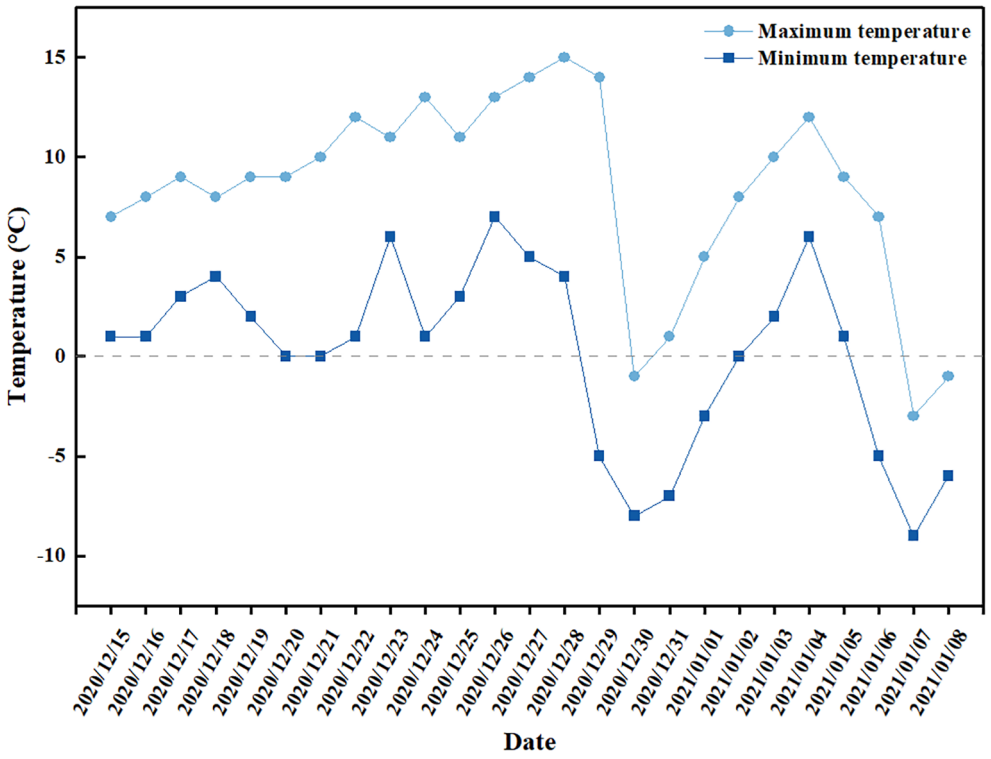


FIGURE 2
The temperature during the experiment in this study.

$$FDI = \frac{\sum_{n}^{i}(PN_i \times i)}{MAX_DL \times TPN}$$

where i is the damage level value, n is the number of damage levels counted in this plot, PN_i is the number of plants at damage level of i in this plot, MAX_DL is the maximum damage level, TPN is the total number of plants in this plot.

2.3 Acquisition and processing of images

Two aerial surveys were performed on Dec 15, 2020, when the temperature had not dropped below 0°C and there was no frost damage, and Jan 8, 2021, when the lettuce had suffered from low temperature and frost damage happened. In each survey, the UAV-based RGB and multispectral images of lettuce field were acquired respectively.

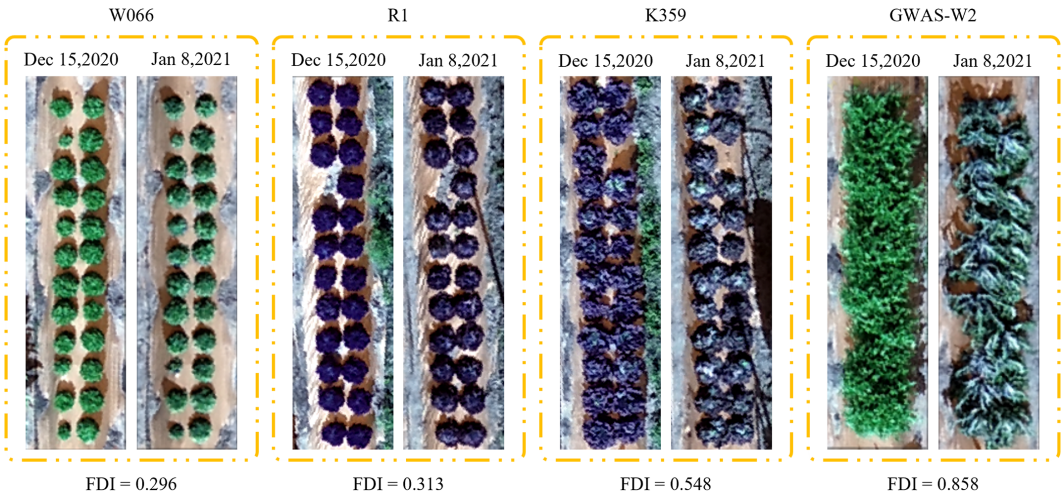


FIGURE 3
Lettuce before and lettuce after frost damage in this experiment. FDI represents frost damage index.

The RGB images were captured by a quadrotor named DJI Phantom 4 RTK (SZ DJI Technology Co., Shenzhen, China), which is a compact and lightweight UAV with a 20-megapixel RGB camera. The flight height of flight route was set as 30 m and the corresponding ground resolution of images was 0.012 m, and the forward and side overlaps were 80% and 70%, respectively.

The multispectral images were acquired using a five-band multispectral camera with a resolution of 1280 × 980 pixels, RedEdge-M (Micasense Inc., Seattle, WA, USA), mounted on a DJI M600 Pro UAV (SZ DJI Technology Co., Shenzhen, China). The central wavelength of each band with the corresponding bandwidth was 475nm (20nm), 560nm (20nm), 668nm (10nm), 717nm (10nm), and 840nm (40nm). The flight height of flight route was set as 30 m and the corresponding ground resolution of images was 0.018 m, and the forward and side overlaps were 85% and 80%, respectively. The images of reference panel which is a gray board with 50% reflectance by the size of 15.5 cm × 15.5 cm were captured before flight missions for radiometric calibration.

The RGB and multispectral images of the whole lettuce field were generated by mosaicking the originally acquired images within the aerial survey area using Pix4Dmapper Pro (PIX4D, Lausanne, Switzerland). The multispectral images were radiometrically corrected before mosaicking according to the images of reference panel. All the mosaiced images were geometrically corrected based on the RGB image of Dec 15, 2020 using ArcGIS (ESRI, Redlands, CA, USA).

The growth of lettuces almost stopped during Dec 15, 2020 to Jan 8, 2021 and their changes were mainly caused by frost damage. On Jan 8, 2021, many lettuces had reduced their coverage and lost biometric features because of frost damage, making it difficult to separate them from the background. Therefore, we chose the image on Dec 15, 2020 to extract pure lettuce regions for image features calculation from images on Jan 8, 2021. As NDVI is one of the most sensitive indexes to vegetation cover(Liu et al., 2020; Murphy et al., 2020), decision tree classification was performed on the multispectral image (MSI) of Dec 15, 2020, when the lettuces were in health status, to separate the lettuce plants from the background by setting NDVI greater than 0.5 as the rule using ENVI (Harris Geospatial Solutions, Inc., Broomfield, CO, America). Majority/Minority analysis was then applied to the classified results to reduce small plaques. Regions of interest of lettuce plants of each plot were generated based on the classification results and converted to shapefiles which were used to extract image and spectral features from the RGB and multispectral images of Jan 8, 2021 through zonal statistics tool.

2.4 Image and spectral features extraction

2.4.1 RGB image features

Frost damage will change the color and texture features of the RGB images of lettuce. These features will, in turn, provide information about the surface characteristics of lettuce after frost damage has occurred. The color features included the digital number (DN) of red, green, and blue channels, which were represented by R, G, and B, respectively. In addition, fifteen vegetation-related color indexes (CIs) were calculated based on DN, as listed in Table 1.

For each channel of the RGB image, 8 texture features, namely, Mean (M), Variance (V), Homogeneity (H), Contrast (Con), Dissimilarity (D), Entropy (E), Second Moment (SM), and Correlation (Cor) were calculated using Gray-level Co-occurrence Matrix (GLCM). Therefore, a total of 24 texture features were calculated and named with the initials of the color channels plus the abbreviation of the texture names. The calculation formulae of texture features are as follows:

$$Mean\ \ (M) = \sum_i \sum_j i \times p(i,j)$$

TABLE 1 Color indexes(CIs).

Index	Acronym	Equation	Reference
Blue Normalized Index	BNI	$\frac{B}{B+G+R}$	
Green Normalized Index	GNI	$\frac{G}{B+G+R}$	
Red Normalized Index	RNI	$\frac{R}{B+G+R}$	
Excess Green Vegetation Index	ExG	$2 \times G - B - R$	(Woebbecke et al., 1995)ADDIN
Visible Atmospherically Resistant Index	VARI	$\frac{G-R}{G+R-B}$	(Gitelson et al., 2002)
Excess Red Vegetation Index	ExR	$1.4 \times R - G$	(Meyer and Neto, 2008)
Excess Blue Vegetation Index	ExB	$1.4 \times B - G$	(Mao et al., 2003)
Excess Green minus Excess Red Vegetation Index	ExGR	ExG – ExR	(Camargo Neto, 2004)
Normalized Green-Red Difference Index	NGRDI	$\frac{G-R}{G+R}$	(Tucker, 1979)
Modified Green Red Vegetation Index	MGRVI	$\frac{G^2-R^2}{G^2+R^2}$	(Tucker, 1979)
Woebbecke Index	WI	$\frac{G-B}{R-G}$	(Woebbecke et al., 1995)
Kawashima Index	IKAW	$\frac{R-B}{R+B}$	(Kawashima and Nakatani, 1998)
Green Leaf Algorithm	GLA	$\frac{2 \times G - R - B}{2 \times G + R + B}$	(Louhaichi et al., 2001)
Red Green Blue Vegetation Index	RGBVI	$\frac{G^2 - B \times R}{G^2 + B \times R}$	(Bendig et al., 2015)
Vegetative	VEG	$\frac{G}{R^a B^{1-a}}$, a = 0.667	(Hague et al., 2006)

$$Variance\ \ (V) = \sum_i \sum_j (i-\mu)^2 p(i,j)$$

$$Homogeneity\ \ (H) = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$$

$$Contrast\ \ (Con) = \sum_{n=0}^{N_g} n^2 \left\{ \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j) \;\middle|\; |i-j| = n \right\}$$

$$Dissimilarity\ \ (D) = \sum_{n=1}^{N_g-1} n \left\{ \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j) \;\middle|\; |i-j| = n \right\}$$

$$Entropy\ \ (E) = -\sum_i \sum_j p(i,j) \log\left(p(i,j)\right)$$

$$Second\ \ Moment\ \ (SM) = \sum_i \sum_j \{p(i,j)\}^2$$

$$Correlation\ \ (Cor) = \frac{\sum_i \sum_j (i,j) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

where $p(i,j)$ is the value of the (i, j)th entry in the gray level cooccurrence matrix; N_g is the number of distinct gray levels in the quantized image; μ_x and σ_x are the mean and standard deviation of x rows in matrix calculation. μ_y and σ_y are the mean and standard deviation of y rows in the matrix calculation.

2.4.2 MSI features

In addition to the changes in surface characteristics, physiological and biochemical indexes of lettuce will also undergo alterations, leading to corresponding changes in the spectral reflectance of different bands. The features extracted from MSI included the spectral reflectance of each band and different VIs. R_{475}, R_{560}, R_{668}, R_{717} and R_{840} represented the reflectance value at specified bands. A total of twenty-three VIs were calculated based on the spectral reflectance at different bands according to the formulation in Table 2. Specifically, for lettuce frost damage evaluation in this study, experiential frost damage VIs (FD_VIs) were constructed by examining the correlativity between FDI and the spectral reflectance of each band, while referring to the existing FDI-sensitive VIs such as NDVI, EVI, and SIPI. These FD_VIs incorporated more bands and different constants, and their sensitivity to FDI was assessed through Pearson correlation analysis. Ultimately, four FD_VIs, namely FD_VI1, FD_VI2, FD_VI3, and FD_VI4, were established.

2.5 Statistical analysis and modeling algorithms

Pearson correlation analysis method was used to select the RGB and multispectral image features. The correlation analysis was employed between these features and FDI in each group. The features that reached a highly significant level (p<0.01) were selected. Three regression algorithms, namely MLR, SVM and neural network (NN), were used to establish the estimation models of lettuce FDI by taking selected color, texture and spectral features as independent variables.

MLR is a basic method in multiple regression analysis and widely used in remote sensing monitoring because of their good theoretical basis(Zhang et al., 2021). In this study, the MLR models were constructed based on the following calculation formula:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \epsilon_i$$

where Y_i, p and ϵ_i are the FDI of the ith sample, the number of independent variables and the ith independent identically distributed normal error; X_{ij} and β_j are respectively the jth independent variable and its coefficient of the ith sample (j =1, 2, …, p).

SVM is one of the commonly used regression methods to predict physiological parameters of crops(Shah et al., 2018; Zhang et al., 2022). The essence of SVM is to construct a set of planes or hyperplanes in a high or infinite dimensional space(Cortes and Vapnik, 1995). In this study, linear kernel function and sequential minimal optimization were chosen to construct the SVM models. The value of kernel scale was 1 and the approximations of box constraint and epsilon ranged from 0.18 to 0.2 and 0.018 to 0.02, respectively.

NN is a mathematical model that simulates the brain for information processing(Agatonovic-Kustrin and Beresford, 2000). The NN models are powerful predictive tools for crop growth status (Romero et al., 2018). The network structure of NN includes the number of hidden layers, the number of nodes in each layer, the initialization of weights, the training algorithm, and the learning rate. The NN models in this study used 10 hidden layers, and the training algorithm was the Levenberg-Marquardt algorithm.

The experimental cultivars contained lettuce in both green and red colors. There may be some influence of lettuce color on the predicted results. To understand this influence, the data was divided into three groups according to the color of lettuce: a group of all the lettuce (ALL), a group of green lettuce (GREEN) and a group of red lettuce (RED). Each group was divided into a training set and a test set by stratified random sampling according to the sample ratio of 7:3. Table 3 describes the statistical characteristics of the FDI of the samples.

The procedure of statistical analysis is summarized by the flowchart in Figure 4. Coefficient of determination (R-squared, R^2), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were used to evaluate the accuracy of the estimation models established with FDI as the dependent variable. The value of R^2

TABLE 2 Vegetation indexes(VIs).

Index	Acronym	Equation	Reference
Frost Damage Vegetation Index 1	FD_VI1	$\frac{R_{840}+R_{717}-R_{668}}{R_{840}+R_{717}+R_{668}}$	
Frost Damage Vegetation Index 2	FD_VI2	$\frac{R_{840}+R_{717}-R_{668}}{R_{840}+R_{717}+R_{668}-R_{475}}$	
Frost Damage Vegetation Index 3	FD_VI3	$\frac{R_{840}+R_{717}-R_{668}}{R_{840}+R_{717}+6\times R_{668}-7.5\times R_{475}+25}$	
Frost Damage Vegetation Index 4	FD_VI4	$\frac{R_{840}+R_{717}+R_{560}-R_{668}}{R_{840}+R_{717}+R_{668}+R_{560}-R_{475}+7}$	
Normalized Difference Vegetation Index	NDVI	$\frac{R_{840}-R_{668}}{R_{840}+R_{668}}$	(Rouse et al., 1973)
Simple Ratio Index	SR	$\frac{R_{840}}{R_{668}}$	(Pearson and Miller, 1972)
Enhanced Vegetation Index	EVI	$\frac{2.5\times(R_{840}-R_{668})}{R_{840}+6\times R_{668}-7.5\times R_{475}+1}$	(Huete et al., 1999)
Atmospherically Resistant Vegetation Index	ARVI	$\frac{R_{840}-(2\times R_{668}-R_{475})}{R_{840}+(2\times R_{668}-R_{475})}$	(Huete et al., 1994)
Red-edge Normalized Difference Vegetation Index	RENDVI	$\frac{R_{840}-R_{717}}{R_{840}+R_{717}}$	(Fitzgerald et al., 2010)
Modified Red Edge Simple Ratio Index	mSR	$\frac{R_{840}-R_{475}}{R_{717}+R_{475}}$	(Sims and Gamon, 2002)
Modified Red Edge Normalized Difference Vegetation Index	mNDVI	$\frac{R_{840}-R_{717}}{R_{840}+R_{717}-2\times R_{475}}$	(Huete et al., 1997)
Red-edge Ratio Vegetation Index	RERVI	$\frac{R_{840}}{R_{717}}$	(Gitelson et al., 2005)
Photochemical Reflectance Index	PRI	$\frac{R_{560}-R_{668}}{R_{560}+R_{668}}$	(Peñuelas et al., 1995)
Structure Insensitive Pigment Index	SIPI	$\frac{R_{840}-R_{475}}{R_{840}+R_{668}}$	(Blackburn, 1998)
Red Green Ratio Index	RG	$\frac{R_{668}}{R_{560}}$	(Gamon and Surfus, 1999)
Plant Senescence Reflectance Index	PSRI	$\frac{R_{717}-R_{475}}{R_{840}}$	(Merzlyak et al., 1999)
Carotenoid Reflectance Index 1	CRI1	$\frac{1}{R_{475}}-\frac{1}{R_{560}}$	(Gitelson et al., 2001b)
Carotenoid Reflectance Index 2	CRI2	$\frac{1}{R_{475}}-\frac{1}{R_{717}}$	(Gitelson et al., 2001b)
Anthocyanin Reflectance Index 1	ARI1	$\frac{1}{R_{560}}-\frac{1}{R_{717}}$	(Gitelson et al., 2001a)
Anthocyanin Reflectance Index 2	ARI2	$(\frac{1}{R_{560}}-\frac{1}{R_{717}})\times R_{840}$	(Gitelson et al., 2006)
Green Normalized Difference Vegetation Index	GNDVI	$\frac{R_{840}-R_{560}}{R_{840}+R_{560}}$	(Gitelson et al., 1996)
Green Ratio Vegetation Index	GRVI	$\frac{R_{840}}{R_{560}}$	(Gitelson et al., 2005)
Normalized Pigment/Chlorophyll Index	NPCI	$\frac{R_{668}-R_{475}}{R_{668}+R_{475}}$	(Peñuelas et al., 1994)

($0\leq R^2 \leq 1$) determines the degree of closeness of correlation. The larger R^2 is, the closer the relationship between dependent and independent variables is. RMSE and MAE are used to measure the deviation between the predicted and actual values. The smaller RMSE and MAE are, the closer the predicted values are to the actual values. Therefore, the closer R^2 is to 1, RMSE is to 0 and MAE is to 0, the higher the accuracy of model will be. The formulas are as follows:

TABLE 3 Descriptive statistics of the FDI of lettuce.

Statistical characteristics	ALL		GREEN		RED	
	Training set	Test set	Training set	Test set	Training set	Test set
Mean	0.525	0.53	0.518	0.517	0.562	0.541
Median	0.533	0.525	0.514	0.51	0.56	0.544
Mode	0.45	0.278	0.45	0.375	0.6	0.714
Standard deviation	0.181	0.189	0.188	0.19	0.164	0.172
Variance	0.033	0.036	0.035	0.036	0.027	0.03
Minimum value	0.2	0.2	0.2	0.214	0.286	0.29
Maximum value	1	1	1	1	0.991	0.842
Sample size	146	63	112	48	35	14

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where y_i, $\hat{y}_i$ and $\bar{y}$ are the actual FDI, the predicted FDI and the average of actual FDI, respectively; n is the number of samples.

3 Results

3.1 Correlation analysis between RGB and multispectral image features and FDI

Supplementary Figure 1 illustrates the correlation between the RGB and multispectral image features and FDI for three distinct lettuce groups. The results reveal that certain CIs demonstrated a significant correlation with FDI, while all texture features exhibited no noticeable correlation with FDI. The majority of MSI features displayed a significant correlation with FDI, and their correlation

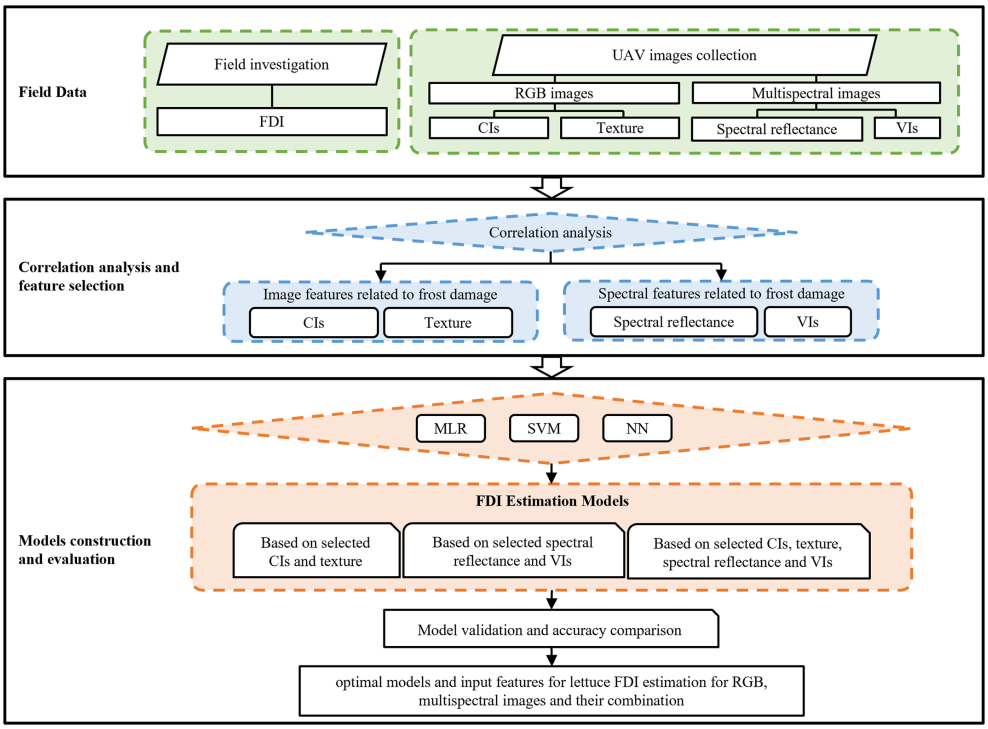


FIGURE 4
Experiment methodology and procedure of statistical analysis in this study. MLR, SVM and NN represent multiple linear regression, support vector machine and neural network, respectively.

trends with lettuce FDI remained largely consistent within each group. Notably, the correlations between certain features, primarily CIs, and FDI exhibited opposing trends for GREEN and RED. Subsequent to the correlation analysis (p<0.01), features with absolute Pearson correlation coefficients (absolute r) greater than 0.181, 0.202, and 0.358 were selected respectively for ALL, GREEN and RED for further analysis, as depicted in Figure 5.

For the ALL group, a total of 32 features were selected, including 15 features from the RGB images and 17 features from the multispectral images. The CIs derived from the RGB images exhibited low correlations with FDI (absolute r< 0.55), with ExR and ExB demonstrating absolute r values higher than 0.5, and the highest absolute r recorded at 0.549. The correlations between various MSI features and FDI displayed considerable disparity (absolute r arranged from 0.183 to 0.717), with R_{840}, the four FD_VIs and EVI surpassing an absolute r of 0.6, and the highest absolute r observed for EVI.

For the group of GREEN, 16 RGB image features and 21 MSI features were selected. The absolute r of RGB image features were all above 0.21 (mostly surpassing 0.55) with ExG recording the highest absolute r of 0.738. Specifically, GNI, ExG, GLA, and RGBVI exhibited absolute r values exceeding 0.72. As for the MSI features, FD_VI3, FD_VI4, EVI, FD_VI1, FD_VI2, NDVI, SIPI, ARVI, and R_840 demonstrated absolute r values greater than 0.65. Among them, FD_VI3 attained the maximum absolute r of 0.773.

For the group of RED, a total of 19 features were selected, including 3 RGB image features and 16 MSI features. The selected RGB image features were B, G and R with absolute r of 0.558, 0.523, and 0.492, respectively. They all showed positive correlations with FDI. The MSI features with absolute r above 0.6 were FD_VI1, FD_VI2, ARVI and NDVI, and FD_VI1 had the highest absolute r of 0.626. Additionally, R_{668} was the only feature that displayed a positive correlation with FDI.

Notably, the four newly proposed FD_VIs were explicitly correlated with FDI and displayed exceptional performance across the three groups of lettuce data, highlighting their broad utility in accurately evaluating the impact of frost damage.

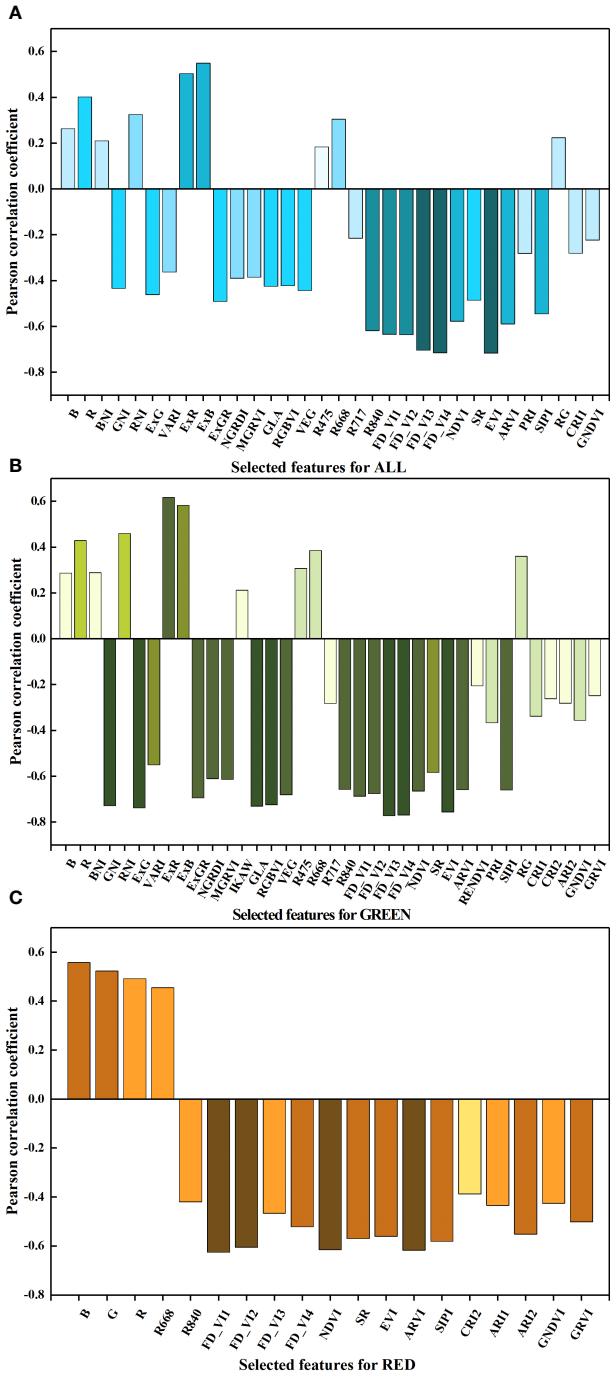


FIGURE 5
The features with significant correlation with FDI for the group of **(A)** ALL, **(B)** GREEN, and **(C)** RED.

3.2 FDI estimation models based on RGB image features

The estimation models were constructed with selected RGB image features as independent variables and FDI as a dependent variable by using MLR, SVM and NN algorithms. The accuracy of models was demonstrated in Table 4 and Figure 6. For the models of the group of ALL, the R^2 and RMSE of training set ranged from 0.465 to 0.56 and 0.014 to 0.132 with MAE between 0.094 and 0.104, and the R^2 and RMSE of test set ranged from 0.437 to 0.545 and 0.019 to 0.142 with MAE between 0.111 and 0.115. The FDI estimation accuracy of models for GREEN were highest among the three groups with the R^2 of training and test sets between 0.517 and 0.637, RMSE between 0.012 and 0.131 and MAE between 0.090 and 0.094. The performance of FDI estimation models for RED was neither sufficient nor stable with R^2 ranging from 0.223 to 0.386 for the training set, but from 0.357 to 0.737 for the test set. For all the three groups, the models using NN algorithm always achieved better accuracy and stability than others with higher R^2, lower RMSE and lower MAE, and the difference of R^2, RMSE and MAE between the training and test set were smaller.

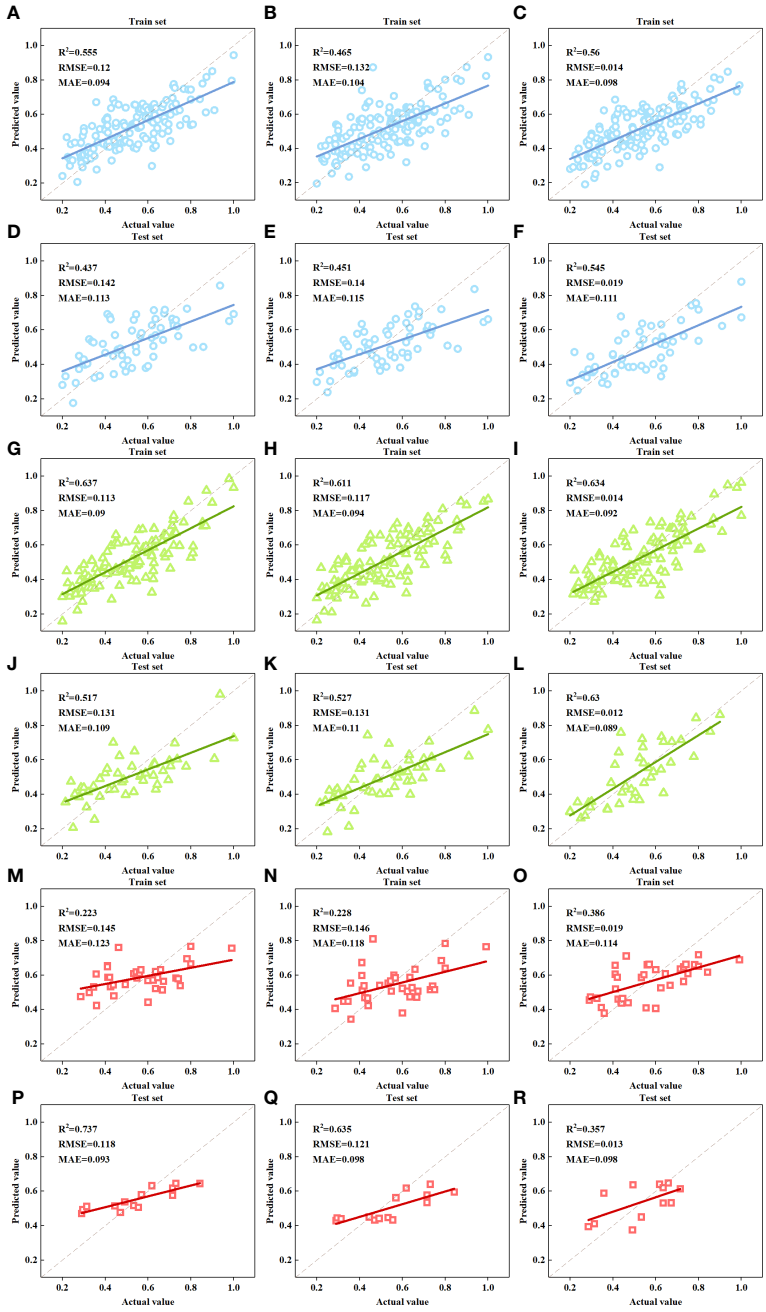


FIGURE 6
Scatter plots of the actual FDI and the predicted FDI. **(A, D).** RGB-ALL-MLR model; **(B, E).** RGB-ALL-SVM model; **(C, F).** RGB-ALL-NN model; **(G, J).** RGB-GREEN-MLR model; **(H, K).** RGB-GREEN-SVM model; **(I, L).** RGB-GREEN-NN model; **(M, P).** RGB-RED-MLR model; **(N, Q).** RGB-RED-SVM model; **(O, R).** RGB-RED-NN model.

3.3 FDI estimation models based on MSI features

The FDI estimation models based on selected MSI features were constructed using MLR, SVM and NN, respectively. Table 5 and Figure 7 illustrate the accuracy of FDI prediction. All the models based on MSI features had better performance than those based on RGB image features. The R^2 of models by different algorithms for the group of ALL in the training set were relatively close, ranging from 0.6 to 0.651; but the R^2 of test set varied from 0.484 to 0.639. The models for GREEN had better accuracy than other groups with R^2 up to 0.718 and 0.665, RSME down to 0.01 and 0.014 and MAE down to 0.080 and 0.090 for the training and test set, respectively. The models for RED were improved by using MSI features, but still not good enough for FDI estimation with R^2 lower than 0.5. As for the algorithms, NN was still superior to MLR and SVM.

TABLE 4 Accuracy of the FDI estimation models based on RGB image features.

Feature Source	Group	Model	Training set			Test set		
			R^2	RMSE	MAE	R^2	RMSE	MAE
RGB	ALL	MLR	0.555	0.120	0.094	0.437	0.142	0.113
		SVM	0.465	0.132	0.104	0.451	0.140	0.115
		NN	0.560	0.014	0.098	0.545	0.019	0.111
	GREEN	MLR	0.637	0.113	0.090	0.517	0.131	0.109
		SVM	0.611	0.117	0.094	0.527	0.131	0.110
		NN	0.634	0.014	0.092	0.630	0.012	0.089
	RED	MLR	0.223	0.145	0.123	0.737	0.118	0.093
		SVM	0.228	0.146	0.118	0.635	0.121	0.098
		NN	0.386	0.019	0.114	0.357	0.013	0.098

3.4 FDI estimation models based on multisource features

In order to make full use of the information from different data sources, the estimation models were constructed using all selected features from both RGB and multispectral images, as shown in Table 6 and Figure 8. Again, NN remained the best modeling algorithm for each group. In the training sets, The R^2 of the NN models for ALL, GREEN, and RED raised to 0.653, 0.722, and 0.592; while the RMSE and MAE decreased to 0.011, 0.009, and 0.009, and to 0.086, 0.074 and 0.082, respectively. Correspondingly, The R^2 of NN models in the test sets for each group increased to 0.694, 0.715, and 0.575, respectively; the RMSE dropped to 0.014, 0.014, and 0.018; and the MAE reduced to 0.097, 0.093 and 0.113.

Table 7 illustrates the average accuracy of FDI estimation models with different sensors and groups. It can be seen that the predictive performance of models with multisource features were better than those with the single-source features. Therefore, it can be concluded that the multisource features are helpful in improving the accuracy of FDI prediction.

4 Discussion

4.1 Response of RGB and multispectral features to FDI of lettuce with different colors

The damage caused by frost on lettuce will change the way how solar radiation interacts with lettuce leaf cells. When frost damage happened, chlorophyll broke down and Photosynthesis was weakened(Choudhury et al., 2019), which would cause the decrease of absorption of visible light, especially the blue and red light which were mainly used for photosynthesis(Gates et al., 1965). In consequence, the reflectance of red and blue bands increased with the FDI. As can be seen in Supplementary Figure 1B and R of RGB image as well as R_{475} and R_{668} of MSI were positively correlated with FDI. For the same reason, most of the VIs (including NDVI, SR, EVI, ARVI, and SIPI) that had been proven to be related to chlorophyll content on(Susantoro et al., 2017; Chen et al., 2021; Shu et al., 2022) had significant negative correlation with FDI.

TABLE 5 Accuracy of the FDI estimation models based on MSI features.

Feature Source	Group	Model	Training set			Test set		
			R^2	RMSE	MAE	R^2	RMSE	MAE
MSI	ALL	MLR	0.639	0.108	0.086	0.484	0.135	0.114
		SVM	0.600	0.114	0.089	0.516	0.131	0.104
		NN	0.651	0.012	0.086	0.639	0.012	0.084
	GREEN	MLR	0.713	0.100	0.080	0.644	0.114	0.095
		SVM	0.638	0.113	0.087	0.624	0.115	0.090
		NN	0.718	0.010	0.081	0.665	0.014	0.090
	RED	MLR	0.472	0.198	0.176	0.362	0.226	0.174
		SVM	0.467	0.119	0.091	0.141	0.159	0.143
		NN	0.485	0.013	0.094	0.421	0.039	0.169

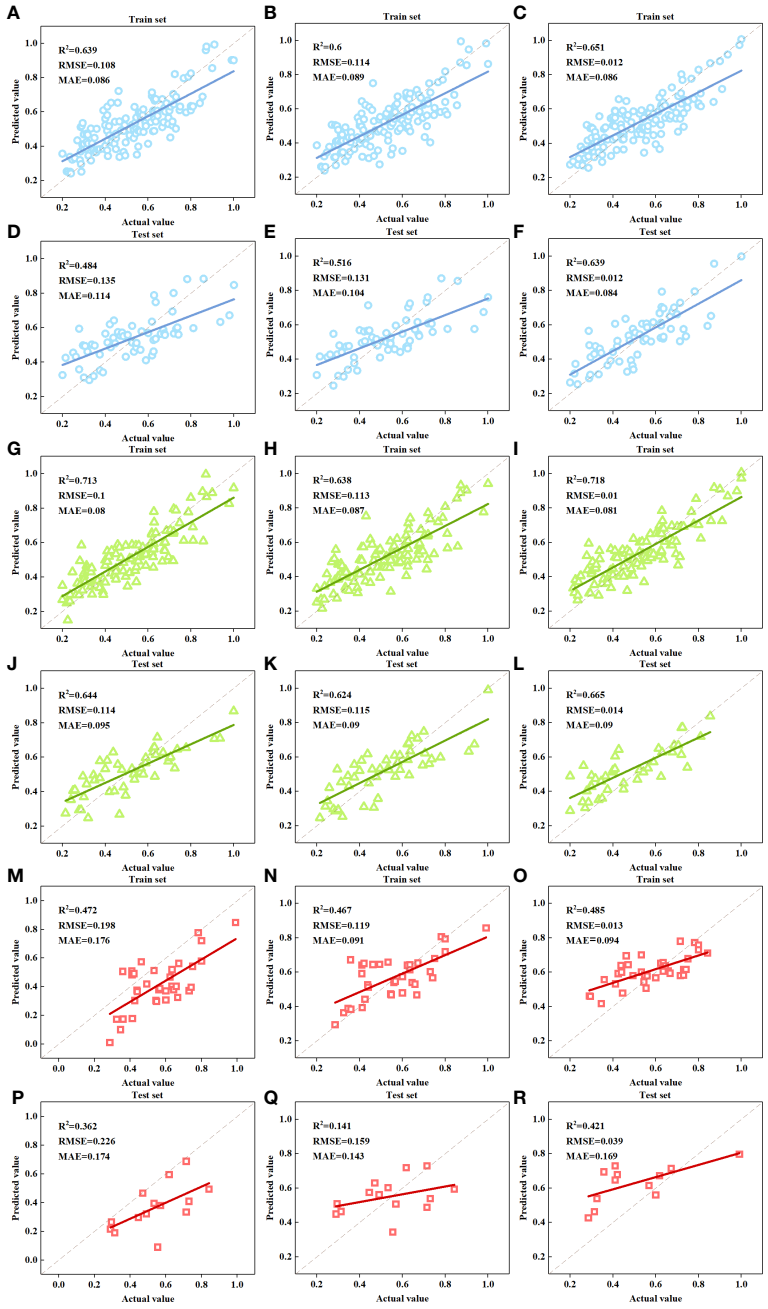


FIGURE 7
Scatter plots of the actual FDI and the predicted FDI. **(A, D).** MSI-ALL-MLR model; **(B, E).** MSI-ALL-SVM model; **(C, F).** MSI-ALL-NN model; **(G, J).** MSI-GREEN-MLR model; **(H, K).** MSI-GREEN-SVM model; **(I, L).** MSI-GREEN-NN model; **(M, P).** MSI-RED-MLR model; **(N, Q).** MSI-RED-SVM model; **(O, R).** MSI-RED-NN model.

The response of green channel of RGB image (G) and reflectance at green band of MSI (R_{560}) responded quite differently for green and red lettuce. The absolute r of G and R_{560} of red lettuce were much higher than that of green one. The possible reason is that red lettuce contained much more anthocyanin which absorbed the green light(Yang et al., 2016); when anthocyanins of lettuce decomposed due to frost damage, the absorption of green light was decreased and the reflection was enhanced; therefore, the positive relations between FDI and G and R_{560} of red lettuce were more significant. For the same reason, the anthocyanin reflectance indexes of MSI, ARI1 and ARI2, which reflected anthocyanin content, was more sensitive to FDI of red lettuce. For green lettuce, the change of color from green to yellow caused by frost damage was obvious. Although G and R_{560} did not achieve a significant correlation with FDI of the green lettuce, this change could still make the CIs(including GNI, ExG, VARI, ExGR, NGRDI, MGRVI, GLA, RGBVI, and VEG) which enhanced the green component show much closer relation to FDI of the green lettuce than that of the red one.

The destruction of cell structure due to frost damage would cause the decrease of reflectance at NIR band(Gates et al., 1965). As

TABLE 6 Accuracy of the FDI estimation models based on multisource features.

Feature Source	Group	Model	Training set			Test set		
			R^2	RMSE	MAE	R^2	RMSE	MAE
MULTISOURCE (RGB and MSI)	ALL	MLR	0.709	0.097	0.079	0.504	0.133	0.103
		SVM	0.660	0.105	0.081	0.581	0.122	0.099
		NN	0.653	0.011	0.086	0.649	0.014	0.097
	GREEN	MLR	0.752	0.093	0.076	0.684	0.106	0.089
		SVM	0.684	0.105	0.082	0.691	0.106	0.088
		NN	0.722	0.009	0.074	0.715	0.014	0.093
	RED	MLR	0.487	0.162	0.141	0.331	0.188	0.188
		SVM	0.497	0.116	0.087	0.201	0.151	0.151
		NN	0.592	0.009	0.082	0.575	0.018	0.113

a result, R_{840} was obviously negative correlation with FDI of all lettuces. The result is similar to that obtained in an experiment evaluating maize frost damage, where it was found that the frost damage caused a sharp decline of reflectance between 720 and 1350 nm(Choudhury et al., 2019).

Although Pearson correlation analysis showed that R_{717} and R_{560} were weakly correlated with FDI, the four proposed FD_VIs also showed high correlations with FDI. These improvements were mainly due to the involvement of red edge band (717nm), green band (560nm) and blue band (475nm), on the basis of the near-infrared band (840nm) and red band (668nm). In particular, the correlations between FD_VI4 of all five bands and FDI of green and red lettuce reached the maximum. The possible reason is that canopy spectrum is the result of comprehensive influence of multiple factors such as internal components of leaves and canopy structure, and there are synergistic changes between FDI and the spectrum of these bands.

The correlations between the texture features and the FDI were not significant. One possible reason is that there were many cultivars with different structures in this experiment, and the difference in texture features among cultivars exceeds the difference in texture caused by frost damage.

4.2 Effect of different sensors and algorithms on FDI estimation models of lettuce with different colors

The lettuce FDI estimation models based on MSI features had better performance than those based on RGB image features, which is consistent with previous studies on the vegetation coverage monitoring and nitrogen accumulation estimation of rice (Zheng et al., 2018; Furukawa et al., 2021). It was because that multispectral imager typically captured more information than RGB cameras(Shu et al., 2022). While RGB cameras capture only three color channels, multispectral imagers capture more pronounced changes in reflectance at multiple bands across the electromagnetic spectrum, including near-infrared bands. This allows for the calculation of vegetation indexes that are highly sensitive to changes in plant health and vigor. This additional information can be leveraged to better detect and quantify frost damage. Some relevant studies have proved that RGB images are generally used to monitor the early growth of crops, and the information about near-infrared band provided by multispectral images is more suitable for the later growth of crops(Marcial-Pablo et al., 2019). In addition, the use of multispectral images in some studies enhanced the monitoring of biodiversity(Tait et al., 2019; Wolff et al., 2023). In this research, the monitored field lettuces were already in the harvesting period, and the changes were no longer obvious as in the early growth stage. As a result, the changes captured by the RGB camera were limited. At the same time, many cultivars of lettuce were selected in this study, the diversity of which was suitable for monitoring with a multispectral camera. However, although multispectral imagers had the advantage of high accuracy, RGB cameras could also be an alternative selection for low-cost monitoring of frost damage, for green lettuce at least. The FDI predictive accuracy was further improved by taking the combination of both RGB and multispectral image features as model independent variables, in which way all the information related to frost damage was fully used and the problem of spectrum saturation can be solved(Zhou et al., 2021; Shu et al., 2022). The conclusion that multisource data fusion can improve model accuracy has also been demonstrated in previous studies on crop monitoring(Jiang et al., 2019; Liu et al., 2019; Zhu et al., 2021).

Compared to the models trained based on RED and ALL, the models trained based on GREEN worked better. The major reason for the low accuracy of RED models was that there were only 49 red lettuce plots, and the sample size of red lettuce was too small which led to severe overfitting in most models for RED. In the future studies, the number of red lettuce cultivars should be increased in the selection of experimental cultivars to improve the sample size of the red lettuce dataset. Since green and red lettuces had different levels of secondary metabolites, they presented different colors and produced different responses after frost damage, as manifested in section 3.1. Therefore, the accuracy of models for ALL were not as good as those for GREEN. But even in the best FDI estimation model for GREEN, the RGB and multispectral image features could

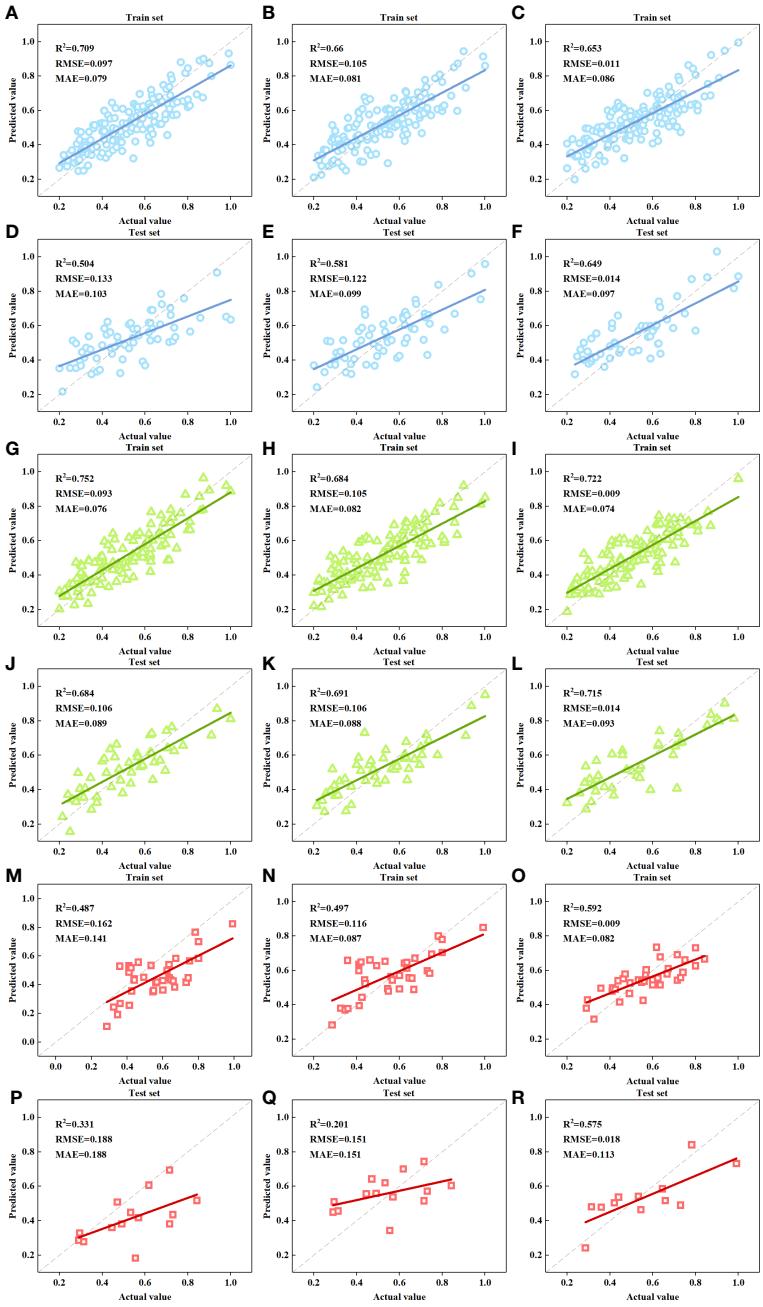


FIGURE 8
Scatter plots of the actual FDI and the predicted FDI. **(A, D).** MULTISOURCE-ALL-MLR model; **(B, E).** MULTISOURCE-ALL-SVM model; **(C, F).** MULTISOURCE-ALL-NN model; **(G, J).** MULTISOURCE-GREEN-MLR model; **(H, K).** MULTISOURCE-GREEN-SVM model; **(I, L).** MULTISOURCE-GREEN-NN model; **(M, P).** MULTISOURCE-RED-MLR model; **(N, Q).** MULTISOURCE-RED-SVM model; **(O, R).** MULTISOURCE-RED-NN model.

explain only about 70% of the variation. The differences between lettuces were not only caused by the different frost damage degrees, but also influenced by the different plant morphologies and green levels among different lettuce cultivars. The accuracy of models built by different regression algorithms did not show significant distinction for ALL and GREEN. NN models tended to be more stable than MLR and SVM with less difference of R^2 between training and test set.

5 Conclusions

This study showed the feasibility of using features derived from RGB and multispectral images collected by a UAV to estimate lettuce FDI. Especially, the accuracy of models with multisource features were higher than those with single-source features. Notably, the four newly proposed FD_VIs had certain universal correlation with frost damage of lettuce with different colors. The

TABLE 7 The average accuracy of FDI estimation models with different sensors and groups.

Feature Source	Group	Training set			Test set		
		R^2	RMSE	MAE	R^2	RMSE	MAE
RGB	ALL	0.527	0.089	0.099	0.478	0.100	0.113
	GREEN	0.627	0.081	0.092	0.558	0.091	0.103
	RED	0.279	0.103	0.118	0.576	0.084	0.096
MSI	ALL	0.630	0.078	0.087	0.546	0.093	0.101
	GREEN	0.690	0.074	0.083	0.645	0.081	0.092
	RED	0.475	0.110	0.120	0.308	0.141	0.162
MULTISOURCE (RGB and MSI)	ALL	0.674	0.071	0.082	0.578	0.089	0.100
	GREEN	0.719	0.069	0.078	0.696	0.075	0.090
	RED	0.525	0.096	0.103	0.369	0.119	0.151

findings could be applied for the prediction and evaluation lettuce resources tolerant to freeze by non-destructive, accurate, and high-throughput identification, providing genetic resources and theoretical basis for the cultivation and genetic improvement of new varieties of lettuce resistant to frost damage in the future. In subsequent works, the impact of differences among cultivars of lettuce on the evaluation of frost damage should be considered, and the sample size of red lettuce should be increased, so as to improve the evaluation models and improve the accuracy of evaluation results.

Data availability statement

Data, models, or codes generated or used in the course of the study are available on GitHub at https://github.com/kwcnmm/predict-FDI.

Author contributions

YL implemented this research and wrote the draft of the manuscript. SB and MT helped formulate research ideas and highlighted research objectives. SW provided research materials and field data. DH and TY provided guidance for technical issues. WL and LL provided constructive suggestions to clarify the expression of ideas. All authors contributed to the article and approved the submitted version.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by Shanghai Agriculture Applied Technology Development Program, China (Grant No. G20220401) and SAAS Program for Excellent Research Team (Grant No. 2022015).

Acknowledgments

We would like to thank the relevant workers of Shanghai Agrobiological Gene Center for their extensive assistance in lettuce planting and field investigation. Also, we would like to thank Institute of Agricultural Science and Technology Information of Shanghai Academy of Agricultural Sciences for its financial and equipment 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.2023.1242948/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Pearson correlation analysis between RGB and multispectral image features and FDI.

SUPPLEMENTARY TABLE 1
Damage level and description for assessing frost damage of lettuce.

References

Agatonovic-Kustrin, S., and Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J. Pharm. Biomed. Anal.* 22 (5), 717–727. doi: 10.1016/S0731-7085(99)00272-1

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

Blackburn, G. (1998). Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. *Int. J. Remote Sens.* 19 (4), 657–675. doi: 10.1080/014311698215919

Camargo Neto, J. (2004). *A combined statistical-soft computing approach for classification and mapping weed species in minimum -tillage systems* (Lincoln, Nebraska, United States: Ph.D, University of Nebraska - Lincoln).

Chen, S., Gao, Y., Fan, K., Shi, Y., Luo, D., Shen, J., 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, Y., Sidhu, H., Kaviani, M., McElroy, M., Pozniak, C., and Navabi, A. (2019). Application of image-based phenotyping tools to identify QTL for in-field winter survival of winter wheat (*Triticum aestivum L.*). *Theor. Appl. Genet.* 132 (9), 2591–2604. doi: 10.1007/s00122-019-03373-6

Choudhury, B., Webster, R., Sharma, V., Goswami, J., Meetei, T., Krishnappa, R., et al. (2019). Frost damage to maize in northeast India: assessment and estimated loss of yield by hyperspectral proximal remote sensing. *J. Appl. Remote Sens.* 13 (4), 44527. doi: 10.1117/1.JRS.13.044527

Concepcion, R., Lauguico, S., Siphengphet, K., Alejandrino, J., Dadios, E., and Bandala, A. (2020). "Variety classification of *Lactuca sativa* seeds using single-kernel RGB images and spectro-textural-morphological feature-based machine learning," in *IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)).* (New York, United States: IEEE).

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20 (3), 273–297. doi: 10.1007/BF00994018

Dammer, K., Möller, B., Rodemann, B., and Heppner, D. (2011). Detection of head blight (Fusarium ssp.) in winter wheat by color and multispectral image analyses. *Crop Prot.* 30 (4), 420–428. doi: 10.1016/j.cropro.2010.12.015

Feng, M., Yang, W., Cao, L., and Ding, G. (2009). Monitoring winter wheat freeze injury using multi-temporal MODIS data. *Agric. Sci. China* 8 (9), 1053–1062. doi: 10.1016/s1671-2927(08)60313-2

Fernandez-Trujillo, J. P., Cano, A., and Artes, F. (1998). Physiological changes in peaches related to chilling injury and ripening. *Postharvest Biol. Technol.* 13 (2), 109–119. doi: 10.1016/s0925-5214(98)00006-4

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 (3), 318–324. doi: 10.1016/j.fcr.2010.01.010

Francini, A., and Sebastiani, L. (2019). Abiotic stress effects on performance of horticultural crops. *Horticulturae* 5 (4), 67. doi: 10.3390/horticulturae5040067

Furukawa, F., Laneng, L. A., Ando, H., Yoshimura, N., Kaneko, M., and Morimoto, J. (2021). Comparison of RGB and multispectral unmanned aerial vehicle for monitoring vegetation coverage changes on a landslide area. *Drones* 5 (3), 97. doi: 10.3390/drones5030097

Gabbrielli, M., Corti, M., Nicola, F., Maro, P., and Bechini, L. (2022a). "Detection of winterkill events of white mustard (*Sinapis alba L.*) by satellite-based remote sensing," in *Book of abstract* (Bolzano, Italy: Unibz), 1–1.

Gabbrielli, M., Corti, M., Perfetto, M., Fassa, V., and Bechini, L. (2022b). Satellite-based frost damage detection in support of winter cover crops management: a case study on white mustard. *Agronomy-Basel* 12 (9), 2025. doi: 10.3390/agronomy12092025

Gamon, J., and Surfus, J. (1999). "Assessing leaf pigment content and activity with a reflectometer," in *Proceedings of the 4th New Phytologist Symposium on At the Crossroads of Plant Physiology and Ecology, at the Meeting of the ENSA-M: New Phytologist.* (Hoboken, New Jersey, United States: Wiley), 105–117.

Gates, D., Keegan, H., Schleter, J., and Weidner, V. (1965). Spectral properties of plants. *Appl. Optics* 4 (1), 11–20. doi: 10.1364/AO.4.000011

Gitelson, A., Kaufman, Y., and Merzlyak, M. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* 58 (3), 289–298. doi: 10.1016/S0034-4257(96)00072-7

Gitelson, A., Kaufman, Y., Stark, R., and Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* 80 (1), 76–87. doi: 10.1016/S0034-4257(01)00289-9

Gitelson, A., Keydan, G., and Merzlyak, M. (2006). Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. *Geophysical Res. Lett.* 33 (11), L11402. doi: 10.1029/2006GL026457

Gitelson, A., Merzlyak, M., and Chivkunova, O. (2001a). Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochem. Photobiol.* 74 (1), 38–45. doi: 10.1562/0031-8655(2001)074<0038:OPANEO>2.0.CO;2

Gitelson, A., Merzlyak, M., Zur, Y., Stark, R., and Gritz, U. (2001b). "Non-destructive and remote sensing techniques for estimation of vegetation status," in *Proceedings of the 3rd European Conference on Precision Agriculture: Papers in Natural Resources.* (Montpelier, France: L'Institut Agro Montpellier), 273.

Gitelson, A., Viña, A., Ciganda, V., Rundquist, D., and Arkebauer, T. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophysical Res. Lett.* 32 (8), L08403. doi: 10.1029/2005GL022688

González-Aguilar, G. A., Tiznado-Hernández, M. E., Zavaleta-Gatica, R., and Martínez-Téllez, M. A. (2004). Methyl jasmonate treatments reduce chilling injury and activate the defense response of guava fruits. *Biochem. Biophys. Res. Commun.* 313 (3), 694–701. doi: 10.1016/j.bbrc.2003.11.165

Goswami, J., Sharma, V., Chaudhury, B., and Raju, P. (2019). Rapid identification of abiotic stress (frost) in in-filed maize crop using UAV remote sensing. *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.* XLII-3/W6, 467–471. doi: 10.5194/isprs-archives-XLII-3-W6-467-2019

Guo, Q., Li, J., Yu, W., and Zhang, H. (2014). Study on hyperspectral sensitivity index of winter wheat after freezing injury at mid winter period. *Chin. J. Agrometeorol.* 35 (06), 708–716. doi: 10.3969/j.issn.1000-6362.2014.06.015

Hague, T., Tillett, N., and Wheeler, H. (2006). Automated crop and weed monitoring in widely spaced cereals. *Precis. Agric.* 7 (1), 21–32. doi: 10.1007/s11119-005-6787-1

Han, Y., Li, Y., Fan, S., Liu, C., Hao, J., Chen, Q., et al. (2014). "Screening and identification of lettuce germplasm for tolerance to high and low temperature," in *Proceedings of the 29th International Horticultural Congress on Horticulture - Sustaining Lives, Livelihoods and Landscapes (IHC) / International Symposium on Plant Breeding in Horticulture* (Belgium: Acta Horticulturae), 381–387.

Huete, A., Justice, C., and Liu, H. (1994). Development of vegetation and soil indices for MODIS-EOS. *Remote Sens. Environ.* 49 (3), 224–234. doi: 10.1016/0034-4257(94)90018-3

Huete, A., Justice, C., and Van Leeuwen, W. (1999). *MODIS vegetation index (MOD13)". (Algorithm Theoretical Basis Document Version 3.1).* (Tucson, USA: Vegetation Index & Phenology Lab). Available at: https://www.cen.uni-hamburg.de/en/icdc/data/land/docs-land/modis-collection6-vegetation-index-atbd-mod13-v03-1.pdf.

Huete, A., Liu, H., Batchily, K., and Van Leeuwen, W. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sens. Environ.* 59 (3), 440–451. doi: 10.1016/S0034-4257(96)00112-5

Jełowicki, L., Sosnowicz, K., Ostrowski, W., Osińska-Skotak, K., and Bakuła, K. (2020). Evaluation of rapeseed winter crop damage using UAV-based multispectral imagery. *Remote Sens.* 12 (16), 2618. doi: 10.3390/rs12162618

Jenni, S. (2005). Rib discoloration: A physiological disorder induced by heat stress in crisphead lettuce. *HortScience* 40 (7), 2031–2035. doi: 10.21273/HORTSCI.40.7.2031

Jiang, Q., Fang, S. H., Peng, Y., Gong, Y., Zhu, R. S., Wu, X. T., et al. (2019). UAV-based biomass estimation for rice-combining spectral, TIN-based structural and meteorological features. *Remote Sens.* 11 (7), 890. doi: 10.3390/rs11070890

Jiang, J. L., Johansen, K., Stanschewski, C. S., Wellman, G., Mousa, M. A. 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 (3), 961–983. doi: 10.1007/s11119-021-09870-3

Johansen, K., Morton, M. J. L., Malbeteau, Y., Aragon, B., Al-Mashharawi, S., Ziliani, M., et al. (2019). Predicting biomass and yield at harvest of salt-stressed tomato plants using UAV imagery. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII-2/W13, 407–411. doi: 10.5194/isprs-archives-XLII-2-W13-407-2019

Kawashima, S., and Nakatani, M. (1998). An algorithm for estimating chlorophyll content in leaves using a video camera. *Ann. Bot.* 81 (1), 49–54. doi: 10.1006/anbo.1997.0544

Kim, C., and Van Iersel, M. W. (2023). Image-based phenotyping to estimate anthocyanin concentrations in lettuce. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1155722

Lassalle, G. (2021). Monitoring natural and anthropogenic plant stressors by hyperspectral remote sensing: Recommendations and guidelines based on a meta-review. *Sci. Total Environ.* 788, 147758. doi: 10.1016/j.scitotenv.2021.147758

Li, W., Huang, J., Yang, L., Chen, Y., Fang, Y., Jin, H., et al. (2021). A practical remote sensing monitoring framework for late frost damage in wine grapes using multi-source satellite data. *Remote Sens.* 13 (16), 3231. doi: 10.3390/rs13163231

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

Liu, Z., Liu, H., Luo, C., Yang, H., Meng, X., Ju, Y., et al. (2020). Rapid extraction of regional-scale agricultural disasters by the standardized monitoring model based on google earth engine. *Sustainability* 12 (16), 6497. doi: 10.3390/su12166497

Louhaichi, M., Borman, M., and Johnson, D. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto Int.* 16 (1), 65–70. doi: 10.1080/10106040108542184

Macedo-Cruz, A., Pajares, G., Santos, M., and Villegas-Romero, I. (2011). Digital image sensor-based assessment of the status of oat (*Avena sativa L.*) crops after frost damage. *Sensors* 11 (6), 6015–6036. doi: 10.3390/s110606015

Mao, W., Wang, Y., and Wang, Y. (2003). "Real-time detection of between-row weeds using machine vision," in *2003 ASAE Annual Meeting*. (Michigan, United States: American Society of Agricultural and Biological Engineers), 031004.

Marcial-Pablo, M. D., Gonzalez-Sanchez, A., Jimenez-Jimenez, S. I., Ontiveros-Capurata, R. E., and Ojeda-Bustamante, W. (2019). Estimation of vegetation fraction using RGB and multispectral images from UAV. *Int. J. Remote Sens.* 40 (2), 420–438. doi: 10.1080/01431161.2018.1528017

Marin, D., Araújo, G., Feraz, S., Schwerz, F., Barata, R., Faria, R., et al. (2022). "Analysis of the potential of remote piloted aircraft in the detection of damage by frost in coffee plants," in *8th International Engineering, Sciences and Technology Conference (IESTEC)*. (Los Alamitos, United States: IEEE Computer Soc), 654–658.

Marin, D., Ferraz, G., Schwerz, F., Barata, R., Faria, R., and Dias, J. (2021). Unmanned aerial vehicle to evaluate frost damage in coffee plants. *Precis. Agric.* 22 (6), 1845–1860. doi: 10.1007/s11119-021-09815-w

Merzlyak, M., Gitelson, A., Chivkunova, O., and Rakitin, V. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum* 106 (1), 135–141. doi: 10.1034/j.1399-3054.1999.106119.x

Meyer, G., and Neto, J. (2008). Verification of color vegetation indices for automated crop imaging applications. *Comput. Electron. Agric.* 63 (2), 282–293. doi: 10.1016/j.compag.2008.03.009

Millan, V., Rankine, C., and Sanchez-Azofeifa, G. (2020). Crop loss evaluation using digital surface models from unmanned aerial vehicles data. *Remote Sens.* 12 (6), 981. doi: 10.3390/rs12060981

Murphy, M., Boruff, B., Callow, J., and Flower, K. (2020). Detecting frost stress in wheat: a controlled environment hyperspectral study on wheat plant components and implications for multispectral field sensing. *Remote Sens.* 12 (3), 477. doi: 10.3390/rs12030477

Pan, L. Q., Zhang, Q., Zhang, W., Sun, Y., Hu, P. C., and Tu, K. (2016). Detection of cold injury in peaches by hyperspectral reflectance imaging and artificial neural network. *Food Chem.* 192, 134–141. doi: 10.1016/j.foodchem.2015.06.106

Pearson, R., and Miller, L. (1972). *Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie* (Ann Arbor, Michigan, United States: Remote Sensing of Environment, VIII: Willow Run Laboratories, Environmental Research Institute of Michigan), 1355.

Peñuelas, J., Filella, I., and Gamon, J. (1995). Assessment of photosynthetic radiation-use efficiency with spectral reflectance. *New Phytol.* 131 (3), 291–296. doi: 10.1111/j.1469-8137.1995.tb03064.x

Peñuelas, J., Gamon, J., Fredeen, A., Merino, J., and Field, C. (1994). Reflectance indices associated with physiological changes in nitrogen-and water-limited sunflower leaves. *Remote Sens. Environ.* 48 (2), 135–146. doi: 10.1016/0034-4257(94)90136-8

Perry, E., Nuttall, J., Wallace, A., Delahunty, A., and Barlow, K. (2017a). "Evaluating canopy reflectance for assessment of frost damage in wheat," in *Proceedings of the 18th ASA Conference*. (Ballarat, Australia: Australian Society of Agronomy), 1–4.

Perry, E., Nuttall, J., Wallace, A., and Fitzgerald, G. (2017b). In-field methods for rapid detection of frost damage in Australian dryland wheat during the reproductive and grain-filling phase. *Crop Pasture Sci.* 68 (6), 516–526. doi: 10.1071/cp17135

Pham, F., Raheja, A., and Bhandari, S. (2019). "Machine learning models for predicting lettuce health using UAV imageries," in *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV*. (Bellingham, Washington, United States: Spie-Int Soc Optical Engineering).

Pirinc, V., and Alas, E. (2021). The effects of applying natural plant antifreeze under low temperature conditions on lettuce (*Lactuca sativa L.*) yield and quality. *Appl. Ecol. Environ. Res.* 19 (4), 2963–2970. doi: 10.15666/aeer/1904_29632970

Porat, R., Pavoncello, D., Peretz, J., Ben-Yehoshua, S., and Lurie, S. (2000). Effects of various heat treatments on the induction of cold tolerance and on the postharvest qualities of 'Star Ruby' grapefruit. *Postharvest Biol. Technol.* 18 (2), 159–165. doi: 10.1016/s0925-5214(99)00075-7

Romani, L., Goncalves, R., Amaral, B., Chino, D., Zullo, J., Traina, C., et al. (2011). "Clustering analysis applied to NDVI/NOAA multitemporal images to improve the monitoring process of sugarcane crops," in *2011 6th International Workshop on the Analysis of Multi-temporal Remote Sensing Images (Multi-Temp)*. (New York, United States: IEEE).

Romanov, P. (2009). "Satellite-derived information on snow cover for agriculture applications in Ukraine," in *Advanced research workshop on using satellite and in situ data to improve sustainability* (Dordrecht, Netherlands: Springer), 81–91.

Romero, M., Luo, Y., Su, B., and Fuentes, S. (2018). Vineyard water status estimation using multispectral imagery from an UAV platform and machine learning algorithms for irrigation scheduling management. *Comput. Electron. Agric.* 147, 109–117. doi: 10.1016/j.compag.2018.02.013

Rouse, J. Jr., Haas, R., Schell, J., and Deering, D. (1973). *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation* (Washington D.C., United States: NASA Technical Reports Server (NTRS)).

Rudorff, B., Aguiar, D., Adami, M., and Salgado, M. (2012). "Frost damage detection in sugarcane crop using MODIS images and SRTM data," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS))*. (New York, United States: IEEE), 5709–5712.

Shah, A., Dubey, A., Hemnani, V., Gala, D., and Kalbande, D. (2018). "Smart farming system: crop yield prediction using regression techniques," in *Proceedings of International Conference on Wireless Communication, ICWICOM 2017*. (Switzerland: Springer Singapore), 49–56.

Shatilov, M., Razin, A., and Ivanova, M. (2019). Analysis of the world lettuce market. *IOP Conf. Ser. Earth Environ. Sci.* 395, 12053. doi: 10.1088/1755-1315/395/1/012053

She, B., Huang, J., Guo, R., Wang, H., and Wang, J. (2015). Assessing winter oilseed rape freeze injury based on Chinese HJ remote sensing data. *J. Zhejiang University-Science B* 16 (2), 131–144. doi: 10.1631/jzus.B1400150

She, B., Huang, J., Zhang, D., and Huang, L. (2017). Assessing and characterizing oilseed rape freezing injury based on MODIS and MERIS data. *Int. J. Agric. Biol. Eng.* 10 (3), 143–157. doi: 10.3965/j.ijabe.20171003.2721

Shi, M., Gu, J., Wu, H., Rauf, A., Emran, T., Khan, Z., et al. (2022). Phytochemicals, nutrition, metabolism, bioavailability, and health benefits in lettuce-a comprehensive review. *Antioxidants* 11 (6), 1158. doi: 10.3390/antiox11061158

Shu, M., Fei, S., Zhang, B., Yang, X., Guo, Y., Li, B., et al. (2022). Application of UAV multisensor data and ensemble approach for high-throughput estimation of maize phenotyping traits. *Plant Phenomics* 2022, 9802585. doi: 10.34133/2022/9802585

Sims, D., and Gamon, J. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.* 81 (2-3), 337–354. doi: 10.1016/S0034-4257(02)00010-X

Soldatenko, A., Pivovarov, V., Razin, A., Meshcheryakova, R., Shatilov, M., Ivanova, M., et al. (2018). The economy of vegetable growing: the state and the present. *Vegetable Crops Russia* 5), 63–68. doi: 10.18619/2072-9146-2018-5-63-68

Susantoro, T., Wikantika, K., Saepuloh, A., and Harsolumakso, A. (2017). "Selection of vegetation indices for mapping the sugarcane condition around the oil and gas field of North West Java Basin, Indonesia," in *Proceedings of the 4th International Symposium on LAPAN-IPB Satellite for Food Security and Environmental Monitoring (LISAT-FSEM): IOP Conference Series-Earth and Environmental Science*. (Amsterdam, Netherlands: Elsevier Science BV), 012001.

Tait, L., Bind, J., Charan-Dixon, H., Hawes, I., Pirker, J., and Schiel, D. (2019). Unmanned aerial vehicles (UAVs) for monitoring macroalgal biodiversity: comparison of RGB and multispectral imaging sensors for biodiversity assessments. *Remote Sens.* 11 (19), 2332. doi: 10.3390/rs11192332

Tao, M., He, Y., Bai, X., Chen, X., Wei, Y., Peng, C., et al. (2022). Combination of spectral index and transfer learning strategy for glyphosate-resistant cultivar identification. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.973745

Tucker, C. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8 (2), 127–150. doi: 10.1016/0034-4257(79)90013-0

Turini, T., Cahn, M., Cantwell, M., Jackson, L., Koike, S., Natwick, E., et al. (2011). *Iceberg lettuce production in California* (California, United States: University of California Agriculture and Natural Resources), 7215. doi: 10.3733/ucanr.7215

Wang, H., Qin, F., Liu, Q., Ruan, L., Wang, R., Ma, Z., et al. (2015). Identification and disease index inversion of wheat stripe rust and wheat leaf rust based on hyperspectral data at canopy level. *J. Spectrosc.* 2015, 1–10. doi: 10.1155/2015/651810

Wang, H., Wang, J., Dong, Y., Gu, X., and Huo, Z. (2014). Monitoring freeze stress levels on winter wheat from hyperspectral reflectance data using principal component analysis. *Spectrosc. Spectral Anal.* 34 (5), 1357–1361. doi: 10.3964/j.issn.1000-0593(2014)05-1357-05

Wang, Z., Yang, R., Liang, Y., Zhang, S., Zhang, Z., Sun, C., et al. (2022). Comparing efficacy of different biostimulants for hydroponically grown lettuce (*Lactuca sativa* L.). *Agronomy-Basel* 12 (4), 786. doi: 10.3390/agronomy12040786

Woebbecke, D., Meyer, G., Von Bargen, K., and Mortensen, D. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Trans. ASAE* 38 (1), 259–269. doi: 10.13031/2013.27838

Wójtowicz, M., Wójtowicz, A., and Piekarczyk, J. (2016). Application of remote sensing methods in agriculture. *Commun. Biometry Crop Sci.* 11 (1), 31–50. Available at: http://agrobiol.sggw.pl/~cbcs/content_11_1.php.

Wolff, F., Kolari, T. H. M., Villoslada, M., Tahvanainen, T., Korpelainen, P., Zamboni, P. A. P., et al. (2023). RGB vs. Multispectral imagery: Mapping aapa mire plant communities with UAVs. *Ecol. Indic.* 148, 110140. doi: 10.1016/j.ecolind.2023.110140

Yang, H., Wu, F., and Cheng, J. (2011). Reduced chilling injury in cucumber by nitric oxide and the antioxidant response. *Food Chem.* 127 (3), 1237–1242. doi: 10.1016/j.foodchem.2011.02.011

Yang, W., Yang, C., Hao, Z., Xie, C., and Li, M. (2019). Diagnosis of plant cold damage based on hyperspectral imaging and convolutional neural network. *IEEE Access* 7, 118239–118248. doi: 10.1109/access.2019.2936892

Yang, X., Zhang, J., Guo, D., Xiong, X., Chang, L., Niu, Q., et al. (2016). "Measuring and evaluating anthocyanin in lettuce leaf based on color information," in *Proceedings of the 5th IFAC Conference on Sensing, Control and Automation Technologies for Agriculture (AGRICONTROL): IFAC-PapersOnLine*. (Amsterdam Netherlands: Elsevier Science BV) 96–99.

Yu, D., and Lee, H. (2020). Evaluation of freezing injury in temperate fruit trees. *Horticulture Environ. Biotechnol.* 61 (5), 787–794. doi: 10.1007/s13580-020-00264-4

Zhang, L., Han, W., Niu, Y., Chávez, J., Shao, G., and Zhang, H. (2021). Evaluating the sensitivity of water stressed maize chlorophyll and structure based on UAV derived vegetation indices. *Comput. Electron. Agric.* 185, 106174. doi: 10.1016/j.compag.2021.106174

Zhang, Y., Zhao, Z., and Zheng, J. (2022). Estimation of reference crop evapotranspiration under different combination of meteorological elements using multivariate adaptive regression splines. *Geomatics Inf. Sci. Wuhan Univ.* 47 (5), 789–798. doi: 10.13203/j.whugis20190337

Zhao, Z., Jiang, W., Cao, J., Zhao, Y., and Gu, Y. (2006). Effect of cold-shock treatment on chilling injury in mango (Mangifera indica L. cv. Wacheng) fruit. *J. Sci. Food Agric.* 86 (14), 2458–2462. doi: 10.1002/jsfa.2640

Zheng, H. B., Cheng, T., Li, D., Zhou, X., Yao, X., Tian, Y. C., et al. (2018). Evaluation of RGB, color-infrared and multispectral images acquired from unmanned aerial systems for the estimation of nitrogen accumulation in rice. *Remote Sens.* 10 (6), 824. doi: 10.3390/rs10060824

Zheng, H., Zhou, X., He, J., Yao, X., Cheng, T., Zhu, Y., et al. (2020). Early season detection of rice plants using RGB, NIR-G-B and multispectral images from unmanned aerial vehicle (UAV). *Comput. Electron. Agric.* 169, 105223. doi: 10.1016/j.compag.2020.105223

Zhou, Y., Lao, C., Yang, Y., Zhang, Z., Chen, H., Chen, Y., et al. (2021). Diagnosis of winter-wheat water stress based on UAV-borne multispectral image texture and vegetation indices. *Agric. Water Manage.* 256, 107076. doi: 10.1016/j.agwat.2021.107076

Zhu, W. X., Sun, Z. G., Huang, Y. H., Yang, T., Li, J., Zhu, K. Y., et al. (2021). Optimization of multi-source UAV RS agro-monitoring schemes designed for field-scale crop phenotyping. *Precis. Agric.* 22 (6), 1768–1802. doi: 10.1007/s11119-021-09811-0

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 January 2024
DOI 10.3389/fpls.2023.1277672

Check for updates

OPEN ACCESS

EDITED BY
Zhenjiang Zhou,
Zhejiang University, China

REVIEWED BY
Haikuan Feng,
Beijing Research Center for Information Technology in Agriculture, China
Simon Pearson,
University of Lincoln, United Kingdom

*CORRESPONDENCE
Geng Bai
gbai2@unl.edu

RECEIVED 15 August 2023
ACCEPTED 12 December 2023
PUBLISHED 08 January 2024

CITATION
Bai G, Koehler-Cole K, Scoby D, Thapa VR, Basche A and Ge Y (2024) Enhancing estimation of cover crop biomass using field-based high-throughput phenotyping and machine learning models.
Front. Plant Sci. 14:1277672.
doi: 10.3389/fpls.2023.1277672

COPYRIGHT
© 2024 Bai, Koehler-Cole, Scoby, Thapa, Basche and Ge. 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 estimation of cover crop biomass using field-based high-throughput phenotyping and machine learning models

Geng Bai[1]*, Katja Koehler-Cole[2], David Scoby[3], Vesh R. Thapa[3], Andrea Basche[3] and Yufeng Ge[1,4]

[1]Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln, NE, United States, [2]Nebraska Extension, University of Nebraska-Lincoln, Ithaca, NE, United States, [3]Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, United States, [4]Center for Plant Science Innovation, University of Nebraska-Lincoln, Lincoln, NE, United States

Incorporating cover crops into cropping systems offers numerous potential benefits, including the reduction of soil erosion, suppression of weeds, decreased nitrogen requirements for subsequent crops, and increased carbon sequestration. The aboveground biomass (AGB) of cover crops strongly influences their performance in delivering these benefits. Despite the significance of AGB, a comprehensive field-based high-throughput phenotyping study to quantify AGB of multiple cover crops in the U.S. Midwest has not been found. This study presents a two-year field experiment carried out in Eastern Nebraska, USA, to estimate AGB of five different cover crop species [canola (*Brassica napus* L.), rye (*Secale cereale* L.), triticale (*Triticale* × *Triticosecale* L.), vetch (*Vicia sativa* L.), and wheat (*Triticum aestivum* L.)] using high-throughput phenotyping and Machine Learning (ML) models. Destructive AGB sampling was performed three times during each spring season in 2022 and 2023. An array of morphological, spectral, thermal, and environmental features from the sensors were utilized as feature inputs of ML models. Moderately strong linear correlations between AGB and the selected features were observed. Four ML models, namely Random Forests Regression (RFR), Support Vector Regression (SVR), Partial Least Squares Regression (PLSR), and Artificial Neural Network (ANN), were investigated. Among the four models, PLSR achieved the highest Coefficient of Determination (R^2) of 0.84 and the lowest Root Mean Squared Error (RMSE) of 892 kg/ha (Normalized RMSE (NRMSE) = 8.87%), indicating that PLSR could be the most appropriate method for estimating AGB of multiple cover crop species. Feature importance analysis ranked spectral features like Normalized Difference Red Edge (NDRE), Solar-induced Fluorescence (SIF), Spectral Reflectance at 485 nm (R485), and Normalized Difference Vegetation Index (NDVI) as top model features using PLSR. When utilizing fewer feature inputs, ANN exhibited better prediction performance compared to other models. Using morphological and spectral parameters as input features alone led to a R^2 of 0.80 and 0.77 for AGB prediction using ANN, respectively. This study demonstrated the

feasibility of high-throughput phenotyping and ML techniques for accurately estimating AGB of multiple cover crop species. Further enhancement of model performance could be achieved through additional destructive sampling conducted across multiple locations and years.

KEYWORDS

aboveground biomass, cover crop, plant phenotyping, machine learning, partial least squares regression, rye

1 Introduction

Cover crops, strategically incorporated into farmland soil during fallow periods between primary cash crop production, serve a multitude of potential benefits. These benefits include the reduction of soil erosion through wind and surface runoff, augmentation of carbon sequestration through biomass accumulation, and the contribution of nitrogen to the soil for subsequent growing seasons (Kaye and Quemada, 2017). In the global context of regenerative and conservation agriculture, cover crops have been identified as a key component for ensuring long-term sustainability and ecological resilience of agroecosystems (Marandola et al., 2019; Aiyer et al., 2022; Inveninato Carmona et al., 2022). While cover crops can advance sustainability goals in crop production, their improper implementation could lead to a reduction in the yield of the primary crop. A large-scale remote sensing study on the yield reduction introduced by cover cropping showed an average yield loss of 5.5% and 3.5% for corn (Zea mays L.) and soybean (Glycine max L.) fields in the U.S. Corn Belt (Deines et al., 2023). Terminating rye cover crop close to the planting date of cash crop was found to increase the seedling root disease, which led to reduced yield for corn production (Acharya et al., 2022). Planting ruzigrass cover crop was found to reduce the soil phosphorus availability (Almeida et al., 2019). Continuing to advance management practices as "systems" that optimize benefits from cover crops while minimizing cash crop yield penalty is a critical element of their success and extended use (Basche and Roesch-McNally, 2017; Koehler-Cole et al., 2023; Popovici et al., 2023).

The accumulation of above ground biomass (AGB) from cover crops serves as the fundamental driving force behind these potential benefits as well as potential negative impacts on cash crop yield (Ruis et al., 2019; Nichols et al., 2020). Consequently, quantitatively monitoring this process for different cover crops becomes an essential step in assessing their performance. Furthermore, such monitoring could empower farmers with digital tools to support their decision-making for important applications, such as establishing data-driven methodologies to determine optimal termination dates of cover crops, and understanding nutrient release and recycling to cash crops. The biomass accumulation process, however, is complex due to the genotype by environment (GxE) interactions created by differing geographical locations, fluctuating climatic conditions year after year, and variability in farmers' practices (Koehler-Cole et al., 2023). As such, intensive research is ongoing to determine the optimal types and cultivars of cover crops, the best termination stage, and the suitable termination method for local farmers (Adetunji et al., 2021). Conventionally, cover crop AGB has been measured through destructive sampling of a specific subarea in the plot. This sampling process is time-consuming and labor-intensive, and poses a challenge in experiments constrained by plot size due to the need to balance treatment levels and field size (O'Brien et al., 2022; Zhang et al., 2022; Capri et al., 2023). Nevertheless, understanding the dynamic nature of cover crop biomass accumulation at high temporal resolution could offer insights for researchers seeking to develop or refine predictive models for cover crop research. However, the use of destructive sampling methods often precludes the possibility of continuous quantification of plant materials over the entire experimental period.

Field-based High-Throughput Plant Phenotyping (FHTPP) systems have made significant progress in recent years, owing to the proliferation of low-cost sensors, ground vehicles, unmanned aerial vehicles (UAV), and high-resolution satellite Constellations (Atefi et al., 2020; Zhang et al., 2020; Feng et al., 2021). Some of these systems have facilitated data acquisition at sub-leaf resolution by measuring plants from a close range (Bai et al., 2018; Bai et al., 2019; Yuan et al., 2019). Various Machine Learning (ML) models leverage morphological and spectral traits extracted from FHTPP systems to quantify important agronomic traits, such as canopy height, flowering date, maturity date, and AGB, among others (Zhou and Nguyen, 2021). While substantial advancements have been made in the development of data collection and processing pipelines in FHTPP, there is a paucity of publications documenting its application for quantifying cover crop AGB. Given the growing popularity and potential benefits of cover crops, FHTPP emerges as a promising tool, capable of quantifying the AGB accumulation process in a non-destructive way with notable efficiency. Multispectral imaging has been employed to establish a correlation with the termination efficiency rating for various cover crop species (Kumar et al., 2023). High-resolution satellite images were used to estimate cover crop biomass, achieving the highest correlation coefficient (R) of 0.74 (Kharel et al., 2023). It is worth noting that some crop species, such as winter wheat, can be utilized as both cash and cover crops, hence a relatively larger volume of

literature is available for them. Multispectral images, gathered by UAVs, have been widely used to estimate the AGB of winter wheat with the aid of ML models (Wang et al., 2022). Similarly, Vegetation Index–based ML models have been developed and compared for the estimation of AGB of potato (Solanum tuberosum L.) canopy (Liu et al., 2022a; Liu et al., 2022b). Both traditional ML and deep learning models have been explored for AGB estimation, the choice of which depends on the size of the available data (Yue et al., 2018; Dong et al., 2020; Whitmire et al., 2021).

We undertook a two-year experiment involving five cover crop species: canola, rye, triticale, vetch, and wheat. The objectives of this study were: 1) to employ a FHTPP system to collect high-resolution phenotypic datasets for the cover crops, and 2) to build and evaluate ML models to quantify AGB of cover crops. This research aims to fill an important gap in the application of FHTPP in cover crop research, potentially leading to new tools for cover crop breeding and field management.

2 Materials and methods

2.1 Experimental design and ground truth data collection

Table 1 provides details of a two-year cover crop experiment conducted on a research farm near Mead, Nebraska, United States (41°08'44" N, 96°26'20" W, and an elevation of 350m). The study spanned the growing seasons of 2022 and 2023, involving the cultivation of five cover crop species, with no irrigation or nitrogen application. The dominant soil types are Filbert silt loam and Yutan silt clay loam. Figure 1 illustrates the field conditions where data collection took place, the experimental designs employed, and the cameras/sensors used in the study. In both years, five cover crop species were planted, with each year having 3 and 4 reps, respectively. In 2023, the experiment included two different rye varieties. This gives a total of 15 plots in the first year, and 24 plots in the second year. The dimensions of the plots were 4.6 m by 6.1 m in 2022 and 1.9 m by 6.1 m in 2023. At each plot, a designated area of 0.5 m^2 was marked using flags. Cover crops and any additional green vegetation within these areas were hand-clipped from the soil surface (Supplementary Figure S1). The harvested fresh biomass was oven-dried at 65°C until a constant weight was achieved for dry AGB determination.

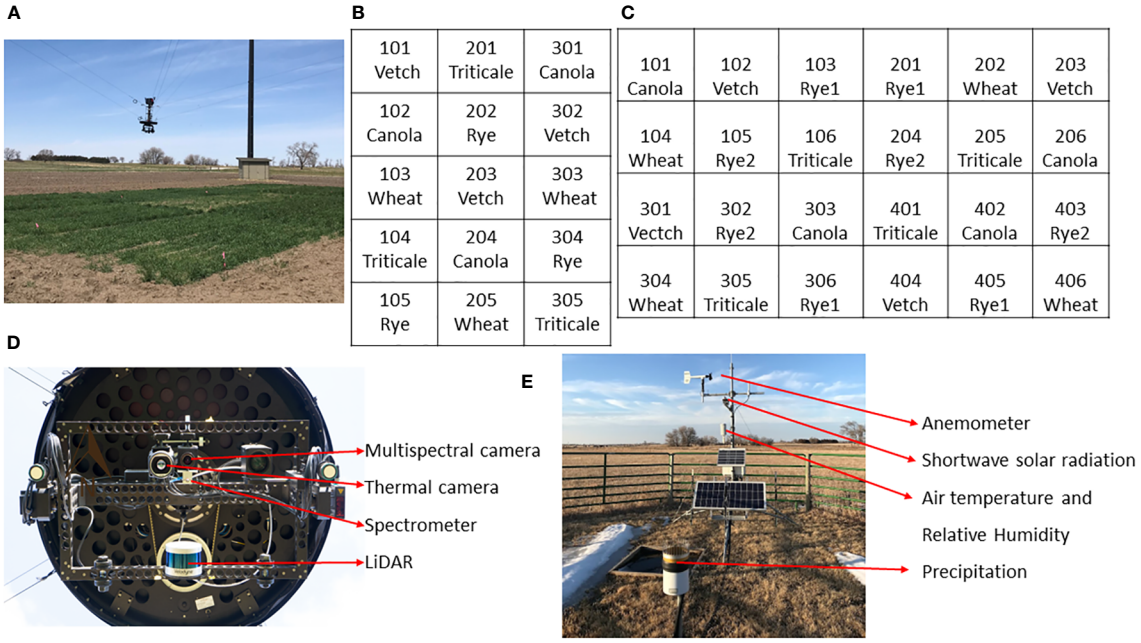
TABLE 1 Information for the two-year cover crop field experiment.

Year	Parameter	Information
2022	Cover crop species and cultivars	Vetch: Hairy vetch Canola: Trophy canola Wheat: Winter wheat Triticale: Triticale Rye: Yankee rye
	Planting date	Sep-26-2021
	Biomass sampling dates	Apr-18-2022 Apr-26-2022 May-09-2022
2023	Cover crop species and cultivars	Vetch: Hairy vetch Canola: Trophy canola Wheat: Winter wheat Triticale: Triticale Rye: Elbon and Yankee rye
	Planting date	Sep-27-2022
	Biomass sampling dates	Apr-07-2023 Apr-27-2023 May-08-2023

All crop species survived the winter of the first year, although there was a notable reduction in canola AGB in the spring. In contrast, triticale, vetch, and canola were winterkilled in the second year, attributable to a uniquely dry and cold regional winter. Destructive AGB sampling was carried out three times each year in the spring, resulting in a total of 45 and 72 AGB samples for the respective years (including winterkilled plots with a yield of 0 kg/ha in 2023). These samples include 36 plots with zero AGB due to winter kill in the spring of 2023. The plot-scale HTPP data, including multispectral images, LiDAR (Light Detection and Ranging) point clouds, thermal images, and spectral reflectance in the visible and near-infrared range, were collected using a cable-suspended field phenotyping system. Additionally, an on-site weather station recorded standard weather data at a 1-minute interval, ensuring an accurate integration of the phenotypic and environmental data.

2.2 Data processing before machine learning

Figure 2 outlines the data preprocessing pipeline employed for feature extraction in the context of ML modeling. Phenotypic data, anticipated to correlate with plant AGB, were derived from raw sensor data following a previously developed data processing protocol (Bai et al., 2019). Environmental data, synchronized with the phenotypic data via timestamps, were also incorporated into the analysis. The phenotypic and environmental features are listed in Figure 2. A set of raw images were also presented alongside their corresponding segmentation results. For each dataset, nadir-captured multispectral and thermal images were obtained using the sensing platform. Green vegetation pixels were identified as foreground through image registration, histogram equalization, and thresholding. Green Pixel Fraction (GPF), plot average temperature (Tp), canopy average temperature (Tc), and soil average temperature (Ts) were computed based on the segmentation outcomes. GPF demonstrated a strong correlation with Normalized Difference Vegetation Index (NDVI) throughout the growing season, suggesting its potential as a robust estimator for quantifying canopy growth before canopy closure (Bai et al., 2016). Tp contributed to the estimation of the Leaf Area Index and was used to detect crop water stress (Irmak et al., 2000; Wang et al., 2021; Cheng et al., 2023). Additionally, Growing Degree Day (GDD), a widely adopted feature in crop growth models, was calculated as an additional environmental feature using weather data from the first day of the year. Two base temperatures were applied for the GDD calculation (Winter wheat: 0°C (Slafer and Savin, 1991); Canola: 5°C (Lawson et al., 2006); Rye: 0°C (Szuleta et al., 2022); Vetch: 5°C (Lawson et al., 2012); Triticale: 0°C (Schwarte et al., 2006). Cover



101 Vetch	201 Triticale	301 Canola
102 Canola	202 Rye	302 Vetch
103 Wheat	203 Vetch	303 Wheat
104 Triticale	204 Canola	304 Rye
105 Rye	205 Wheat	305 Triticale

101 Canola	102 Vetch	103 Rye1	201 Rye1	202 Wheat	203 Vetch
104 Wheat	105 Rye2	106 Triticale	204 Rye2	205 Triticale	206 Canola
301 Vectch	302 Rye2	303 Canola	401 Triticale	402 Canola	403 Rye2
304 Wheat	305 Triticale	306 Rye1	404 Vetch	405 Rye1	406 Wheat

FIGURE 1
Field condition, experimental layout, and instrumentation of the experiment. **(A)** Field photo of the data collection; **(B, C)** Experimental layout in 2022 and 2023 seasons, respectively; **(D)** Sensors onboard the NU-Spidercam platform; **(E)** Instrumentation of the on-site weather station.

crops sharing the same base temperatures had identical GDD values (Supplementary Equation S1). A comprehensive list of equations used for calculating Vegetation Indices (VIs) is provided in the Supplementary File (Supplementary Table S1), sourced from prior publications (Rouse et al., 1974; Rondeaux et al., 1996; Gamon et al., 1997; Barnes et al., 2000; Suárez et al., 2008; Meroni et al., 2009; Badgley et al., 2017).

2.3 Machine learning models for aboveground biomass estimation

Feature reduction was carried out to mitigate strong multicollinearity among features, as well as to exclude features that showed a weak correlation with AGB ($|R| < 0.45$). Subsequently, one-hot encoding was applied to convert the crop

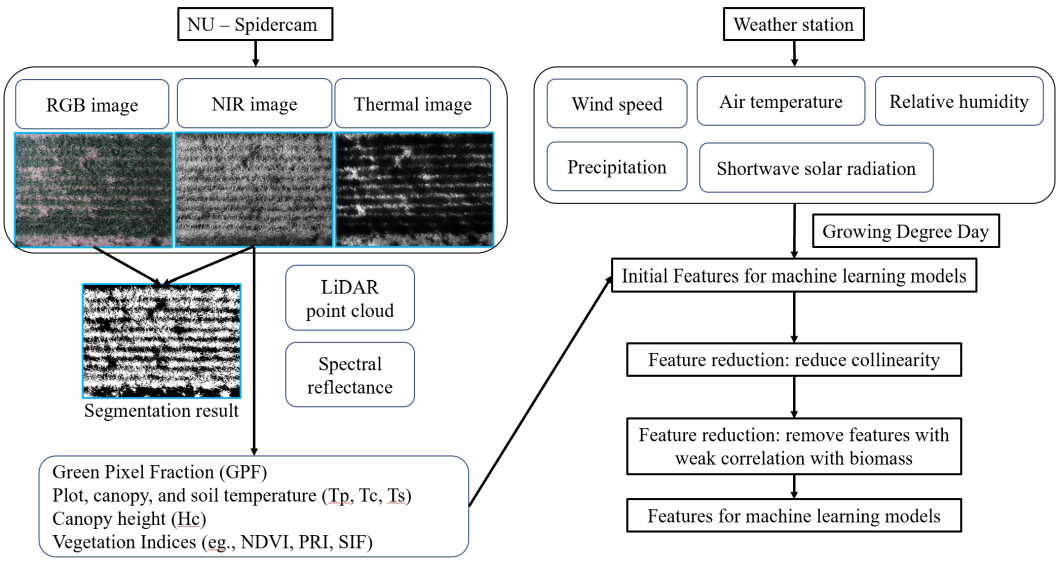


FIGURE 2
Data processing pipeline for phenotypic and environmental features. The left panel visualizes the data processing of NU-Spidercam data, including raw images and a list of phenotypic parameters; the right panel lists the raw environmental data collected from the on-site weather station and the feature reduction process before the development of machine learning models.

species into a numerical feature. Model performance was investigated using all features first and using different categories of features. The input features for model training were categorized into phenotypic and environmental features. Phenotypic features can quantify the variance among individual plots due to growth heterogeneity caused by differences in soil, water, and nutrients. Phenotypic features were further grouped into morphological (GPF and Canopy Height (Hc)), spectral (VIs), and thermal (Tp) groups, which were captured using different onboard instruments (Figures 1, 2).

Given the small data size, we used non-deep learning models to avoid the risk of overfitting to a certain degree. These models include Random Forests Regression (RFR), Support Vector Regression (SVR), Partial Least Squares Regression (PLSR), and Artificial Neural Network (ANN). RFR is an ensemble learning method that constructs multiple decision trees during training and outputs the mean prediction of the individual trees for AGB modeling. This method not only offers robustness against overfitting but also provides an inherent feature importance evaluation, enabling an understanding of which predictors are most influential in the modeling process. SVR is a regression adaptation of support vector machines, which operates by identifying an optimal hyperplane that functions as a decision boundary. It strives to ensure that most data points are within a certain margin of this hyperplane, maximizing the margin while limiting the regression errors. This method is effective in providing robust predictions, particularly for datasets with noisy observations. PLSR is a modeling method specializing in handling highly correlated input features. It works by transforming the original input features into a smaller set of uncorrelated components, capturing the most variance in the dataset. These components are formed from linear combinations of the original variables and are used in building the model to efficiently deal with multicollinearity and reducing dimensionality. ANNs are powerful computational models that excel at capturing complex non-linear relationships. They function by adjusting weights and biases during training to minimize a specific loss function and improve the prediction accuracy. This process involves multiple training cycles of forward and backward propagation. By iteratively refining these parameters, ANNs learn high-level patterns and correlations that are often missed by more traditional statistical methods. During the development of the ANN model, only two hidden layers were used, both incorporating potential L2 regularizations to reduce overfitting. The primary parameters explored during hyperparameter tuning are detailed in Supplementary Table S2. The dataset was randomly split into training (80%) and test (20%) sets. Data standardization was carried out for each feature independently. During model training, ten-fold cross-validation was implemented across all models. The optimized models with the best hyperparameters were selected based on the highest R^2 observed during cross validation. The chosen optimal model was then implemented for predicting AGB in the test set, and the predictive performance was assessed using R^2, Root Mean Squared Error (RMSE), and Normalized RMSE (NRMSE). Scikit-learn, a Python library, was used for training and evaluating ML models. Mean decrease impurity, coefficient magnitudes, and PLS loadings were employed to evaluate feature importance in RFR, SVR, and PLSR models, respectively.

3 Results and discussion

3.1 Aboveground biomass and data pre-processing

Descriptive statistics of AGB for each crop type in the whole, training, and testing data sets are illustrated in the Supplementary Document (Supplementary Figure S2). Figure 3 shows the temporal variations of AGB and selected sensor-based phenotypic features (Hc, NDVI1, and GPF) on the three sampling dates in both years. All parameters exhibited a consistent increase over time, aligning with the observed AGB growth, thus suggesting their potential as features to estimate cover crop AGB. In both years, rye consistently displayed the highest biomass, surpassing other cover crops significantly (Figures 3A, E). Notably, rye stands out as the most widely grown cover crop species in the U.S. Midwest (Nichols et al., 2020). The capacity to produce exceptional AGB under the local climate has prompted the wide adoption of rye cover crop (Ruis et al., 2019; Koehler-Cole et al., 2020). In addition, rye's capability to significantly suppress weed growth leads to reduced management costs (Oliveira et al., 2019; Rosa et al., 2021). A field study found that a minimum AGB of 5000 kg/ha was required to achieve a 75% reduction in weed biomass (Nichols et al., 2020). Based on the biomass data obtained in this study, it is evident that rye was the only species consistently achieving this threshold in both years, typically by the end of April or the beginning of May, given a planting date at the end of September. In 2022, wheat and triticale outperformed vetch and canola in terms of biomass production. A substantial portion of the canola perished in the spring of 2022, resulting in the lowest biomass yield.

Rye consistently demonstrated higher Hc values than other cover crops, and all species exhibited increasing Hc values throughout the experimental period (Figure 3B). At the first and second sampling dates of 2022, rye had the highest NDVI1 and GPF values (Figures 3C, D). Discrepancies in VIs among the cover crop species diminished by the second date, ultimately reaching the same level by the third sampling date in 2022. By the final sampling day of 2022, NDVI1 and GPF for different crop species were not significantly different. However, these two parameters for vetch and canola plots experienced a faster increase than those of others during the investigated period. In 2023, triticale, vetch, and canola experienced winterkill, potentially attributable to diminished snow cover during the coldest periods. A few triticale plants survived in 2023 (Supplementary Figure S3), resulting in slightly higher NDVI1 and GFP than vetch and canola, which were essentially bare soil (Figures 3G, H). This underscores a potential risk associated with growing less winter-hardy crops, such as legumes and brassicas, in the study region. Between rye and wheat, all assessed phenotypic parameters showed higher values for rye than wheat in 2023 (Figures 3F–H). The development of additional features can be found in the Supplementary Document (Supplementary Figures S4, S5).

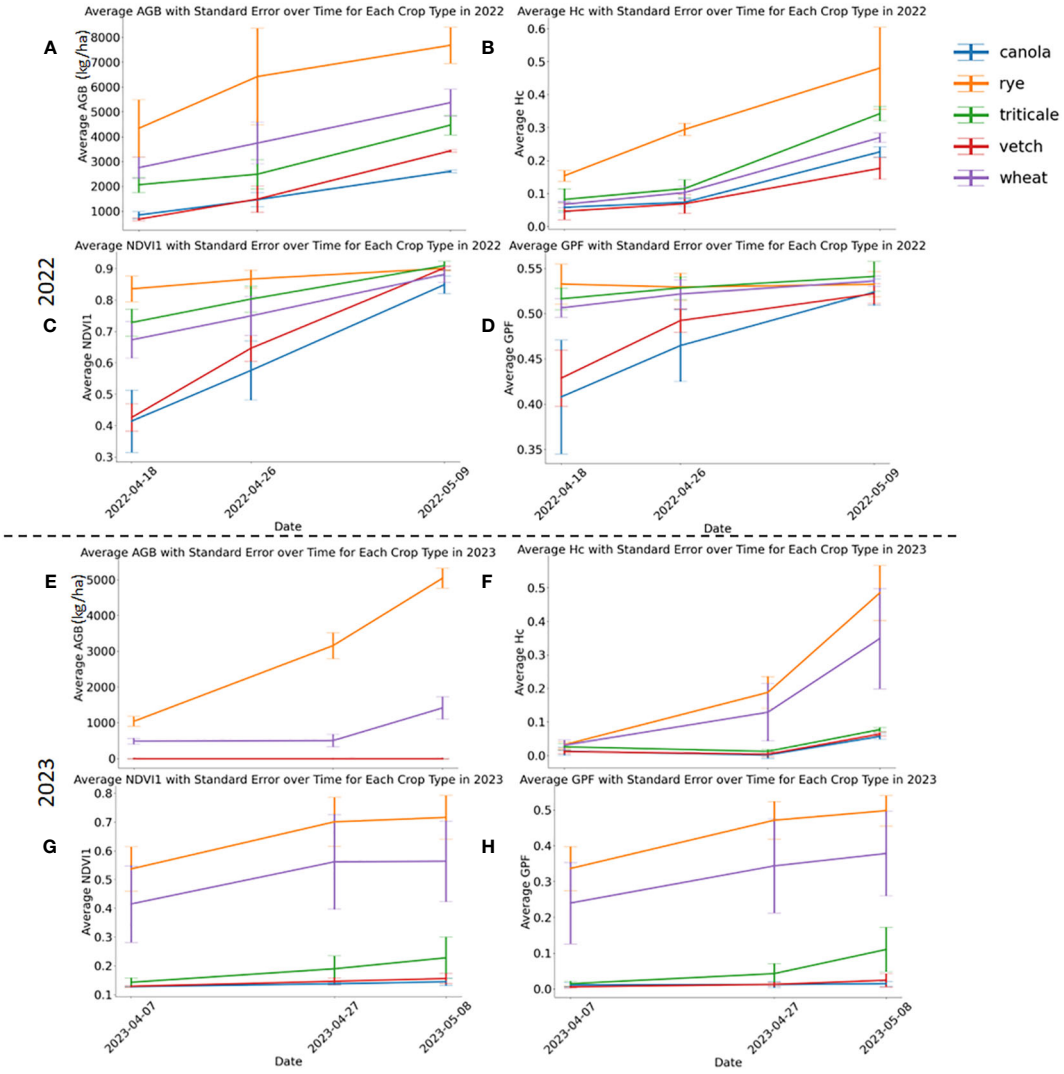


FIGURE 3
Aboveground biomass and selected phenotypic parameters in 2022 **(A–D)** and 2023 **(E–H)**. Three phenotypic parameters include Hc, NDVI1, and GPF. Standard errors are marked with vertical bars.

3.2 Correlation analysis

The correlation matrix of all features and AGB before the feature reduction for all crop species is shown in Figure 4A. We categorized these features into four groups for further analysis: morphological (GPF and Hc), spectral (all VIs), thermal (Tp, Tc, and Ts), and environmental (all features from the on-site weather station). The first feature reduction aimed to eliminate features with strong collinearity while retaining at least one feature from each category. Tp, the average temperature of all canopy and vegetation pixels, was retained as the representative thermal feature for Tc and Ts due to their strong positive correlations. Further investigation revealed a notable temperature difference in mid April between Tc and Ts, as opposed to early May (Supplementary Figure S6). However, Tp consistently showed very strong correlations with Tc and Ts. Another rationale for retaining Tp as the thermal feature, rather than Tc or Ts, was its relatively easier measurability without a need to distinguish between canopy and soil pixels. A previous study showed that thermal parameters improve ML-based prediction of the Leaf Area Index for maize (*Zea mays* L.) (Wang et al., 2021). Therefore, we expected that integrating Tp would improve the model's performance in estimating AGB. R485 (Blue) and R675 (Red) represent the strong absorption bands by the crop canopy, while R550 (Green) denotes the most reflected band in the visible range. Due to their strong positive correlations, reflectance in the blue wavelength (R485) was kept to represent spectral reflectance in the visible wavelength range. All NDVIs, Normalized Difference Red Edge (NDRE), and Optimized Soil Adjusted Vegetation Index (OSAVI) exhibited strong positive correlations due to the similar wavelengths used in their calculations. Thus, NDVI1 and NDRE were retained for NDVI-related VIs. NDRE was also kept due to its lesser susceptibility to saturation, as evidenced by its behavior at the final sampling date (Supplementary Figures S4, S5). Among the ten Solar-induced Fluorescence (SIF) features, only SIF1, SIF8, and SIF10 were kept, as the remaining SIFs had strong positive

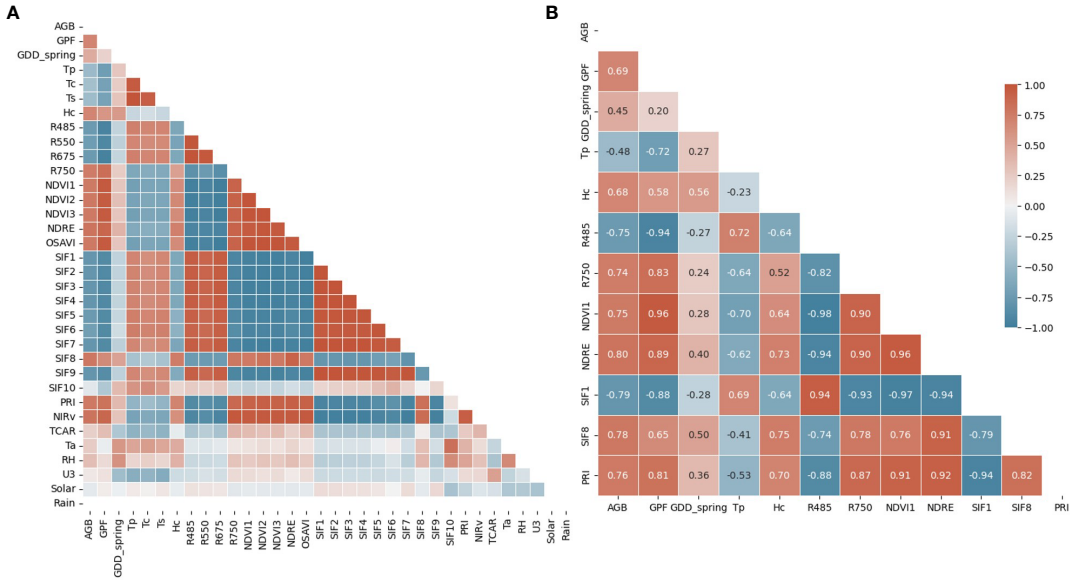


FIGURE 4
Correlation matrix of all **(A)** and reduced features **(B)**, including morphological (GPF and Hc), spectral (VIs), thermal (Tp, Tc, and Ts), and environmental features (GDD_spring, Ta, RH, U3, Solar, and Rain). The correlation coefficient is shown in the correlation matrix of reduced features.

correlations with SIF1. These strong correlations could be attributed to similarities in their calculation equations. Notably, SIF8 uses reflectance values from two NIR bands, which is distinct from the equations of other SIFs. While most SIF features leverage reflectance near an oxygen absorption band (O2-B) around 687 nm, SIF10 utilizes reflectance near another oxygen absorption band (O2-A) around 760 nm. The results of the first feature reduction are available in the Supplementary Document (Supplementary Figure S7).

The second feature reduction involved the removal of features with a weak correlation to AGB ($|R| < 0.45$). In this step, all environmental features were filtered out except for GDD from the beginning of the year (GDD_spring). This indicates that instantaneous environmental features exerted a weaker influence on AGB than phenotypic parameters in this study. The only accumulative parameter, GDD_spring, demonstrated relatively stronger predictive power for AGB (R = 0.45). SIF10 and Transformed Chlorophyll Absorption in Reflectance (TCAR) were also excluded due to their weak correlation with AGB. When considering crop species as a single feature, the feature reduction process resulted in the selection of 12 features for ML model training. Figure 4B illustrates the correlation matrix of the final features, including an environmental feature (GDD_spring), two morphological features (GPF and Hc), seven spectral features (seven VIs), and a thermal feature (Tp). These features served as inputs for ML models, although strong correlations persisted among some of them.

Figure 5 presents the sorted R values between feature inputs and AGB for all crop species (Figure 5A), as well as for each specific crop species (Figures 5B–F). In general, the top five features with stronger correlations were VIs calculated from spectral reflectance. Specifically, NDRE (R = 0.8), SIF1 (R = -0.79), SIF8 (R = 0.78), PRI (R = 0.76), and NDVI1 (R = 0.75) exhibited the highest correlations with AGB when considering all crop species. For individual crop species, additional features were included in the top five features, including R485, R750, GDD_spring, and Hc. A weighted scoring system was used to rank the overall top five predictors, assigning decreasing scores (from 5 to 1) based on the features' ranks (1 to 5) for each crop species. The results reaffirmed that VIs exhibit relatively stronger linear correlations with AGB, with a slightly varied ranking compared to Figure 5A, where the top five predictors were SIF1, PRI, NDRE, NDVI1, and R485.

3.3 Machine learning model performance

Figure 6 illustrates the model performance using all feature inputs across different models. The optimal hyperparameters for each model are listed in Supplementary Table S3. R^2 and RMSE values are also presented in the corresponding subfigures. The PLSR model (R^2 = 0.84, RMSE = 892 kg/ha, NRMSE = 9%) exhibited the best performance when compared with models that could better capture non-linear interrelationships between the features and AGB. The RFR model (R^2 = 0.69, RMSE = 1246 kg/ha, NRMSE = 12%) showed the poorest performance, while SVR (R^2 = 0.78, RMSE = 1037 kg/ha, NRMSE = 10%) and ANN (R^2 = 0.81, RMSE = 981 kg/ha, NRMSE = 10%) showed comparable performance with each other. These results indicate that AGB can be estimated using linear ML models with good performance. The correlation analysis above also confirmed strong linear correlations between features and AGB with feature collinearity at a certain degree (Figures 4, 5), indicating that PLSR could be the most suitable model for this scenario. Previous work concluded that PLSR showed promising performance in AGB estimation for winter wheat at different growth stages (Wang et al., 2022). Thus, our work further establishes the promising performance of PLSR in AGB

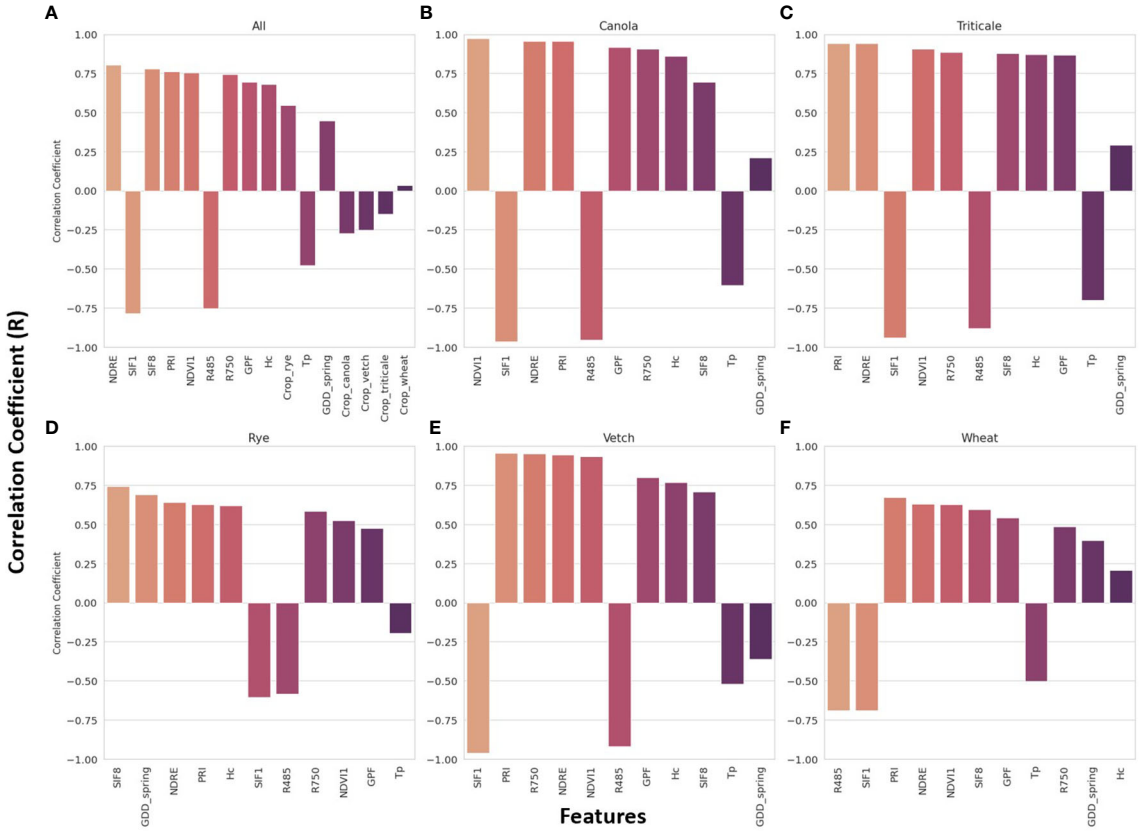


FIGURE 5
Pearson's correlation coefficient (R) between all features extracted from the images and sensor data and the aboveground biomass for all crop species **(A)** and each specific crop species **(B–F)**.

estimation for multiple cover crop species. Another study explored AGB estimation for multiple oats (*Avena sativa* L.) cultivars at different locations using RFR, SVR, PLSR, and ANN models. The results showed that no single model performed best across all locations, and PLSR outperformed other models at certain experimental locations (Sharma et al., 2022). Furthermore, a systematic investigation of model performance involving eight ML models concluded that PLSR was among the best models for ABG estimation in winter wheat (Yue et al., 2018). These studies reconfirmed that linear-based models such as PLSR can outperform more complex, non-linear models in AGB estimation. To the best of our knowledge, no prior results are available for comparison with our model performance across multiple cover crop species. Therefore, we exclusively compared the model performance for available crop species, such as winter wheat. Comparable model performance was observed when using drone-based multispectral camera systems and PLSR to estimate winter wheat AGB [R^2 = 0.75, (Wang et al., 2021)]. A higher model performance was achieved when using VIs from a close-range field spectrometer and PLSR for the same purpose [R^2 = 0.89, (Yue et al., 2018)]. Additionally, comparable performance was achieved for potato AGB estimation using RFR at the tuber-growth stage [R^2 = 0.68, (Liu et al., 2022a)].

Figure 7 shows the feature importance of RFR, SVR, and PLSR models. Morphological features (Hc and GPF) ranked among the top five important features of the RFR model. In SVR, Hc and GPF ranked 9th and 6th, while in PLSR, they ranked 16th and 6th. SVR and PLSR shared the top five features from VIs with different ranking orders, including NDRE, NDVI1, R485, SIF1, and SIF8. Among these five VIs, SIF1, R485, and SIF8 were also among the top five important features of RFR. The environmental feature, GDD_spring, ranked 7th, 8th, and 7th for RFR, SVR, and PLSR models, respectively, indicating its significant contribution to model performance. The thermal feature, Tp, ranked 12th, 10th, and 9th in the three models, respectively, suggesting a slightly lesser contribution to model performance in this study. Generally, the crop species feature held lower importance than other features, with average rankings at 13th, 13th, and 12th for the three models, respectively.

Table 2 shows the model performance using different feature categories. Firstly, only morphological features were used in the model development. Except for the ANN model (R^2 = 0.80, RMSE = 986 kg/ha, NRMSE = 9.81%), other models yielded the lowest R^2 (0.34 - 0.50) with the highest RMSE and NRMSE (1573 - 1802 kg/ha and 15.64% - 17.92%) compared to other conditions. ANN appears to be the only method capable of estimating AGB for multiple cover crop species with GPF and Hc alone. When using only spectral parameters as feature inputs, ANN still demonstrated the highest R^2 (0.77), although with a slightly reduced margin compared to other methods (RFR: 0.66; SVR: 0.62; PLSR: 0.68). The relatively strong performance of using spectral features alone ($R^2 \geq 0.62$) confirms

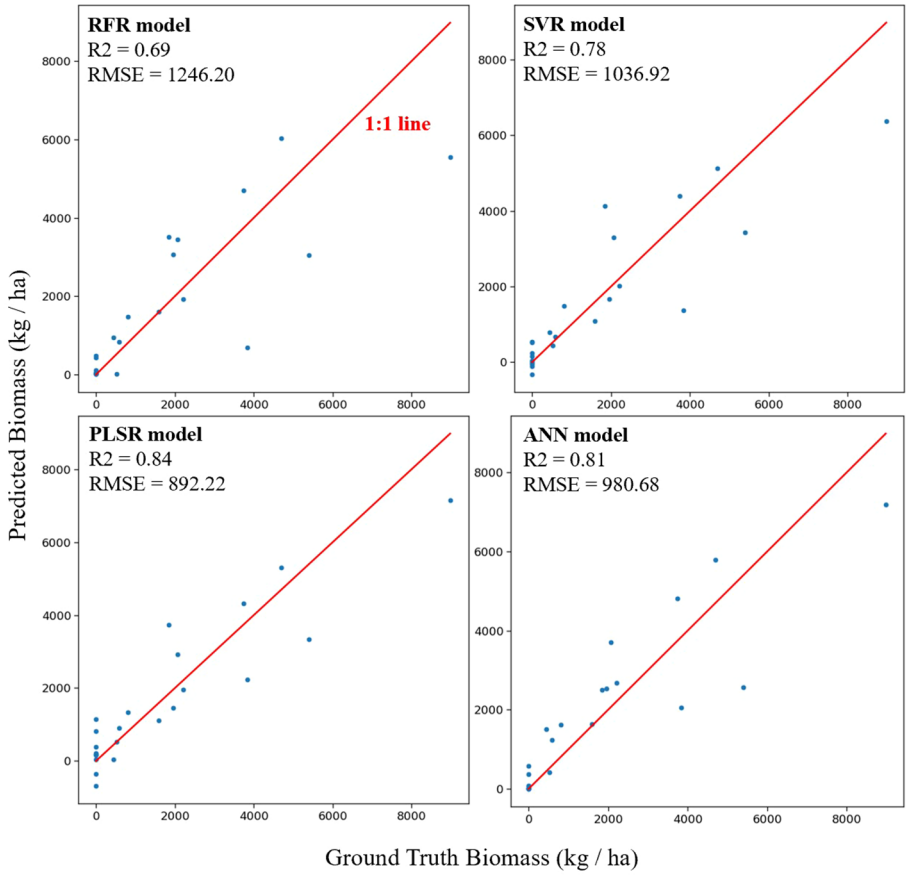


FIGURE 6
Scatter plots of predicted vs. actual aboveground biomass using all features extracted from the cameras and sensors by four machine learning techniques. Model names and performance metrics (R^2 and RMSE) are shown in the corresponding subplots.

the application of aerial sensing platforms with a single multispectral camera (e.g., drones and high-resolution satellite imagery) for cover crop AGB estimation. The difference in model performance between the models utilizing spectral features and those using both morphological and spectral features was small ($\Delta R^2 = \pm 0.2$). This result indicates that GPF and Hc did not improve the model performance when spectral features were already utilized. A possible explanation is that the spectral features already embedded the useful information of GPF and Hc. Upon excluding the crop species feature, model performance declined except for the ANN model. The feature importance of crop species for RFR, SVR, and PLSR models also explains to a certain degree why adding this feature improves the performance of the overall model for all five crop species. The impact of the thermal feature, Tp, on model performance was also investigated by estimating model performance without Tp. The results indicated that Tp contributed the least to the improvement of R^2 and RMSE for RFR, SVR, and PLSR models. Notably, including Tp led to a higher RMSE for ANN. Therefore, the instantaneous thermal features (e.g., Tp) may offer limited improvement in this application. Overall, we concluded that accurate AGB estimation for multiple cover crop species can be achieved using drone systems or high-resolution satellite constellations, which provide spectral and morphological features at the plot level. Also, more accurate prediction models can be generated from drone systems due to their higher spatial resolution of images, while satellite systems offer benefits in terms of higher throughput and coverage (Sankaran et al., 2015; Zhang et al., 2020; Kharel et al., 2023). Furthermore, accurate canopy height and thermal features could improve the model performance to a certain extent.

3.4 Model limitation

Based on Table 2, the PLSR model built on all features yields the highest R^2 value of 0.84, with an RMSE of 892 kg/ha. Given this RMSE value, the current model's accuracy may require further improvement to assist producers in predicting aboveground performance from advanced technology. Additionally, understanding the quantitative influence of the AGB of different cultivars of promising crops (e.g., various rye cultivars) on potential benefits is necessary. This effort will help establish clear target RMSE ranges of the models. The models developed in this study were trained and tested on multiple cover crop types using data from two years. However, the study was limited to a single site in Nebraska with specific soil and weather conditions. As the models are purely data-driven, we acknowledge that their performance

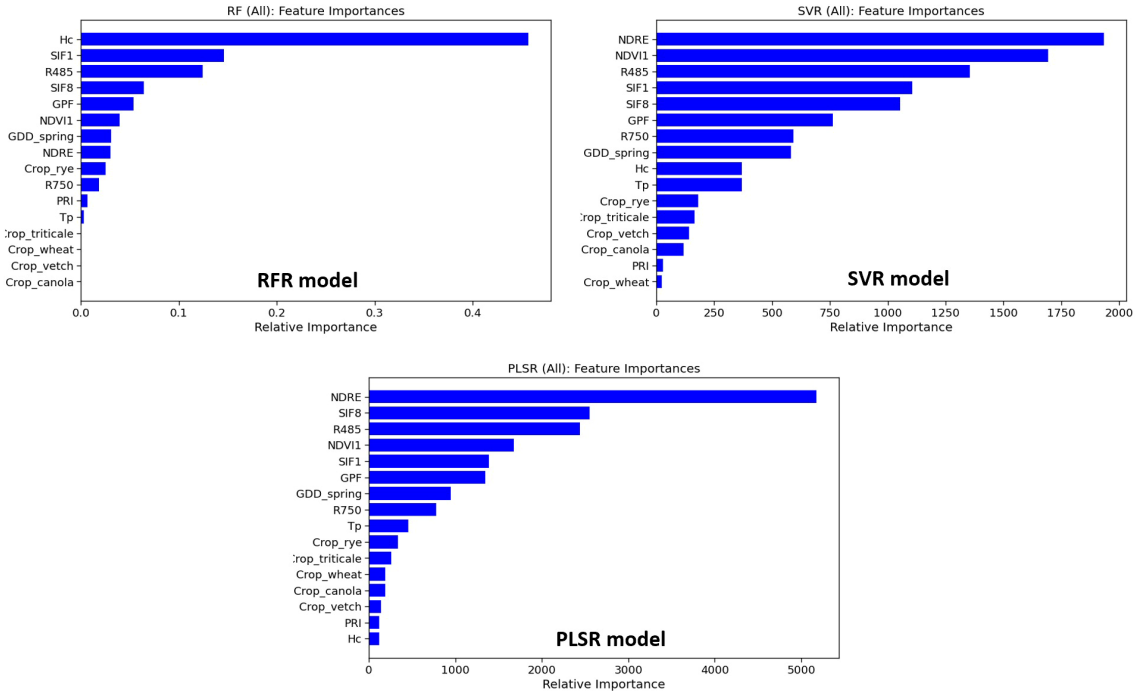


FIGURE 7
Feature importance for three investigated models, including RFR, SVR, and PLSR.

may not be robust when predicting cover crop AGB in environments beyond the investigated ranges, even within the U.S. Midwest and Great Plains. Further validation of the model's robustness and transferability is required by testing with truly independent data sets. For instance, a multi-location experiment across Nebraska, from west to east, would provide a much broader range of growing environments for candidate cover crops, aiding in testing and model improvement.

4 Conclusion

High-throughput plant phenotyping offers a non-destructive and efficient approach for estimating plant AGB. This study encompassed a two-year field experiment aiming at developing ML models for AGB estimation of five cover crop species - rye, vetch, canola, winter wheat, and triticale - from phenotypic and environmental data in the U.S.

TABLE 2 Comparison of model performance for cover crop aboveground biomass estimation by categorizing model inputs into morphological, spectral, and thermal groups. The units of RMSE and NRMSE are kg/ha and %, respectively.

Model	Metric	Feature Selection					
		Morphological (GPF, Hc)	Spectral	Morphological And Spectral	All but crop species	All but thermal (Tp)	All
RFR	R^2	0.34	0.66	0.67	0.65	0.68	0.69
	RMSE	1802	1287	1279	13201	1253	1246
	NRMSE	17.92	12.80	12.72	131.28	12.46	12.39
SVR	R^2	0.50	0.62	0.63	0.67	0.75	0.78
	RMSE	1573	1365	1355	1273	1109	1037
	NRMSE	15.64	13.57	13.48	12.66	11.03	10.31
PLSR	R^2	0.44	0.68	0.67	0.72	0.82	0.84
	RMSE	1664	1259	1281	1176	947	892
	NRMSE	16.55	12.52	12.74	11.69	9.42	8.87
ANN	R^2	0.80	0.77	0.79	0.84	0.81	0.81
	RMSE	986	1063	1008.05	895	958	981
	NRMSE	9.81	10.57	10.02	8.90	9.53	9.76

Midwest region. The raw dataset included multispectral imagery, LiDAR point clouds, spectral reflectance, thermal images, and environmental data, complemented by the AGB data obtained via destructive sampling. Consistent with prior research in similar climates, rye outperformed other species in terms of AGB accumulation in both years. Leveraging morphological, spectral, environmental, crop species, and thermal features extracted from the raw dataset, we employed four ML techniques - Random Forest Regression (RFR), Support Vector Regression (SVR), Partial Least Squares Regression (PLSR), and Artificial Neural Network (ANN) - to predict cover crop AGB. PLSR emerged as the best approach (R^2 = 0.84, RMSE = 892 kg/ha, NRMSE = 8.87%) when utilizing all feature inputs. Thus, this study highlights linear models, like PLSR, are on par with non-linear models in capturing the fundamental relationship between features and AGB for cover crop research, especially when dealing with feature collinearity. All feature categories contributed to the performance of RFR, SVR, and PLSR models, while spectral features alone exhibited the strongest performance for RFR, SVR, and PLSR models. Morphological features alone yielded satisfactory results when trained with the ANN model. Instantaneous thermal features made a marginal contribution to the model performance in this study. Despite achieving high testing accuracy in this study, we suggest that further training and validation of the models, using larger datasets and various data splitting techniques, could enhance model robustness. Consistent with challenges identified in previous research on AGB estimation for various crops, the scarcity of ground truth data continues to be a significant obstacle in developing more accurate and robust models, owing to the inherent constraints of field experiments and limited resources. Collecting more biomass data under different soil and climate conditions could further improve the model performance. Continuous development of more universal and robust image processing algorithms for vegetation segmentation, especially in later growing stages with dense vegetation canopy, could improve GPF quantification. Convolutional neural networks could be leveraged to replace manual feature extraction if a much larger biomass data set is available.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors.

Author contributions

GB: Data curation, Formal analysis, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. KK-C: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Writing – review & editing. DS: Conceptualization, Data curation, Formal analysis, Methodology, Writing – review & editing. VT: Visualization, Writing – review & editing. AB: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Writing – review & editing. YG: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, 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 a grant (Award No. 2020-68013-32371) from the US Department of Agriculture – National Institute of Food and Agriculture. This work was also supported by a grant (Award No. NR216526XXXXC021, 'Cover Crop Initiative: A Collaborative Project to Advance Knowledge and Utilization of Cover Crops for Conservation Measures in Nebraska') from the US Department of Agriculture - Natural Resources Conservation Service.

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.2023.1277672/full#supplementary-material

References

Acharya, J., Moorman, T. B., Kaspar, T. C., Lenssen, A. W., Gailans, S., and Robertson, A. E. (2022). Effect of planting into a green winter cereal rye cover crop on growth and development, seedling disease, and yield of corn. *Plant Dis.* 106, 114–120. doi: 10.1094/PDIS-04-21-0836-RE

Adetunji, A. T., Ncube, B., Meyer, A. H., Olatunji, O. S., Mulidzi, R., and Lewu, F. B. (2021). Soil pH, nitrogen, phosphatase and urease activities in response to cover crop species, termination stage and termination method. *Heliyon* 7, e05980. doi: 10.1016/j.heliyon.2021.e05980

Aiyer, H., Fofana, B., Fraser, T., Caldwell, C., Mckenzie-Gopsill, A., Mills, A., et al. (2022). Choice of cover crop influences soil fungal and bacterial communities in Prince Edward Island, Canada. *Can. J. Microbiol.* 68, 465–482. doi: 10.1139/cjm-2021-0360

Almeida, D. S., Menezes-Blackburn, D., Zhang, H., Haygarth, P. M., and Rosolem, C. A. (2019). Phosphorus availability and dynamics in soil affected by long-term ruzigrass cover crop. *Geoderma* 337, 434–443. doi: 10.1016/j.geoderma.2018.09.056

Atefi, A., Ge, Y., Pitla, S., and Schnable, J. (2020). Robotic detection and grasp of maize and sorghum: stem measurement with contact. *Robotics* 9. doi: 10.3390/robotics9030058

Badgley, G., Field, C. B., and Berry, J. A. (2017). Canopy near-infrared reflectance and terrestrial photosynthesis. *Sci. Adv.* 3, e1602244. doi: 10.1126/sciadv.1602244

Bai, G., Ge, Y., Hussain, W., Baenziger, P. S., and Graef, G. (2016). A multi-sensor system for high throughput field phenotyping in soybean and wheat breeding. *Comput. Electron. Agric.* 128, 181–192. doi: 10.1016/j.compag.2016.08.021

Bai, G., Ge, Y., Scoby, D., Leavitt, B., Stoerger, V., Kirchgessner, N., et al. (2019). NU-Spidercam: A large-scale, cable-driven, integrated sensing and robotic system for advanced phenotyping, remote sensing, and agronomic research. *Comput. Electron. Agric.* 160, 71–81. doi: 10.1016/j.compag.2019.03.009

Bai, G., Jenkins, S., Yuan, W., Graef, G. L., and Ge, Y. (2018). Field-based scoring of soybean iron deficiency chlorosis using RGB imaging and statistical learning. *Front. Plant Sci.* 9, 1002. doi: 10.3389/fpls.2018.01002

Barnes, E. M., Clarke, T. R., Richards, S. E., Colaizzi, P. D., Haberland, J., Kostrzewski, M., et al. (2000). *Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data* (Madison: American Society of Agronomy).

Basche, A. D., and Roesch-McNally, G. E. (2017). Research topics to scale up cover crop use: Reflections from innovative Iowa farmers. *J. Soil Water Conserv.* 72, 59A–63A. doi: 10.2489/jswc.72.3.59A

Capri, C., Gatti, M., Fiorini, A., Ardenti, F., Tabaglio, V., and Poni, S. (2023). A comparative study of fifteen cover crop species for orchard soil management: water uptake, root density traits and soil aggregate stability. *Sci. Rep.* 13, 721. doi: 10.1038/s41598-023-27915-7

Cheng, M., Yin, D., Wu, W., Cui, N., Nie, C., Shi, L., et al. (2023). A review of remote sensing estimation of crop water productivity: definition, methodology, scale, and evaluation. *Int. J. Remote Sens.* 44, 5033–5068. doi: 10.1080/01431161.2023.2240523

Deines, J. M., Guan, K., Lopez, B., Zhou, Q., White, C. S., Wang, S., et al. (2023). Recent cover crop adoption is associated with small maize and soybean yield losses in the United States. *Glob Chang Biol.* 29, 794–807. doi: 10.1111/gcb.16489

Dong, L., Du, H., Han, N., Li, X., Zhu, D. E., Mao, F., et al. (2020). Application of convolutional neural network on Lei bamboo above-ground-biomass (AGB) estimation using worldview-2. *Remote Sens.* 12, 958. doi: 10.3390/rs12060958

Feng, L., Chen, S., Zhang, C., Zhang, Y., and He, Y. (2021). A comprehensive review on recent applications of unmanned aerial vehicle remote sensing with various sensors for high-throughput plant phenotyping. *Comput. Electron. Agric.* 182. doi: 10.1016/j.compag.2021.106033

Gamon, J. A., Serrano, J., and Surfus, J. S. (1997). PRI_an optical indicator of photosynthetic RUE across species, functional types, and nutrient levels. *Oecologia* 112, 492–501. doi: 10.1007/s004420050337

Inveninato Carmona, G., Robinson, E., Tonon Rosa, A., Proctor, C. A., and Mcmechan, A. J. (2022). Impact of cover crop planting and termination dates on arthropod activity in the following corn. *J. Econ Entomol* 115, 1177–1190. doi: 10.1093/jee/toac090

Irmak, S., Haman, D. Z., and Bastug, R. (2000). Determination of crop water stress index for irrigation timing and yield estimation of corn. *Agron. J.* 92, 1221–1227. doi: 10.2134/agronj2000.9261221x

Kaye, J. P., and Quemada, M. (2017). Using cover crops to mitigate and adapt to climate change. A review. *Agron. Sustain. Dev.* 37, 4. doi: 10.1007/s13593-016-0410-x

Kharel, T. P., Bhandari, A. B., Mubvumba, P., Tyler, H. L., Fletcher, R. S., and Reddy, K. N. (2023). Mixed-species cover crop biomass estimation using planet imagery. *Sens (Basel)* 23. doi: 10.3390/s23031541

Koehler-Cole, K., Basche, A., Thompson, L., and Rees, J. (2023). Comparing cover crop research in farmer-led and researcher-led experiments in the Western Corn Belt. *Front. Sustain. Food Syst.* 7. doi: 10.3389/fsufs.2023.1064251

Koehler-Cole, K., Elmore, R. W., Blanco-Canqui, H., Francis, C. A., Shapiro, C. A., Proctor, C. A., et al. (2020). Cover crop productivity and subsequent soybean yield in the western Corn Belt. *Agron. J.* 112, 2649–2663. doi: 10.1002/agj2.20232

Kumar, V., Singh, V., Flessner, M. L., Haymaker, J., Reiter, M. S., and Mirsky, S. B. (2023). Cover crop termination options and application of remote sensing for evaluating termination efficiency. *PloS One* 18, e0284529. doi: 10.1371/journal.pone.0284529

Lawson, A. N., Acker, V., and Friesen, L. F. (2006). Emergence timing of volunteer canola in spring wheat fields in Manitoba. *Weed Sci.* 54, 873–882. doi: 10.1614/WS-05-169.I.1

Lawson, A., Fortuna, A. M., Cogger, C., Bary, A., and Stubbs, T. (2012). Nitrogen contribution of rye–hairy vetch cover crop mixtures to organically grown sweet corn. *Renewable Agric. Food Syst.* 28, 59–69. doi: 10.1017/S1742170512000014

Liu, Y., Feng, H., Yue, J., Fan, Y., Jin, X., Zhao, Y., et al. (2022a). Estimation of potato above-ground biomass using UAV-based hyperspectral images and machine-learning regression. *Remote Sens.* 14. doi: 10.3390/rs14215449

Liu, Y., Feng, H., Yue, J., Li, Z., Jin, X., Fan, Y., et al. (2022b). Estimation of aboveground biomass of potatoes based on characteristic variables extracted from UAV hyperspectral imagery. *Remote Sens.* 14. doi: 10.3390/rs14205121

Marandola, D., Belliggiano, A., Romagnoli, L., and Ievoli, C. (2019). The spread of no-till in conservation agriculture systems in Italy: indications for rural development policy-making. *Agric. Food Econ* 7, 7. doi: 10.1186/s40100-019-0126-8

Meroni, M., Rossini, M., Guanter, L., Alonso, L., Rascher, U., Colombo, R., et al. (2009). Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications. *Remote Sens. Environ.* 113, 2037–2051. doi: 10.1016/j.rse.2009.05.003

Nichols, V., Martinez-Feria, R., Weisberger, D., Carlson, S., Basso, B., and Basche, A. (2020). Cover crops and weed suppression in the U.S. Midwest: A meta-analysis and modeling study. *Agric. Environ. Lett.* 5, e20022. doi: 10.1002/ael2.20022

O'Brien, P. L., Emmett, B. D., Malone, R. W., Nunes, M. R., Kovar, J. L., Kaspar, T. C., et al. (2022). Nitrate losses and nitrous oxide emissions under contrasting tillage and cover crop management. *J. Environ. Qual* 51, 683–695. doi: 10.1002/jeq2.20361

Oliveira, M. C., Butts, L., and Werle, R. (2019). Assessment of cover crop management strategies in Nebraska, US. *Agriculture* 9. doi: 10.3390/agriculture9060124

Popovici, R., Ranjan, P., Bernard, M., Usher, E. M., Johnson, K., and Prokopy, L. S. (2023). The social factors influencing cover crop adoption in the Midwest: A controlled comparison. *Environ. Manage.* 3, 614–629. doi: 10.1007/s00267-023-01823-y

Rondeaux, G., Steven, M., and Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sens. Environ.* 55, 95–107. doi: 10.1016/0034-4257(95)00186-7

Rosa, A. T., Creech, C. F., Elmore, R. W., Rudnick, D. R., Lindquist, J. L., Butts, L., et al. (2021). Contributions of individual cover crop species to rainfed maize production in semi-arid cropping systems. *Field Crops Res.* 271. doi: 10.1016/j.fcr.2021.108245

Rouse, J. W. Jr., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). "Monitoring vegetation systems in the Great Plains with ERTS," in *NASA. Goddard Space Flight Center 3d ERTS-1 Symp.* (Washington, D.C.: NASA).

Ruis, S. J., Blanco-Canqui, H., Creech, C. F., Koehler-Cole, K., Elmore, R. W., and Francis, C. A. (2019). Cover crop biomass production in temperate agroecozones. *Agron. J.* 111, 1535–1551. doi: 10.2134/agronj2018.08.0535

Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., et al. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. *Eur. J. Agron.* 70, 112–123. doi: 10.1016/j.eja.2015.07.004

Schwarte, A. J., Gibson, L. R., Karlen, D. L., Dixon, P. M., Liebman, M., and Jannink, J.-L. (2006). Planting date effects on winter triticale grain yield and yield components. *Crop Sci.* 46, 1218–1224. doi: 10.2135/cropsci2005.06-0130

Sharma, P., Leigh, L., Chang, J., Maimaitijiang, M., and Caffe, M. (2022). Above-ground biomass estimation in oats using UAV remote sensing and machine learning. *Sens (Basel)* 22. doi: 10.3390/s22020601

Slafer, G. A., and Savin, R. (1991). Developmental base temperature in different phenological phases of wheat (Triticum aestivum). *J. Exp. Bot.* 42, 1077–1082. doi: 10.1093/jxb/42.8.1077

Suárez, L., Zarco-Tejada, P. J., Sepulcre-Cantó, G., Pérez-Priego, O., Miller, J. R., Jiménez-Muñoz, J. C., et al. (2008). Assessing canopy PRI for water stress detection with diurnal airborne imagery. *Remote Sens. Environ.* 112, 560–575. doi: 10.1016/j.rse.2007.05.009

Szuleta, E., Phillips, T., Knott, C. A., Lee, C. D., and Van Sanford, D. A. (2022). Influence of planting date on winter rye performance in Kentucky. *Agronomy* 12. doi: 10.3390/agronomy12112887

Wang, L., Li, J., Zhao, L., Zhao, B., Bai, G., Ge, Y., et al. (2021). Investigate the potential of UAS-based thermal infrared imagery for maize leaf area index estimation. *SPIE Defense + Commercial Sens.* SPIE. 11747. doi: 10.1117/12.2586694

Wang, F., Yang, M., Ma, L., Zhang, T., Qin, W., Li, W., et al. (2022). Estimation of above-ground biomass of winter wheat based on consumer-grade multi-spectral UAV. *Remote Sens.* 14. doi: 10.3390/rs14051251

Whitmire, C. D., Vance, J. M., Rasheed, H. K., Missaoui, A., Rasheed, K. M., and Maier, F. W. (2021). Using machine learning and feature selection for Alfalfa yield prediction. *AI* 2, 71–88. doi: 10.3390/ai2010006

Yuan, W., Wijewardane, N. K., Jenkins, S., Bai, G., Ge, Y., and Graef, G. L. (2019). Early prediction of soybean traits through color and texture features of canopy RGB imagery. *Sci. Rep.* 9, 14089. doi: 10.1038/s41598-019-50480-x

Yue, J., Feng, H., Yang, G., and Li, Z. (2018). 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

Zhang, Z., Kaye, J. P., Bradley, B. A., Amsili, J. P., and Suseela, V. (2022). Cover crop functional types differentially alter the content and composition of soil organic carbon in particulate and mineral-associated fractions. *Glob Chang Biol.* 28, 5831–5848. doi: 10.1111/gcb.16296

Zhang, C., Marzougui, A., and Sankaran, S. (2020). High-resolution satellite imagery applications in crop phenotyping: An overview. *Comput. Electron. Agric.* 175, 105584. doi: 10.1016/j.compag.2020.105584

Zhou, J., and Nguyen, T. H. (2021). *High-Throughput Crop Phenotyping* (Cham: Springer).

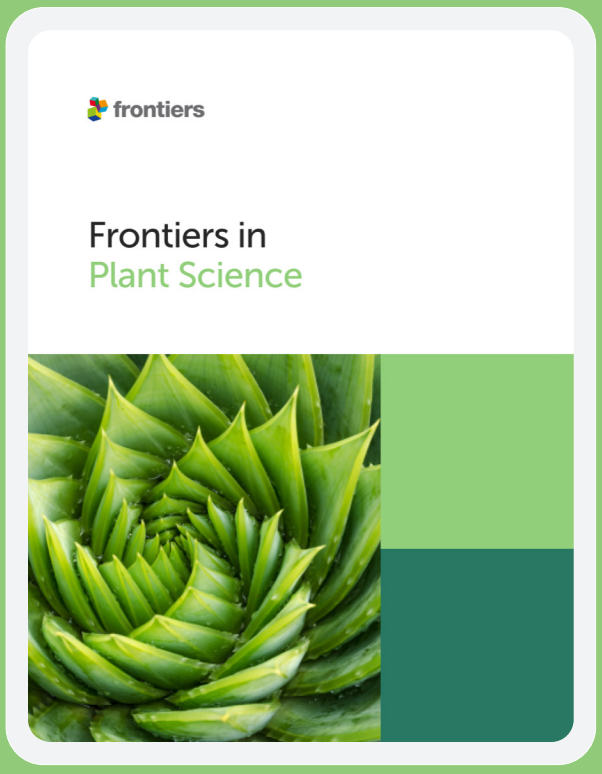
Frontiers in
Plant Science

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