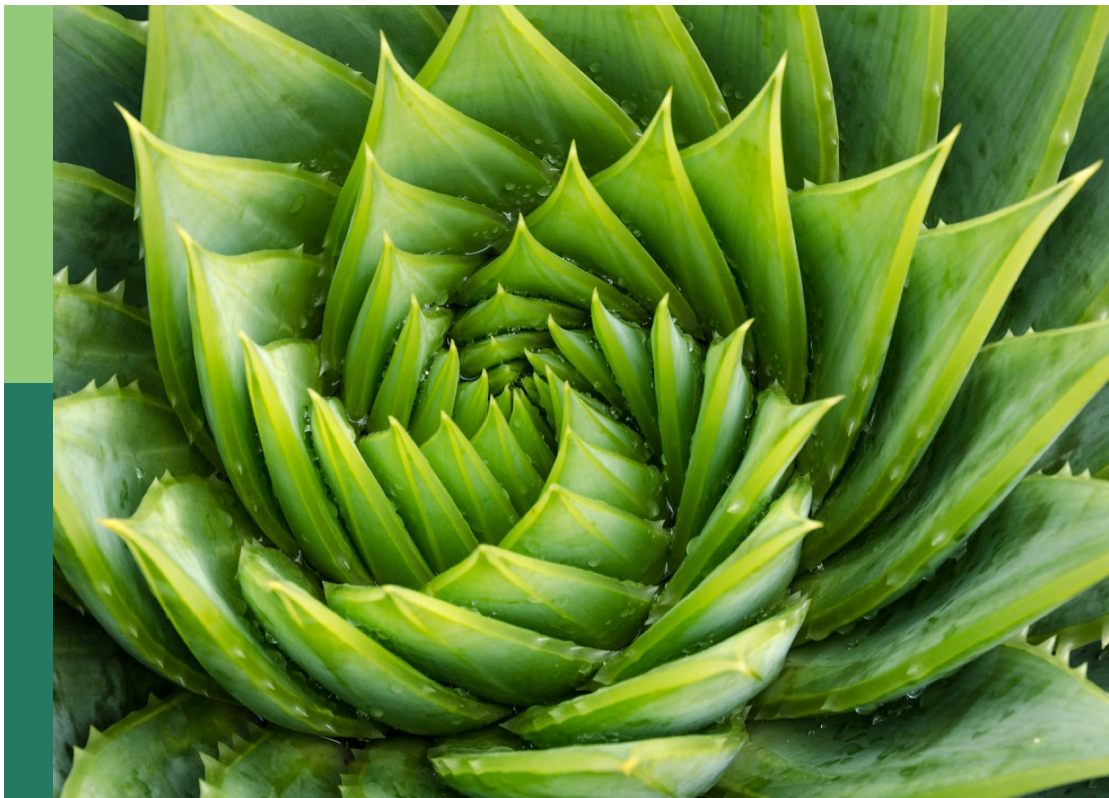
frontiers | Research Topics

Advanced methods, equipment and platforms in precision field crops protection, volume II

Edited by
Jun Ni, Peter Yuen, Jingcheng Zhang and Ning Yang

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 978-2-8325-7019-7
DOI 10.3389/978-2-8325-7019-7

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

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

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

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

Advanced methods, equipment and platforms in precision field crops protection, volume II

Topic editors

Jun Ni — Shanghai University of Engineering Science, China
Peter Yuen — Cranfield University, United Kingdom
Jingcheng Zhang — Hangzhou Dianzi University, China
Ning Yang — Jiangsu University, China

Citation

Ni, J., Yuen, P., Zhang, J., Yang, N., eds. (2025). *Advanced methods, equipment and platforms in precision field crops protection, volume II*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-7019-7

Table of contents

05 Editorial: Advanced methods, equipment and platforms in precision field crops protection, volume II
Zhaoyuan Zhang, Ning Yang, Peter Yuen, Jingcheng Zhang and Jun Ni

08 An integrated framework for UAV-based precision plant protection in complex terrain: the ACHAGA solution for multi-tea fields
Pengyang Zhang, Yangyang Liu and Hongbin Du

28 Motion behavior of droplets on curved leaf surfaces driven by airflow
Zhou-Ming Gao, Wei Hu, Xiao-Ya Dong, Xiao-Yuan Zhao, Song Wang, Jian Chen and Bai-Jing Qiu

43 Intelligent agriculture: deep learning in UAV-based remote sensing imagery for crop diseases and pests detection
Hongyan Zhu, Chengzhi Lin, Gengqi Liu, Dani Wang, Shuai Qin, Anjie Li, Jun-Li Xu and Yong He

67 Design and validation of a real-time cassava planter seed quality monitoring system based on optical fiber sensors and rotary encoders
Bin Yan, Zhende Cui, Ganran Deng, Guojie Li, Shuang Zheng, Fengguang He, Ling Li, Pinlan Chen, Xilin Wang, Sili Zhou, Ye Dai, Shuangmei Qin and Zehua Liu

79 A novel lightweight YOLOv8-PSS model for obstacle detection on the path of unmanned agricultural vehicles
Zhijian Chen, Yijun Fang, Jianjun Yin, Shiyu Lv, Farhan Sheikh Muhammad and Lu Liu

95 Effect of operational parameters on droplet deposition characteristics using an unmanned aerial vehicle for banana canopy
Jiaxiang Yu, Xing Xu, Jieli Duan, Yinlong Jiang, Haotian Yuan, Huazimo Liang, Shuaijie Jing and Zhou Yang

110 A novel method based on lesion expansion to assess plant disease severity
Feng Qin, Haiguang Wang, Qian Jiang and Hongli Wang

131 Deep learning-based target spraying control of weeds in wheat fields at tillering stage
Haiying Wang, Yu Chen, Shuo Zhang, Peijie Guo, Yuxiang Chen, Guangrui Hu and Yuxuan Ma

146 Plant stem and leaf segmentation and phenotypic parameter extraction using neural radiance fields and lightweight point cloud segmentation networks
Gaofei Qiao, Zhibin Zhang, Bin Niu, Sijia Han and Enhui Yang

165 VMamba for plant leaf disease identification: design and experiment
Hewei Zhang, Shengzhou Li, Jialong Xie, Zihan Chen, Jiyang Chen and Jianwen Guo

182 Optimization of spray operation parameters of unmanned aerial vehicle confers adequate levels of control of fall armyworm (*Spodoptera frugiperda*)
Ying Liu, Xiao Liang, Chunling Wu, Xingkui An, Mufeng Wu, Zihua Zhao, Zhihong Li and Qing Chen

195 Deep learning-based anomaly detection for precision field crop protection
Cheng Wei, Yifeng Shan and MengZhe Zhen

218 Research on fault prediction and speed control system for unmanned combine harvesters based on IPSO-SVM and fuzzy logic
Shaocen Zhang, Chongquan Zang, Zhang Yang, Lingyu Tang, Kun Wang, Anzhe Wang, Wenming Chen, Qi Song and Xinhua Wei

235 Quantifying assessment of American ginseng (*Panax quinquefolius L.*) main root bruising based on FEM
Pengcheng Jia, Guixuan Zhu, Junzhi Chen, Dong Wang, Han Tang and Jinwu Wang

251 Deep learning-based time series prediction for precision field crop protection
Tao He, Meijin Li and Dong Jin

274 Harnessing large vision and language models in agriculture: a review
Hongyan Zhu, Shuai Qin, Min Su, Chengzhi Lin, Anjie Li and Junfeng Gao

299 FHB-Net: a severity level evaluation model for wheat *Fusarium* head blight based on image-level annotated aerial RGB images
Shuxin Zhu, Huayong Li, Shun Zou, Huanliang Xu and Zhaoyu Zhai

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 07 October 2025
DOI 10.3389/fpls.2025.1704271

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Lei Shu,
Nanjing Agricultural University, China

*CORRESPONDENCE
Ning Yang
yangning7410@163.com
Peter Yuen
p.yuen@cranfield.ac.uk
Jingcheng Zhang
zhangjcrs@hdu.edu.cn
Jun Ni
nijun1218@sues.edu.cn

RECEIVED 12 September 2025
ACCEPTED 26 September 2025
PUBLISHED 07 October 2025

CITATION
Zhang Z, Yang N, Yuen P, Zhang J and Ni J (2025) Editorial: Advanced methods, equipment and platforms in precision field crops protection, volume II.
Front. Plant Sci. 16:1704271.
doi: 10.3389/fpls.2025.1704271

COPYRIGHT
© 2025 Zhang, Yang, Yuen, Zhang and Ni. 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: Advanced methods, equipment and platforms in precision field crops protection, volume II

Zhaoyuan Zhang[1], Ning Yang[1*], Peter Yuen[2*], Jingcheng Zhang[3*] and Jun Ni[4*]

[1]School of Electrical and Information Engineering, Jiangsu University, Zhenjiang, Jiangsu, China, [2]Centre for Defence Engineering and Physical Science, Cranfield University, Bedfordshire, United Kingdom, [3]School of Automation, Hangzhou Dianzi University, Hangzhou, Zhejiang, China, [4]Laboratory of Intelligent Control and Robotics, Shanghai University of Engineering Science, Shanghai, China

KEYWORDS

precision perception, crop protection, intelligent decision-making and control, multi-source data fusion, smart agriculture, UAVs

Editorial on the Research Topic
Advanced methods, equipment and platforms in precision field crops protection, volume II

1 Introduction

The proliferation of pests, diseases, and weeds constitutes a primary constraint on agricultural productivity. Recently, leveraging modern information technology to realize precision crop protection has become a key area of research within the domain of smart agriculture. This Research Topic focuses on novel sensor technologies for early detection and identification of biotic stresses, artificial intelligence-based methods for intelligent diagnostics and phenotypic analysis, and the development of precision equipment and systems for variable-rate crop protection strategies. Concurrently, this Research Topic delves into the innovative integration of digital twin models, the Internet of Things (IoT), and cloud-based platforms into crop protection paradigms, while also offering a perspective on the future trajectories of this research field.

2 Research topic overview

We have compiled a collection of two review articles and fifteen original research studies focusing on the focal points of this Research Topic. The authors' contributions predominantly encompass the following research domains: the intelligent perception and identification of plant pests and diseases; automation in agricultural operations

and precision control systems; the precise quantification of plant phenotypes and disease severity using deep learning; autonomous sensing for intelligent farm machinery coupled with lightweight model optimization; the application of Unmanned Aerial Vehicles (UAVs) for crop disease detection; and the development of a macro-scale agricultural decision support system leveraging multi-modal data fusion. This editorial elucidates the significant progress in intelligent monitoring and precision protection for field crops, underscoring the transformative influence of these modern technologies on global agriculture.

3 Precision sensing and monitoring of crop diseases

In crop protection, precision perception and monitoring are fundamental steps, encompassing tasks such as pest and disease identification, disease severity assessment, and crop phenotype extraction. Qiao et al. proposed a method for 3D crop reconstruction and parameter extraction that combines Neural Radiance Fields (NeRF) with a lightweight point cloud segmentation network. This study achieved high-precision segmentation in maize 3D reconstruction, and the proposed method outperformed five existing mainstream networks. Concurrently, the maize stem thickness, plant height, and leaf parameters extracted by this method showed high consistency with manual measurements, demonstrating its reliability and applicability. Turning to wheat stripe rust, Qin et al. introduced a severity assessment method based on lesion expansion. Through experiments with nine method combinations, the authors found that the optimal combination achieved an accuracy of 96.16% in severity assessment—significantly outperforming traditional visual methods and non-expansion approaches—and resolved the discrepancy between grading standards and actual lesion areas. A review by Zhu et al. revealed that Unmanned Aerial Vehicles (UAVs) equipped with multispectral, RGB, and thermal imaging sensors, when integrated with deep learning algorithms, can achieve rapid pest and disease identification at the field scale with accuracy significantly higher than traditional manual surveys. For instance, deep convolutional networks exceeded 95% accuracy in detecting maize leaf blight, whereas manual visual inspection was only approximately 80% accurate. Furthermore, Zhu et al. noted that Large Vision and Language Models (LVMs/LLMs) exhibit "zero-shot" and "few-shot" learning capabilities in agricultural multi-modal data fusion, holding promise for substantially enhancing monitoring performance in agricultural scenarios that lack large-scale annotated data. Their systematic review emphasized the potential of these models in understanding remote sensing images and generating agricultural decisions. The perception and monitoring of pests, diseases, and weeds is evolving from 2D imaging to 3D modeling, lesion quantification, and large model-driven multi-source fusion, thereby greatly enhancing the accuracy and automation of crop health monitoring.

4 Intelligent decision-making and control for crop protection

Building upon monitoring data, intelligent decision-making and control technologies facilitate proactive and intelligent crop protection by predicting and identifying potential threats. He et al. proposed the deep learning-based time-series forecasting models, SADF-Net and the RAADA network, which fuse satellite imagery, sensor data, and meteorological data to enable early warnings for crop disease risks. This approach achieved significantly higher temporal forecasting accuracy than LSTM and GRU models, with an improvement of approximately 8%–10% in both prediction accuracy and stability. Wei et al. introduced the IMSFNet+AROS framework, which performs multi-modal anomaly detection using data from UAVs, satellites, and ground-based sensors to aid in the early discovery of in-field anomalies. In the domain of intelligent control for agricultural equipment, Chen et al. developed the YOLOv8-PSS, a lightweight obstacle detection model. Compared to the original YOLOv8, this model reduced the number of parameters by 55.8% and computational overhead by 51.2%, while maintaining a mean Average Precision (mAP) of 90.6%. Localization error was controlled within a range of 2.73%–4.44%, significantly enhancing the safety of unmanned agricultural machinery in complex field environments. Meanwhile, research by Zhang et al. demonstrated that by improving the IPSO-SVM algorithm and fuzzy logic, it is possible to achieve fault prediction and adaptive speed regulation for unmanned combine harvesters, thereby enhancing the equipment's operational stability and level of intelligence. In summary, intelligent decision-making and control are progressively actualizing a closed-loop management paradigm of "prediction-diagnosis-control," providing a reliable foundation for proactive defense and intelligent execution in crop protection.

5 Precision operations and intelligent equipment

Operation optimization is critical for translating monitoring and decision-making outcomes into practical field applications. Current research primarily focuses on the optimization of Unmanned Aerial Vehicle (UAV) spraying and the mechanisms of droplet deposition. In the field of UAV spraying, the ACHAGA algorithm proposed by Zhang et al. was found to optimize UAV flight paths in complex tea plantation terrains, reducing flight distance and the number of turns. This approach improved efficiency by approximately 20%–30% compared to manual planning, thereby enhancing crop protection efficacy. A study by Yu et al. demonstrated that different flight altitudes and droplet sizes significantly impacted spray deposition distribution on banana canopies, with an altitude of 4 m and a droplet size of 100 μm achieving optimal deposition. This proves the significant influence of operational parameters on deposition uniformity and penetration. Liu et al. optimized UAV spraying parameters for the control of the fall armyworm (Spodoptera frugiperda), showing

that a control efficacy of over 90% was achieved using an XR110015VS nozzle, a spray volume of 37.5 L/ha, and a flight height of 2.5 m. This performance was comparable to that of traditional knapsack sprayers but with significantly reduced pesticide dosage and labor intensity. Wind tunnel experiments by Gao et al. indicated that the critical wind speed for droplets on curved leaf surfaces under airflow is significantly correlated with droplet diameter and leaf curvature. As leaf curvature increases, droplets are more easily dislodged, with the difference reaching 24.8%. Furthermore, the acceleration difference for large droplets can be as high as 68%, revealing the influence of airflow and leaf structure on droplet deposition. Meanwhile, research by Wang et al. on targeted spraying in wheat fields showed that the detection accuracy of their improved YOLO model reached 95.6% during the tillering stage—an improvement of 7.3% over the original YOLOv5. This also led to a 40% reduction in pesticide usage, indicating that intelligent spraying can effectively lower the environmental burden while ensuring control efficacy. In conclusion, research on equipment optimization not only enhances the adaptability of UAVs in complex field environments but also promotes improvements in pesticide use efficiency, achieving the goal of "high efficacy with low pesticide volume" in crop protection.

6 Conclusion

In summary, current research has established an integrated technological chain for crop protection, progressing from precision perception and monitoring to intelligent decision-making and control, and culminating in precision operation and equipment optimization. At the perception level, a significant leap has been made from 2D imaging to 3D reconstruction, with large models enhancing the capacity for multi-source data fusion. At the decision-making level, artificial intelligence algorithms have substantially improved the accuracy of prediction and diagnosis, enabling autonomous operation of agricultural machinery in complex environments. Subsequently, at the operational level, advancements in UAV path planning and spray parameter optimization have markedly increased pesticide use efficiency and operational throughput. Nevertheless, several challenges persist, including insufficient real-time processing capabilities, limited model generalizability, and the need for enhanced equipment stability in complex operational settings. Future research should therefore focus on the deeper integration of advanced sensors, intelligent algorithms, and digital twin platforms to construct a more efficient, environmentally friendly, and intelligent system for field crop protection.

Author contributions

ZZ: Writing – original draft, Writing – review & editing. NY: Writing – original draft, Writing – review & editing. PY: Writing – review & editing. JZ: Writing – review & editing. JN: Writing – original draft, Writing – review & editing.

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 author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

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

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

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 26 September 2024
DOI 10.3389/fpls.2024.1440234

Check for updates

OPEN ACCESS

EDITED BY
Jun Ni,
Nanjing Agricultural University, China

REVIEWED BY
Amila Thibbotuwawa,
University of Moratuwa, Sri Lanka
Morufu Olalekan Raimi MNES,
REHO, LEHO, FAIWMES,
Federal University, Nigeria
Elio Romano,
Centro di ricerca per l'Ingegneria e le Trasformazioni Agroalimentari (CREA-IT), Italy

*CORRESPONDENCE
Pengyang Zhang
zpy19982222@163.com
Hongbin Du
ly994199903@163.com

RECEIVED 29 May 2024
ACCEPTED 27 August 2024
PUBLISHED 26 September 2024

CITATION
Zhang P, Liu Y and Du H (2024) An integrated framework for UAV-based precision plant protection in complex terrain: the ACHAGA solution for multi-tea fields.
Front. Plant Sci. 15:1440234.
doi: 10.3389/fpls.2024.1440234

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

An integrated framework for UAV-based precision plant protection in complex terrain: the ACHAGA solution for multi-tea fields

Pengyang Zhang[1,2]*, Yangyang Liu[3] and Hongbin Du[1,2]*

[1]College of Horticulture and Forestry, Tarim University, Alar, China, [2]Xinjiang Production and Construction Corps Key Laboratory of Facility Agriculture, Alar, China, [3]College of Engineering, Anhui Agricultural University, Hefei, China

UAV-based plant protection represents an efficient, energy-saving agricultural technology with significant potential to enhance tea production. However, the complex terrain of hilly and mountainous tea fields, coupled with the limited endurance of UAVs, presents substantial challenges for efficient route planning. This study introduces a novel methodological framework for UAV-based precision plant protection across multiple tea fields, addressing the difficulties in planning the shortest routes and optimal flights for UAVs constrained by their endurance. The framework employs a hyperbolic genetic annealing algorithm (ACHAGA) to optimize UAV plant protection routes with the objectives of minimizing flight distance, reducing the number of turns, and enhancing route stability. The method involves two primary steps: cluster partitioning and sortie allocation for multiple tea fields based on UAV range capabilities, followed by refining the UAV's flight path using a combination of hyperbolic genetic and simulated annealing algorithms with an adaptive temperature control mechanism. Simulation experiments and UAV route validation tests confirm the effectiveness of ACHAGA. The algorithm consistently identified optimal solutions within an average of 40 iterations, demonstrating robust global search capabilities and stability. It achieved an average reduction of 45.75 iterations and 1811.93 meters in the optimal route, with lower variation coefficients and extreme deviations across repeated simulations. ACHAGA significantly outperforms these algorithms, GA, GA-ACO, AFSA and BSO, which are also heuristic search strategies, in the multi-tea field route scheduling problem, reducing the optimal routes by 4904.82 m, 926.07 m, 3803.96 m and 800.11 m, respectively. Field tests revealed that ACHAGA reduced actual flight routes by 791.9 meters and 359.6 meters compared to manual and brainstorming-based planning methods, respectively. Additionally, the algorithm reduced flight scheduling distance and the number of turns by 11 compared to manual planning. This study provides a theoretical and technical foundation for managing large-scale tea plantations in challenging landscapes and serves as a reference for UAV precision operation planning in complex environments.

KEYWORDS

UAV-based plant protection, multi-tea field plant protection, unmanned aerial vehicle, replenishment -scheduling optimization, scheduling route planning

1 Introduction

In recent years, Unmanned Aerial Vehicles (UAVs) have emerged as a crucial tool in modern agriculture, offering efficient and eco-friendly solutions tailored to specific needs. The tea industry, with its intricate field layouts, diverse topography, and varied cultivation patterns, exemplifies the unique challenges of plant protection that UAV technology can address (Zhang and Zhu, 2023; Paraforos et al., 2022). Historically, agriculture has witnessed a gradual but significant shift towards automation and technological integration. UAVs, once a novelty, are now at the forefront of this evolution, playing a pivotal role in crop monitoring, pest control, and resource management (Bao et al., 2023; Verma et al., 2023). Their ability to navigate difficult terrains and distribute treatments with unprecedented precision has not only enhanced crop yield and quality but also reduced environmental impact.

Despite the clear advantages, scheduling and routing UAVs in the tea industry remain complex tasks. Traditional methods, heavily reliant on human experience and subjective judgment, fall short in addressing the non-linear and dynamic nature of tea field layouts. The existing literature on UAV route optimization in complex terrains highlights several approaches, such as the A* algorithm, Rapidly-exploring Random Trees (RRT), and various meta-heuristic algorithms (Ait Saadi et al., 2022; He et al., 2024). However, these methods often struggle with the high computational demands and adaptability required for real-time operations in uneven and unpredictable environments. Additionally, the limited battery life of UAVs imposes stringent constraints on flight distance and operational time. This necessitates highly efficient and adaptive route optimization to ensure UAVs can complete their tasks within a single charge. The non-linear and dynamic nature of tea field layouts further complicates the planning process, as routes must continuously adapt to the intricate and changing topography to ensure comprehensive coverage and efficient resource use.

This paper explores the UAV multi-tea field resupply and dispatch route planning challenge, encompassing both site selection and route optimization. Extensive research has been conducted globally on UAV dispatch center location planning. For instance, Bian et al. (Bian et al., 2022). introduced a UAV base station positioning approach utilizing the spiral algorithm, which formulates an air-to-ground propagation model reflecting the actual geographic environment to identify optimal UAV base station sites in complex settings. However, this approach primarily targets communication optimization rather than the logistical challenges specific to agricultural operations. Famili et al. (Famili and Stavrou, 2022) acknowledged the constraints of UAV battery life, proposing an optimization framework based on an approximation algorithm to ascertain the minimal number of charging stations needed for sustained flight capability. This study, while addressing the critical issue of battery life, does not fully tackle the complexities of dynamic route planning in varied terrains. Saavedrai et al. (Saavedra et al., 2021) developed an adaptive and comprehensive capacity-constrained localization-routing (CLRP) model for UAV identification in post-disaster relief, effectively pinpointing ideal locations and routes for UAV hubs in preparation for disasters. Although this model is robust for emergency scenarios, it does not account for the continuous and repetitive nature of agricultural UAV operations. Li et al. (Li et al., 2019) enhanced the firefly algorithm for logistics distribution center siting focuses on maximizing benefits based on the performance characteristics and limitations of logistics UAVs. This approach is effective for static distribution centers but lacks the adaptability required for real-time agricultural resupply and dispatch. While these studies above primarily focused on identifying optimal locations for specific purposes such as communication, charging, and rescue, the current paper addresses the distinct challenge of planning dynamic multi-tea field supply and dispatch routes. This entails UAV route optimization for efficient resupply and dispatch across tea fields, ensuring effective operations and comprehensive coverage—a novel and complex aspect of aerial plant protection planning.

The scheduling of UAV route tasks is a significant research area, involving the development of optimal flight schemes for UAVs across various application scenarios. Researchers worldwide have devised numerous route planning algorithms for this domain, engaging in comprehensive discussions centered on diverse optimization goals. For instance, Li et al. (Li et al., 2022c) redefined the challenge of optimal task distribution among multiple UAVs as a combinatorial optimization problem and developed a sequence-independent enumeration algorithm, significantly reducing flight frequency. However, this method does not integrate resupply point planning, which is crucial for large-scale agricultural settings. Xu et al. (Xu et al., 2020) addressed the assignment and sequencing of UAV operational tasks by formulating a bi-objective model, prioritizing non-operational flight distance and total operational time, and introduced an enhanced MOSFLA algorithm. This model is effective for minimizing travel distances, but it does not consider the dynamic nature of resupply in multi-field environments. Sun et al. (Sun et al., 2020) presented a dragonfly-inspired scheduling method for agricultural UAVs, focusing on plant protection and charging operations, which efficiently identifies near-optimal schedules. Despite its efficiency, it lacks a comprehensive approach to resupply point selection. Li et al. (Li et al., 2022b) proposed a novel approach for route planning and task allocation during UAV fly-over operations, optimizing task distribution through a particle swarm algorithm with flight control time as the objective, constrained by UAV battery life and payload capacity. This approach, while innovative, does not address the specific requirements of multi-field agricultural resupply. Fesenko et al. (Fesenko et al., 2020) responded to the nascent state of UAV monitoring technology for nuclear power plants by proposing an algorithm for automated battery exchange at aerial deployment sites, facilitating the development of UAV flight schedules and automatic replacement systems. This algorithm is tailored for fixed installation sites and does not accommodate the fluid resupply needs of agricultural UAVs. While these aforementioned algorithms achieve cost-effective and concise scheduling solutions for their respective problems, they diverge from the multi-tea field

route scheduling problem based on supply partitioning addressed in this paper. Primarily, the problem models employed by the aforementioned methods differ from the one in this study. Moreover, previous methods have not considered UAV resupply point planning and selection, focusing instead on route optimization for small-scale operations with predetermined fixed resupply points. In contrast, this study's multi-tea field scheduling route planning problem also encompasses the UAV multi-tea field resupply challenges including resupply location selection and multi-task allocation planning. This research area has seen limited exploration, thus this study contributes to bridging a notable gap in the field.

Despite its efficiency, it lacks a comprehensive approach to resupply point selection. To overcome these limitations, this paper introduces a novel scheduling and routing methodology for UAVs in multi-tea field operations. By analyzing the operational area's characteristics and requirements, we have developed a model that not only identifies optimal supply points but also provides a comprehensive route plan for multi-field plant protection. The non-linear and dynamic nature of tea field layouts, combined with the limited battery life of UAVs, poses significant challenges in ensuring efficient and continuous operations. This study redefines the UAV multi-field tea plantation resupply and routing challenge as a multiple traveling salesman problem (mTSP), offering a tailored solution to the unique distribution and routing complexities of tea fields. Through this innovative approach, we aim to facilitate swift, effective, and consistent plant protection operations, setting a new benchmark for precision agriculture in the tea industry.

2 Details of optimization techniques

2.1 Environmental projection of tea fields

This study aims to facilitate the scheduling and route planning of UAV operations across multiple tea fields, necessitating environmental modeling of the tea plantation to acquire geographic coordinates of individual tea field. As depicted in Figures 1A, B, the orthophoto map of the tea plantation is derived using a remote sensing dataset, which processes and translates the geographic coordinates into planar coordinates via the Mercator projection method:

The Mercator projection is a way to represent the curved surface of the Earth on a flat plane. It preserves the right angles between latitude and longitude lines, making it useful for navigation. However, it can cause distortions, especially near the poles. In this method, points on the Earth's surface with coordinates $(0, \lambda_0)$ are transformed to planar coordinates (x, y), where x represents horizontal distances and y is determined by the following Equations 1 and 2:

$$x = R(\lambda - \lambda_0)\cos(\varphi_0) \tag{1}$$

$$y = R\ln[\tan(\frac{\pi}{4} + \frac{\varphi}{2\sec(\varphi_0)}] \tag{2}$$

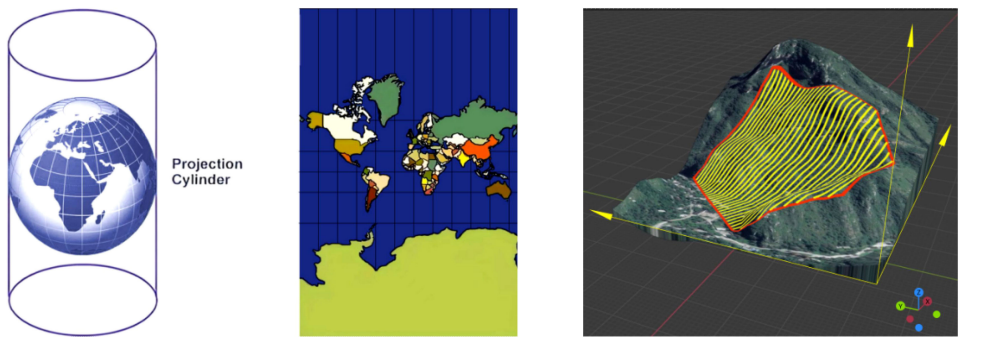
Given known planar coordinates, conversion back to geographic coordinates is achievable through Equations 3 and 4.

$$\varphi = 2\tan^{-1}(e^y) - \frac{1}{2}\pi \tag{3}$$

$$\lambda = x + \lambda_0 \tag{4}$$

Where x is the projected horizontal coordinate; y is the projected vertical coordinate; R is the Earth's radius; λ_0 is the central meridian's longitude; φ_0 is the central meridian's latitude.

UAV route planning for plant protection can be categorized into two approaches: single-area planning and multi-field planning. Single-area planning focuses solely on operational width to determine the shortest coverage route, while multi-field planning involves projecting regional boundaries and establishing a planar coordinate system. This study, using a hilly mountainous region as a reference, selects several polygonal fields at random for the operational area. These fields reside within the eastern longitudinal and northern latitudinal zones. Employing the two fields with the minimal latitude and longitude as the reference



(a) Mercator Projection Method **(b)** 3d Schematic Diagram of Tea Plantation

FIGURE 1
Schematic diagram of projected coordinate system of tea plantation. **(A)** Mercator Projection Method. **(B)** 3d Schematic Diagram of Tea Plantation.

point, all operational area boundaries fall within the first quadrant, leading to the establishment of a multi-tea field environmental coordinate system, depicted in Figure 2.

Previous research results indicate that by employing a coordinate system—where the boundaries of an operational area define the x-axis origins—we can calculate and compare to ascertain the most efficient full-coverage routes within a single operational area (Liu et al., 2022). These routes minimize the journey length, reduce excess coverage, and ensure that there is exactly one optimal full-coverage route per operational area, characterized by unique starting and ending points. For multi-area scheduling, the voyage encompasses the distance from the endpoint of one field to the starting point of the next, as depicted in Figure 3. This paper introduces the midpoint V3, between the start point V1 and the endpoint V2 of a single area, as the defining vertex representing the area. By extracting these characteristic vertices from multiple fields, we generate a discrete point distribution graph, thereby transforming the multi-tea field scheduling problem into a multiple traveling salesman problem model.

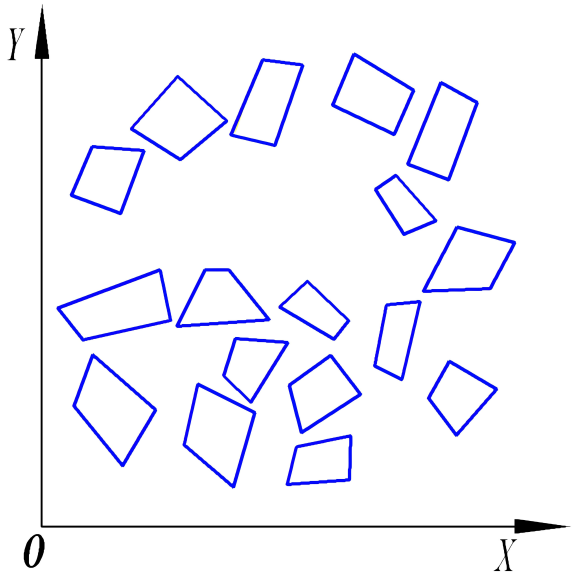


FIGURE 2
Constructing the coordinate system of multi-tea field operation area.

This study investigates the UAV planting operation planning in hilly, mountainous tea fields, a variant of the multiple traveling salesman problem with unmanned aerial vehicle (MTSPU) (Zhang et al., 2023; Yan et al., 2024). It accounts for the dynamic distances between the supply center and each plant protection operational areas. To address this intricate issue, we propose a hybrid optimization approach combining the K-means clustering and heuristic optimization algorithms. The methodology unfolds in two phases: initially, the K-means algorithm partitions the tea fields, and then supplemental locations and UAV schedules are determined based on workload. Subsequently, the heuristic algorithm orders the operational sequences within each sector, formulating the UAVs' planned routes.

2.2 Efficient hierarchical clustering assignment algorithms

2.2.1 Partitioning of operational areas using clustering algorithms

The division clustering method primarily computes distances between sample data to facilitate uniform partitioning of dispersed field vertices based on proximity (Das et al., 2023; Kuo et al., 2016). This study employs the K-means clustering algorithm to segment tea fields into sub-areas and strategize the replenishment locations according to workload. Silhouette coefficient, as defined in Equations 5 and 6, are calculated to ascertain the optimal number of clusters, thereby enhancing UAV operational efficiency and partitioning quality.

$$S(i) = \frac{b(i) - a(i)}{max(a(i), b(i))} \tag{5}$$

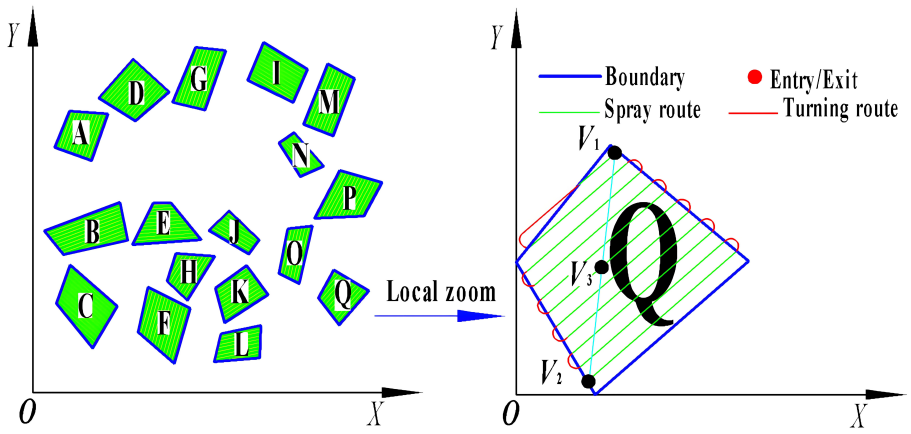


FIGURE 3
Extract field feature vertices.

$$a(i) = \frac{a_1 + a_2 + a_3 \ldots a_n}{n_x}, \quad b(i) = \frac{b_1 + b_2 + b_3 \ldots b_n}{n_y} \tag{6}$$

Where $a(i)$ is the denseness of clusters; $b(i)$ is the dispersion between clusters; $S(i)$ is the silhouette coefficient; n_x is the count of vertices within the data of a specific cluster; n_y is the count of vertices within the nearest neighboring cluster.

2.2.2 Resupply location planning algorithm

For selecting supply points post-clustering for each plant protection partition, it is essential to consider the return points optimally. As delineated in Equations 7 and 8, the planning of supply points is conceptualized as an optimization problem. The objective is to minimize the aggregate distance from the target point coordinates c to the surrounding point coordinates c_i, thereby identifying the most proximal points. The algorithm's pseudo-code flow is detailed in Appendix A.

$$f = \min \sum_{i=1}^{n} D_{c_i c} \tag{7}$$

$$D_{c_i c} = \sqrt{(x_i - x)^2 + (y_i - y)^2} \tag{8}$$

Where $D_{c_i c}$ is the distance from random point coordinates to the target point, m; c_i is the position of the discrete point; c is the target position; x_i, y_i are the coordinates of the discrete point; x, y are the coordinates of the target point.

2.2.3 Allocation of unmanned aerial vehicle operational sorties

Prior assessment of pest severity or crop fertilizer requirements allows for the pre-setting of the UAV's spraying flow rate per hectare F, operating speed v, and width d. The operational area formula is presented in Equation 9, while the UAV payload formula is detailed in Equation 10.

$$\begin{cases} R = \sum_{1}^{n} R_1 + R_2 + \ldots R_n \\ S = Rd \end{cases} \tag{9}$$

$$L_j = \frac{S_n - S_{n-1}}{dv} F \tag{10}$$

Prior assessment of pest severity or crop fertilizer requirements allows for the pre-setting of the UAV's spraying flow rate per hectare F, operating speed v, and width d. The operational area formula is presented in Equation 9, while the UAV payload formula is detailed in Equation 10.

Where, R is the total range of plant protection operations on fields, m; n is the number of UAV sorties, sorties; S is the total area of UAV operations, hm^2; S_n is the area covered in the nth sortie, hm^2; S_{n-1} is the area covered in the $(n-1)$th sortie, hm^2; L_j is the payload of the jth sortie kg; F is the flow rate of spraying L/min.

Utilizing the optimal route from the coverage algorithm, the workload for each plant protection area and the requisite load per sortie are computed, with quantitative loading guided by Equation 10.

Given that the operational area may not be an exact multiple of a single sortie's area, pre-allocation of the load and strategic arrangement of racks and payload are essential to conserve energy and minimize sorties. Equation 11 outlines two methods for rack allocation.

$$j = \begin{cases} 1, S_m < S_d \\ \text{int}\left(\frac{S_m}{S_d}\right) + 1, \frac{S_m}{S_d} \notin Z, S_m \geq S_d \end{cases} \tag{11}$$

Where S_d is the maximum area covered in a single UAV sortie in hectares, hm^2; S_m is the cumulative area of plant protection operations within the operational area, hm^2; Z is a positive integer.

As stipulated in Equation 11, there are two scenarios for UAV sortie allocation:

1. If the operational area is less than the maximum coverage of a single sortie, the UAV does not require recharging.
2. If the operational area exceeds the maximum coverage, the UAV must recharge at least once. The number of flights is the quotient of the operational area and the maximum single sortie area, rounded to the nearest whole number. Typically, the number of flights ranges from $1 < j < \text{int}\left(\frac{S_m}{S_d}\right) + 1$.
3. The algorithm's pseudo-code flow is provided in Appendix B.

2.3 Design of route planning algorithm for multi-tea field scheduling

2.3.1 Optimization of genetic algorithms

To effectively encode the characteristic points of tea fields, this study utilizes integer coding (Asim and Abd El-Latif, 2023; Yan et al., 2024). Within genetic algorithms, the fitness function gauges an individual's environmental adaptability, with higher values denoting superior individuals, that is, shorter routes. Here, we consider a coded chromosome represented by $| k_1 | k_2 | \ldots | k_i | \ldots | k_n |$, with its fitness function defined in Equation 12.

$$f_n = \frac{1}{\sum_{i=1}^{n} D_{k_i k_j}} \tag{12}$$

Where the $D_{k_i k_j}$ is the distance from operational field i to operational field j, m; f_n is the reciprocal of the distance to the start vertex after completing the circuit.

To enhance the genetic algorithm's search efficiency across various evolutionary stages, this study introduces the Hyperbolic Tangent Mapping Crossover and Fitness-Inverse Adjusted Mutation, predicated on fitness values, as delineated in Equations 13 and 14:

$$P_c' = \begin{cases} \frac{1}{2}\left(1 + \tanh\left(\frac{2f' - f_{\min} - f_{\max}}{f_{\max} - f_{\min}}\right)\right), & f' \geqslant f_{avg} \\ 1, & f' < f_{avg} \end{cases} \tag{13}$$

$$P_m = \begin{cases} \frac{k_2(f - f_{\min})}{f_{\max} - f_{avg}}, & f \geq f_{avg} \\ \frac{k_3}{f}, & f < f_{avg} \end{cases} \quad (14)$$

Where f' is the mean fitness value of the individuals undergoing crossover; $f_{\min}$, $f_{\max}$ are the minimum and maximum fitness value respectively; k_1 is a constant set to 0.5; f is the variant individual's fitness; k_2, k_3 are constants; $k_2 \in [0,0.001]$; k_3=0.2.

In the early stages of the algorithm, exploration of new solution regions yields more varied results. To enhance optimization efficiency during this phase, we directly apply the crossover operation ($P_c' = 1$) to less adapted routes. Additionally, we set a fixed variance probability (k_3=0.2) inversely related to fitness. This encourages less adapted individuals to explore new solutions aggressively, preventing premature convergence to local optima. Furthermore, we prevent computationally expensive over-exploration by highly fit individuals within better candidate solutions.

As the algorithm progresses, the adaptability of route results improves in later stages. Our study aims to strike a balance between diversity (highly adapted individuals) and retention (better-adapted individuals). To achieve this, we introduce a hyperbolic tangent function with Steeper Gradient and Smoothness properties during the mid- and late algorithm stages. This function controls the transition from high to low crossover probability. Additionally, we dynamically adjust the difference between an individual's maximum fitness and average fitness, fine-tuning it to the mutation probability using a small interval ($k_2 \in [0, 0.001]$). Our designed crossover and mutation operators adapt the population's fitness distribution across different periods, promoting efficient convergence to the optimal solution.

2.3.2 Optimization of simulated annealing algorithms

In the standard simulated annealing approach, a new solution is accepted if it results in a decrease in system energy; otherwise, its acceptance is determined by a predefined probability (Li et al., 2022a; Sajid et al., 2022). This acceptance criterion dictates the transition probability between the current and new solutions, impacting the algorithm's optimization performance and convergence velocity. Accordingly, this paper proposes an optimized Metropolis criterion, formulated in Equations 15 and 16.

$$P = \begin{cases} 1, & E_{t+1} < E_t \\ p = e^{-\frac{E_{t+1} - E_t}{T_m T}}, & E_{t+1} \geq E_t \end{cases} \quad (15)$$

$$T_m = e^{-\frac{S(x_{\max}) - S(x_{\min})}{T_0}} \quad (16)$$

Where E_{t+1}, E_t are the new and preceding solutions, respectively; T_m is the temperature adjustment coefficient; T is the current temperature; $S(x_{\max})$, $S(x_{\min})$ are the maximum and minimum values of the objective function corresponding to the N feasible solutions randomly selected from the top 20% of the pre-algorithmic solution set, respectively; T_0 is the initial temperature.

If the solution set derived from the genetic algorithm serves as the initial solution for the simulated annealing algorithm, the discrepancy between E_{t+1} and E_t may not correspond to the current temperature T, potentially leading to an impractical selection probability under the Metropolis criterion. This discrepancy can influence the algorithm's optimization efficacy and computational duration. The dynamic genetic algorithm introduced in this study adjusts the Metropolis criterion with T_m by equating the extreme difference $S(x_{\max}) - S(x_{\min})$ of the feasible solutions within the top 20% of the solution set to T_0, enabling the Metropolis criterion to adapt to solution space fluctuations and maintain the stability of probability P.

In the adaptive Metropolis mechanism, the cooling rate must be modulated based on the proportion ΔN of suboptimal solutions accepted in the inner loop. Equation 17 demonstrates that ΔN assesses the appropriateness of the cooling rate; a value approaching 0 implies excessive rapidity necessitating a decrease, whereas a value nearing 1 suggests insufficient speed, requiring an increase to enhance algorithmic efficiency.

$$\Delta N = \frac{N_2}{N_1} \quad (17)$$

Where ΔN is the cooling rate; N_2 is the count of inferior solutions accepted within N_1 iterations; N_1 is the total number of instances where $E_{t+1} \geq E_t$ at temperature T.

Given the complexity of the UAV plant protection scheduling problem with multiple traveling salesman problem studied in this paper, an efficient parallel algorithm is needed to solve the traveling salesman problem in each partition. Thus, this paper introduces an adaptive cooling function that amalgamates logarithmic and exponential functions to define the temperature function, as depicted in Equations 18 and 19.

$$T_k = \frac{T_s}{\log(k)}, k \leq m \quad (18)$$

$$T_k = \begin{cases} T_s * 0.99^k, & k > m, \Delta N \in (0.75, 1) \\ \frac{T_s}{\log(k)}, & k > m, \Delta N \in (0, 0.25) \end{cases} \quad (19)$$

Where m is a predetermined constant; T_s is the set initial temperature; k is the current number of iterations.

As shown in Figure 4, the intersection of the logarithmic and exponential functions determines m, with the derivation of the logarithmic function yielding m = 161.8473. Initially, the algorithm employs the logarithmic function for swift cooling to approximate the optimal solution's vicinity. Subsequently, the cooling magnitude is dynamically adjusted based on the acceptance rate of suboptimal solutions ΔN, with real-time modulation of the cooling pace achieved through alternating use of the exponential and logarithmic functions.

2.4 Design of the algorithmic fusion approach

Figure 5 illustrates the optimization search processes of the genetic algorithm (GA) and the simulated annealing algorithm (SA), likened to a mountain climbing strategy. In GA, an individual's position, indicated by a solid arrow, often converges

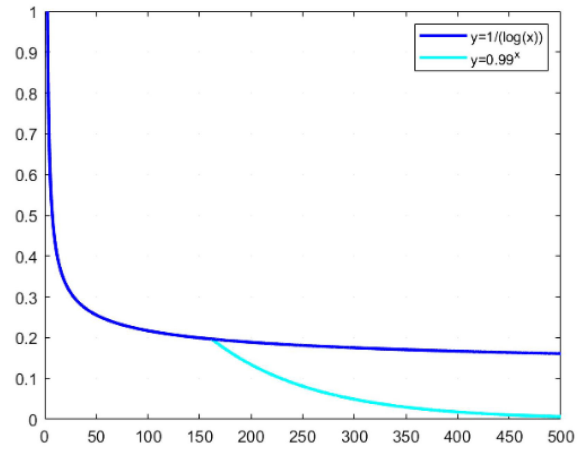


FIGURE 4
Optimized cooling curve intervals for the simulated annealing algorithm.

prematurely to a local optimum. This paper integrates the concept of simulated annealing into GA, as depicted by dashed arrows, which promotes the pursuit of the global optimum by probabilistically pushing GA beyond the local minima through the jump mutation property of adaptive annealing itself (Marjit et al., 2023; Ozkan, 2021).

Upon examining the optimization trajectories of GA and SA, it becomes evident that they offer complementary advantages in exploring the solution space's breadth and depth. This synergy suggests the potential for a hybrid approach, leveraging GA's prowess in global search and SA's finesse in local refinement (Santé et al., 2016). The dynamic crossover and mutation mechanisms in GA promote robust, undirected population evolution, while SA's targeted filtering fine-tunes the search direction. This study proposes amalgamating this enhanced algorithmic framework with clustering techniques and supply point strategies to devise a proficient planning method tailored to the multiple traveling salesman problem with unmanned aerial vehicle (MTSPU) in multi-tea field scenarios. The workflow of the ACHAGA fusion algorithm is presented in Figure 6.

2.5 Design of the experiment

In this experiment, the DJI T20p plant protection UAV served as a model to simulate and evaluate its operational efficiency within a tea cultivation region in Dadugang Township, Jinghong City, Xishuangbanna Prefecture, Yunnan Province, China. As depicted in Figure 7A, the selected tea cultivation area spans from 100°43' to 101°12' east longitude and 22°30' north latitude, encompassing 200 tea fields with a cumulative area of 28.2 hectares (hm²), averaging 0.3 hm² per field. The T20p's operational efficiency, as per official specifications, is 1.67 hm² per flight. For this analysis, the tea regions were digitally rendered using OWI 3D mapping software (Version 9.1.6 X64), illustrated in Figure 7B, where the blue bold line demarcates the tea field boundaries, and the green shaded regions represent the fields themselves.

2.5.1 Adaptive clustering hyperbolic annealing genetic algorithm for multi-field plant protection operation scheduling route simulation experiments

1. This study aims to evaluate the enhancements made to the Hyperbolic Genetic Algorithm (HGA) and the Adaptive Simulated Annealing (ASA), and to assess the performance of the newly developed Adaptive Clustering Hyperbolic Annealing Genetic Algorithm (ACHAGA). To achieve this, we conducted a comparative evaluation via computer simulation, analyzing the search accuracy and optimization capabilities of each algorithm iteration. The simulations were performed on a test PC with an Intel (R) Core(TM) i7-6700HQ CPU @ 2.60GHz, 8 GB RAM, running on a Windows 10 platform using Matlab-2018a, focusing on the multi-tea field problem model. The selected hardware and software specifications ensure a balance between computational efficiency and accessibility, providing a realistic benchmark for practical

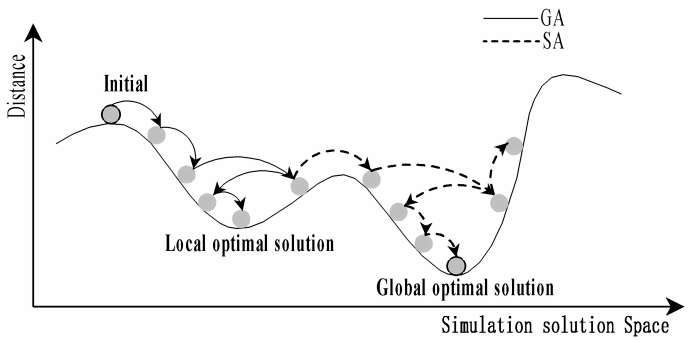


FIGURE 5
Schematic representation of the solution space for Genetic Algorithm (GA) and Simulated Annealing (SA).

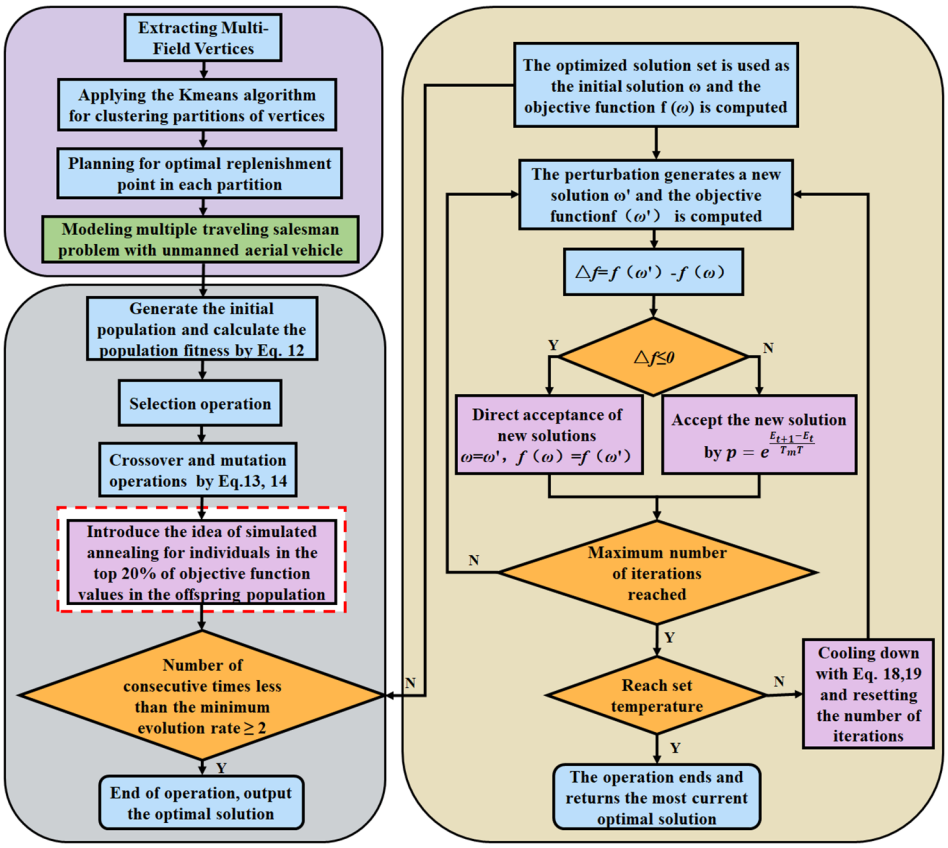


FIGURE 6
Search flowchart of Adaptive Clustering Hyperbolic Annealing Genetic Algorithm (ACHAGA).

applications. For consistency, the maximum number of iterations was set at 200 and the population size at 50. These parameters were chosen based on preliminary tests that indicated they provide a good balance between computational time and optimization performance. A higher number of iterations allows the algorithm to explore the solution space more thoroughly, increasing the likelihood of finding a global optimum. Meanwhile, a population size of 50 ensures sufficient genetic diversity without overwhelming computational resources. Additional algorithm parameters are detailed in Table 1, which includes specifics on mutation rates, crossover probabilities, and the cooling coefficient for the ASA component. These parameters were fine-tuned through a series of preliminary experiments to ensure optimal performance of ACHAGA in solving the multi-tea field problem. The software specifications are listed in Table 2.

2. To evaluate the optimization efficacy of the proposed ACHAGA on the multi-tea field problem model, it was benchmarked against established swarm intelligence algorithms, namely the Brainstorming Algorithm (BSO) and the Artificial Fish Swarm Algorithm (AFSA) (Qiu et al., 2015; Neshat et al., 2014).

(a) Modeled Distribution of Multi-Tea Fields **(b)** Aerial Diagram of Multi-Tea Field Distribution

FIGURE 7
Schematic of the model distribution for the multi-tea field problem. **(A)** Modeled Distribution of Multi-Tea Fields. **(B)** Aerial Diagram of Multi-Tea Field Distribution.

TABLE 1 Various algorithm parameters.

Algorithm	Parameters
AFSA	Population size: n=100, Maximum number of iterations: Max_gen=200, Maximum number of attempts: trynumber=500, Perception distance: Visual=16, Crowding factor: deta=0.8
BSO	Population size: n=100, Number of clusters: cluster_num=x, Maximum number of iterations: Max_gen=200, Probability of selecting one cluster: p_one=0.5, Probability of selecting two clusters: p_two=1-p_one
GA-ACO	Population size: n=100, Maximum number of iterations: Max_gen=200, Crossover probability: Pc=0.8, Mutation probability: Pm=0.2, Importance of pheromone: Alpha=1, Importance of heuristic factor: Beta=5, Pheromone enhancement coefficient: Rho=0.1, Pheromone enhancement coefficient: Q=100
ACHAGA	Population size: n=100, Crossover probability: as in Equation 13, Mutation probability: as in Equation 14, Maximum number of iterations: Max_gen=200, Initial temperature: T0 = 100, Final temperature: Tend=1e-8, Selection probability: as in Equation 15, Cooling coefficient: alpha as in Equations 18, 19
GA	Population size: n=100, Crossover probability: Pc=0.8, Mutation probability: Pm=0.2, Maximum number of iterations: Max_gen=200

Each algorithm underwent 20 trials, maintaining identical iteration and population constraints as specified in Table 1, with the mean outcome representing the final result. As illustrated in Figures 8A, B, two distinct test scenarios were selected: a self-constructed 200-tea field problem model and the eil51 problem from the TSPLIB standard database, featuring uniformly distributed nodes. The eil51 nodes were presumed to be ideal tea field vertices, averaging 0.3 hectares (hm^2) per vertex, totaling 15.3 hm^2.

To optimize the preprocessing clustering division scheme for simulation trials of the multi-tea field problem model and the eil51 problem, it is necessary to determine the optimal silhouette coefficient distributions across various UAV sortie ranges, depicted in Figures 9A, B. The UAV sortie range is calculable from Equation 11, considering the UAV's working efficiency (S_d) and the working area (S_m) for each problem, as detailed in Section 2.5. The silhouette coefficient distributions for the refined multi-tea field problem model and the eil51 problem fall within the ranges of (1, 17] and (1, 10], respectively.

TABLE 2 Test platform and software version.

Test platforms	Platform configuration parameters
CPU	Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz, 2.59 GHz
GPU	GTX 1060 6GB
RAM	8GB
Operating system version	Windows-10
Software version	Matlab-2018(a)

Analysis of Figures 9A, B indicates that the silhouette coefficient for both problem types peak at k = 5 and k = 3. Consequently, the k-means algorithm was utilized to segment the two problem types into 5 and 3 plant protection work areas, respectively, as shown in Figures 10A, B. In these figures, dots represent vertices within the tea fields, colors denote work zones, and stars symbolize supply points.

3. To evaluate the effectiveness of the ACHAGA algorithm on the multi-tea field problem model and the eil51 problem, a Wilcoxon rank-sum test was conducted to perform a non-parametric statistical comparison with four other related algorithms (Kochengin et al., 2019). A significance level of α= 5% was established; a p-value less than 5% signifies a statistically significant difference, while a value greater than or equal to this threshold indicates no significant difference. To provide a more rigorous statistical analysis, confidence intervals were calculated to estimate the precision of the observed effect sizes. The effect sizes were also reported to quantify the magnitude of the differences between the ACHAGA algorithm and the other algorithms.

2.5.2 Multi-tea field plant protection site test based on adaptive clustering hyperbolic annealing genetic algorithm

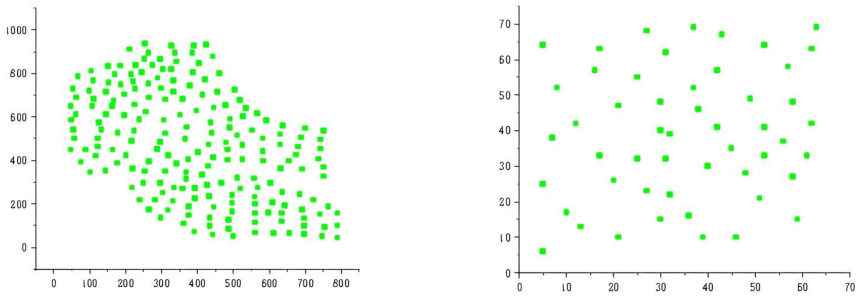
This study introduces an algorithm designed for path planning across multi-tea field in hilly, mountainous terrains. To validate its feasibility and accuracy, the algorithm was evaluated based on four criteria: operational voyage, coverage, excess coverage, and dispatch path length. Additionally, a field trial was conducted to compare the actual operating conditions with simulation results and to assess the experimental design's soundness. Drawing on the methodologies of scholars both domestic and international (Tian et al., 2023; Li et al., 2022c), the study utilized a DJI Phantom 4 Pro UAV, boasting a top horizontal velocity of 72 km/h and a maximum endurance of approximately 30 minutes. It is equipped with a 42° tilt capability and a satellite positioning system supporting GPS/GLONASS dual-mode for superior navigational precision. The route planning was facilitated by the RockyCapture flight control system, which offers features such as free planning, one-key control, and real-time tracking, as illustrated in Figures 11A, B.

2.5.2.1 Trial Site Conditions

The trial was conducted from February 16 to 21, 2023, at the Dashanmiao Tea Farm in Shitang Town, Feidong County, Hefei City, Anhui Province, China, featuring a tea plantation with an approximate 7° slope. The weather conditions during the test period included sunny skies, an average temperature of 17°C, relative humidity of 38%, and wind speeds ranging from 1 to 2 on the Beaufort scale. Given the UAV's operational range limitations, the selected test area comprised 15 fields with an average area of 0.12 hectares (hm^2), detailed in Figure 12, with the geographical coordinates provided in Table 3.

2.5.2.2 Field trial program

The test flowchart, depicted in Figure 13, outlines the specific test procedures:



(a) Vertex Distribution of Multi-tea Field Problem Model

(b) Vertex Distribution of the Eil51 Problem Model

FIGURE 8
Problem objects of the optimization performance test. **(A)** Vertex Distribution of Multi-tea Field Problem Model. **(B)** Vertex Distribution of the Eil51 Problem Model.

Step 1: Utilizing OWI Interactive Map Software (Version 9.1.6 X64), the test area's coordinates, boundaries, and elevation data were acquired to establish a planar right-angle coordinate system. Subsequently, the k-means algorithm and Mercator inverse formula were applied to determine and label the test site's supply points on the map. Each algorithm was then used to devise a corresponding route scheme, culminating in the export of a KML file of the planned route.

Step 2: The ground station flight control platform was accessed to import the route by selecting the import function, locating the recently exported KML file, and integrating it into the platform. The route's accuracy was verified, parameters and settings were adjusted, and preparations were made for UAV flight control.

Step 3: The flight control platform was employed for testing: the UAV was connected to the ground station flight control platform, initiated, and navigated according to the imported route. The UAV's status and data were monitored throughout the flight, with test outcomes and issues being meticulously recorded.

2.5.2.3 Pre-Treatment Clustering Division of the Trial Area

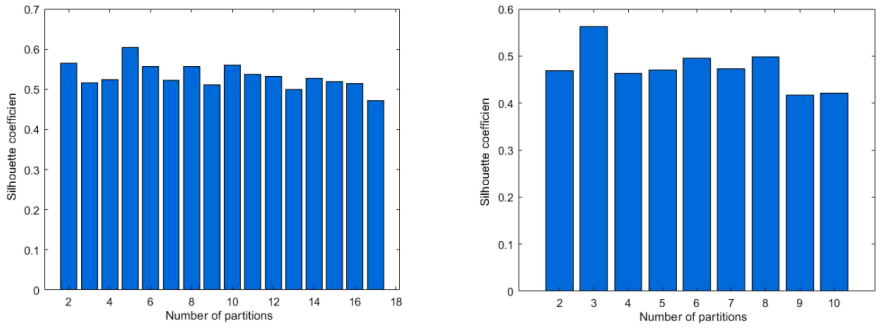
To enhance the clustering partitioning effect, the study introduced a contour coefficient calculation method prior to partitioning. This yielded a contour coefficient comparison chart for varying cluster numbers k, as illustrated in Figures 14A, B.

The analysis of Figure 14A reveals that the contour coefficient is maximized when k=2. Consequently, the study opted to segment the multi-tea field into two task areas, achieving the results displayed in Figure 14B through clustering division.

2.5.2.4 Trial Methods and Data Calculation

To assess the algorithm's performance and feasibility, site experiments were conducted for both the test group (BSO) and the control group (ACHAGA). The test group employed manual empirical planning and the brainstorming algorithm, while the control group utilized the clustering division planning algorithm proposed in this study for site operation scheduling route planning. Table 1 presents the theoretical parameters of the algorithms. Accounting for navigation errors and wind speed effects, five field trials were executed for each algorithm, selecting the optimal voyages and routes for comparative analysis. Subsequently, the optimal route data were derived and subjected to computational analysis.

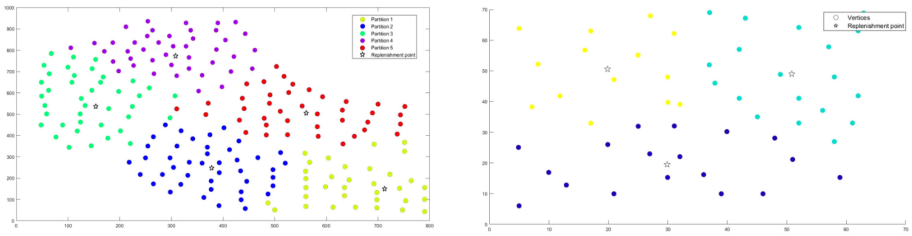
For evaluating the accuracy of full-coverage routes, the index of excess coverage was employed. The excess coverage rate is calculated using Equation 20, where the product of the total operating range L and the operating width B constitutes the operating area. In full-coverage



(a) Distribution of Silhouette Coefficient for the Multi-tea Field Problem Model

(b) Distribution of Silhouette Coefficient for the Eil51 Problem

FIGURE 9
Comparative histogram of silhouette coefficient for each problem. **(A)** Distribution of Silhouette Coefficient for the Multi-tea Field Problem Model. **(B)** Distribution of Silhouette Coefficient for the Eil51 Problem.



(a) Clustering Distributions for Multi-tea Field Problem Model **(b)** Clustering Distributions for the Eil51 Problem

FIGURE 10
Clustering Distributions for Multi-tea Field Problem Model and Eil51 Problem. **(A)** Clustering Distributions for Multi-tea Field Problem Model. **(B)** Clustering Distributions for the Eil51 Problem.

operations, a smaller operating area correlates with a lower excess coverage rate, indicating greater accuracy of the coverage area.

$$\eta = \left(\left|\frac{LB}{S_0} - 1\right|\right) \times 100\,\% \tag{20}$$

Where η is excess coverage; L is the total operational range, m.

3 Results

3.1 Multi-tea field plant protection site trials using adaptive clustering hyperbolic annealing genetic algorithm

The field trial outcomes are presented in Tables 4 and 5. The black line in the planning route column (left side of the table) represents the full-coverage operation route at 2.5-meter intervals, with an operational width set at 5 meters. The pink line indicates the dispatch route, and the 'H' point marks the return to the supply point. In the optimal flight path column (left side of the table), the red line signifies the full-coverage operation route, the yellow line the scheduling route, and the coordinates for the two supply points in the south and north are determined using the inverse Mercator projection formula as (117°39′49.31″, 31°51′2.66″) and (117°39′51.31″, 31°51′6.70″), corresponding to the circular charge markers in the north and south, respectively.

4 Discussion

4.1 Convergence curve performance test of hyperbolic annealing genetic algorithm for adaptive clustering

4.1.1 Robustness and optimization capability analysis of the proposed algorithm's iterative curve

The analysis of iteration curves in Figures 15A, B demonstrates that both the Hyperbolic Annealing Genetic Algorithm (HGA) and the Genetic Algorithm (GA) exhibit superior optimization efficiency in the initial phase. HGA presents smoother iteration curves and achieves a more refined global optimal solution. This suggests that early direct crossover operations and optimization strategies, such as hyperbolic tangent mapping crossover and fitness-inverse adjusted mutation, expedite the removal of low-fit individuals, curtail ineffective searches, and enhance HGA's efficiency and stability, thereby broadening its optimization search space. The Adaptive Simulated Annealing (ASA) and Simulated Annealing (SA) algorithms display greater overall optimization

(a) Test site **(b)** Rockycapture Ground Station Platform for Flight Control

FIGURE 11
Field validation tests. **(A)** Test site. **(B)** Rockycapture Ground Station Platform for Flight Control.

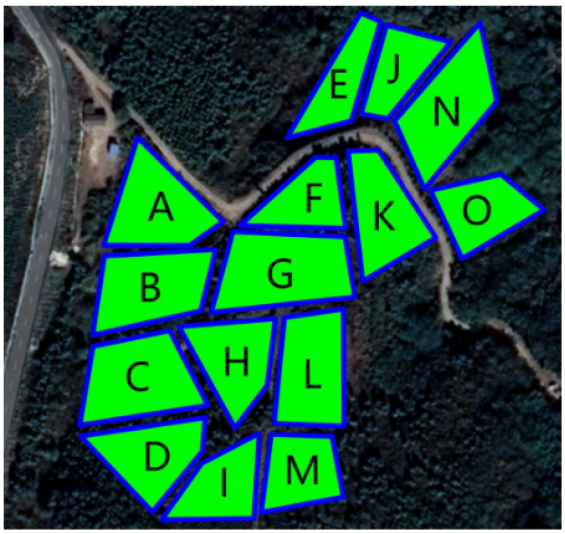


FIGURE 12
Diagram of the test site.

capability with minimal fluctuations in the optimal solution. The iteration curves for ACHAGA and GASA are characterized by smoothness and stability, with an observable "discontinuity" in the early stages and a transition from rapid to gradual convergence within the first thirty iterations. This indicates that the ACHAGA algorithm, which integrates HGA and ASA, significantly improves robustness, stability, and optimization efficiency. Furthermore, the intersection of dark blue and green iteration curves in Figures 15A, B reveals overlapping segments during the iterative process, particularly between HGA and ASA post-improvement (15-165 iterations). This overlap effectively balances the performance of both algorithms, mitigating the impact of individual algorithm performance and substantiating the efficacy of incorporating hyperbolic genetic algorithm concepts into adaptive annealing, thereby leveraging the strengths of both approaches.

TABLE 3 Longitudinal and latitudinal coordinates of field plots in the test area.

Field	Field apex coordinates	Area (m^2)	Projected perimeter (m)
A	117°39′47.56″; 31°51′5.65″	1400	161.21
B	117°39′47.50″; 31°51′4.51″	1500	164.29
C	117°39′47.27″; 31°51′3.17″	1600	167.63
D	117°39′47.50″; 31°51′2.06″	1200	155.06
E	117°39′50.75″; 31°51′7.57″	900	152.08
F	117°39′50.22″, 31°51′5.77″	800	134.76
G	117°39′49.73″; 31°51′4.75″	1800	181.25
H	117°39′49.00″; 31°51′3.32″	1000	136.84
I	117°39′48.69″; 31°51′1.54″	900	140.34
J	117°39′51.66″; 31°51′7.61″	800	128.29
K	117°39′51.42″, 31°51′5.62″	1300	157.69
L	117°39′50.20″; 31°51′3.22″	1400	157.75
M	117°39′50.04″; 31°51′1.83″	900	127.84
N	117°39′52.56″; 31°51′7.11″	1600	179.24
O	117°39′53.18″; 31°51′5.63″	1000	133.38

4.1.2 Optimization accuracy and stability analysis of the proposed algorithm

The analysis of iteration curves for each partition, as shown in Tables 6 and 7, reveals that while all five algorithms successfully generate flight routes covering all fields, they exhibit significant variance in search accuracy, convergence speed, and iteration performance. The algorithm introduced in this paper demonstrates a pronounced ability to converge to the optimal solution within an average of 40 iterations for the multi-tea field problem model. Notably, the ACHAGA algorithm's average number of iterations and optimal route lengths are reduced by 113.32 iterations and 1811.93 meters, respectively, compared to other biomimetic algorithms. This efficiency indicates that ACHAGA can achieve superior solutions with fewer iterations. Further examination of Table 8 indicates that the mean optimal solution lengths for ACHAGA are 9693.27 meters for the multi-tea field problem model and 461.26 meters for the standard eil51 problem. These results are significantly lower than those achieved by the other four biomimetic algorithms. For the multi-tea field problem model, ACHAGA's results show substantial reductions of 4904.82, 926.07, 3803.96, and 800.11 meters compared to the optimal solutions of GA, Genetic ACO Fusion Algorithm, AFSA, and BSO, underscoring ACHAGA's considerable advantage in model optimization accuracy. Additionally, ACHAGA exhibits the smallest polar deviation and coefficient of variation in data for both the multi-tea field problem model and the standard eil51 problem, with values of 56.29 and 4.56, and coefficients of variation of 0.00406241 and 0.00326200, respectively. These metrics suggest a low dispersion in test results, affirming that the outcomes are not due to random chance and that the data's reliability is robust. The narrowest confidence interval for ACHAGA not only suggests more reliable results compared to other algorithms but also confirms its stability in delivering consistent outcomes.

4.1.3 Disparities in algorithmic optimization performance across various problem complexities

Analysis of Table 8 indicates minor discrepancies in the simulation outcomes for the standard eil51 problem across all algorithms. However, when addressing the multi-tea field problem model with expanded solution spaces, AFSA and GA, which typically yield superior results, exhibit a marked decline in performance. Conversely, other algorithms demonstrate less variability in solution quality across different problems but still show significant divergence from ACHAGA's outcomes. This variation is likely attributable to the distinct mechanisms and coding methodologies of each algorithm, which may limit their ability to derive reasonable results within a finite number of iterations. The p-values in Table 9, being under 5%, substantiate the enhanced performance resulting from the optimization research conducted in this study. ACHAGA has

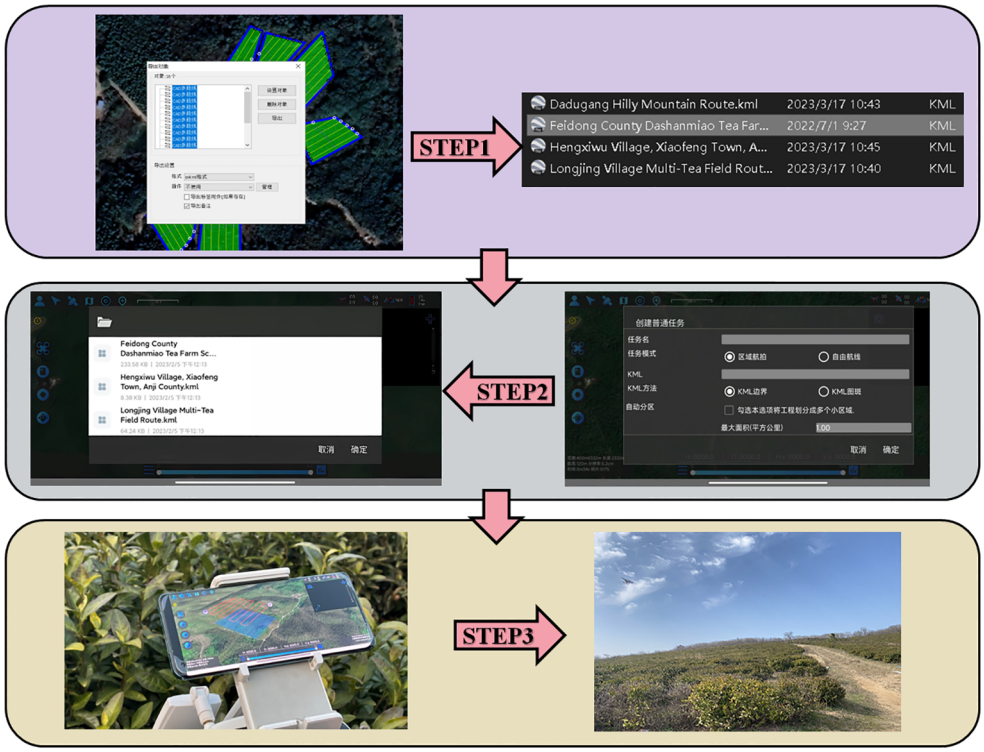
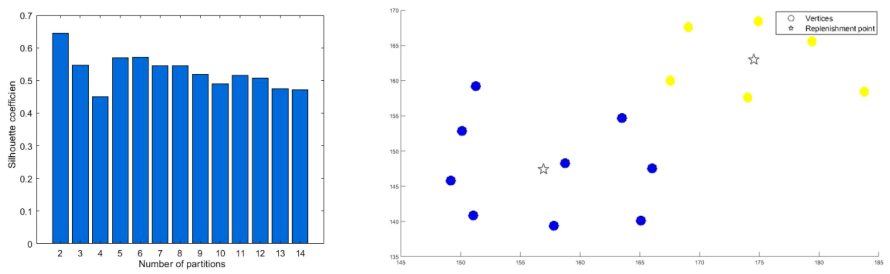


FIGURE 13
Importing the planned route KML file into the ground station operation process.

proven its efficacy in identifying optimal solutions for the problem models within 200 iterations, showcasing robust adaptability suitable for more intricate scenarios. The Pearson correlation coefficients in Table 9 provide insight into the relationship between the performance metrics of ACHAGA and the other algorithms. For the multi-tea field problem and eil51, the coefficients are generally low, with both positive and negative values, indicating weak correlations. This suggests that the performance patterns of ACHAGA are somewhat independent of those of the other algorithms. For eil51, the coefficient variance is relatively greater. The negative correlations with GA and GA-ACO suggest that as ACHAGA's performance improves, the performance of these algorithms tends to worsen, and vice versa. The results demonstrate that ACHAGA significantly outperforms the other algorithms across both problem instances. The statistical significance of the p-values confirms the robustness of ACHAGA's performance improvements. The Pearson correlation coefficients further illustrate the nature of these performance differences, highlighting the unique strengths of ACHAGA in addressing the complexities of both the multi-tea field problem and eil51.

4.1.4 Comparative analysis of the proposed algorithm with other algorithms

The performance analysis of ACHAGA indicates that it surpasses traditional heuristic algorithms in several key performance indicators.



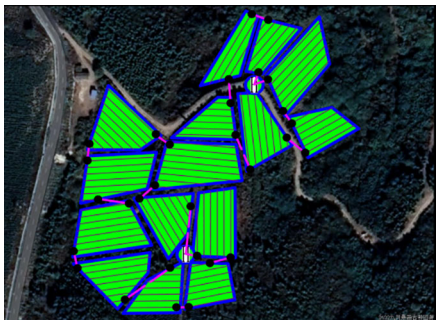
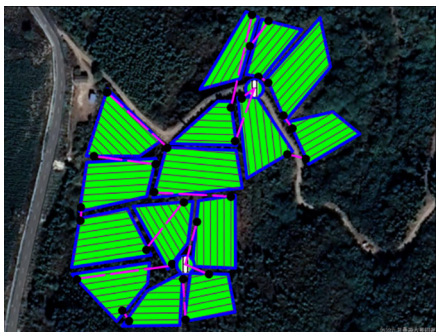
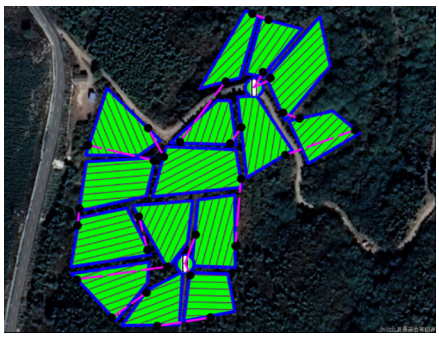
(a) Distribution of Silhouette Coefficient in the Test Area

(b) Partitioning Results based on Optimal Silhouette Coefficient

FIGURE 14
Clustering of the trial area. **(A)** Distribution of Silhouette Coefficient in the Test Area. **(B)** Partitioning Results based on Optimal Silhouette Coefficient.

10.3389/fpls.2024.1440234

TABLE 4 Comparison of various route planning methods and actual flight routes.

Methodology	Planning the optimal route range (m)	Actual flight range (m)
ACHAGA		
	3861.2	4385.2
BSO		
	4275.9	4744.8
AE		
	4358.3	5177.1

To understand why, we examine the principles and characteristics of various algorithms. GA, which relies on random mutations and crossover, often faces issues with slow convergence and premature convergence. ACHAGA addresses these issues with its hybrid GA-Simulated Annealing (SA) approach, combining GA's global search capabilities with SA's local precision for faster convergence. The Genetic ACO Fusion Algorithm, while enhancing exploration by integrating GA and Ant Colony Optimization (ACO), suffers from higher iteration counts and complex parameter control. ACHAGA's adaptive crossover and mutation rates dynamically adjust based on fitness, ensuring efficient convergence with fewer iterations. AFSA's performance is hindered by sensitivity to parameter settings and slower convergence in complex spaces. ACHAGA's adaptive Metropolis criterion, which incorporates temperature and solution space fluctuations, balances exploration and exploitation effectively, preventing premature convergence. BSO, which mimics brainstorming, generates diverse solutions but shows higher variance and slower convergence. ACHAGA's efficient cooling strategy, using logarithmic and exponential functions, enhances convergence speed and accuracy, avoiding local minima and ensuring robust optimization. These optimization features enable ACHAGA to achieve superior performance with shorter optimal solution lengths, fewer iterations, and higher stability, making it an efficient solution for complex optimization problems in a multi-tea field environment.

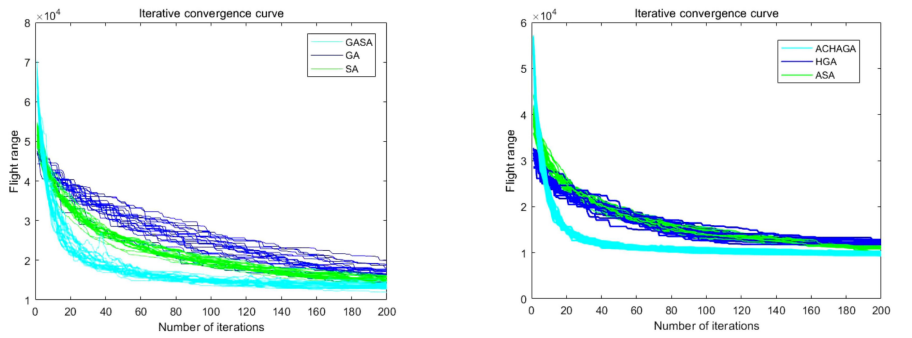
4.2 Field trial results analysis

Table 4 delineates the full-coverage and scheduling routes devised by three methodologies. The algorithm from this study yields more evenly distributed routes with reduced crossover,

TABLE 5 Comparison of optimal results from five planning routes by different methods.

Planning methodology	ACHAGA		AE		BSO	
Export Route	Planned Routes	Flight routes	Planned Routes	Flight routes	Planned Routes	Flight routes
Minimum dispatch range (m)	277.1	322.7	501.9	592.2	591.8	689.1
Number of U-turns (times)	83		94		83	
Percentage of non-operational time	–	8.40%	–	12.60%	–	15.10%
Rate of coverage	–	9.48%	–	13.85%	–	9.36%

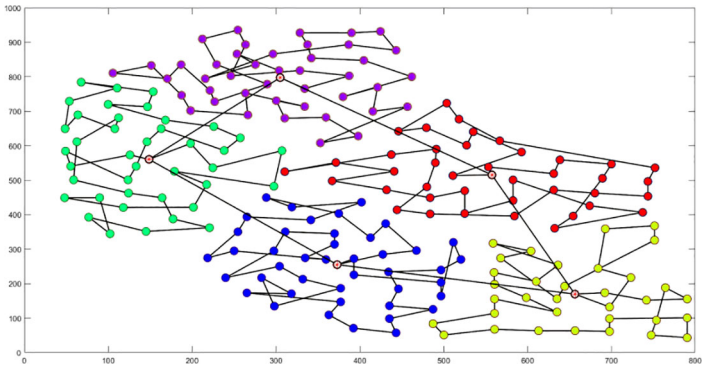
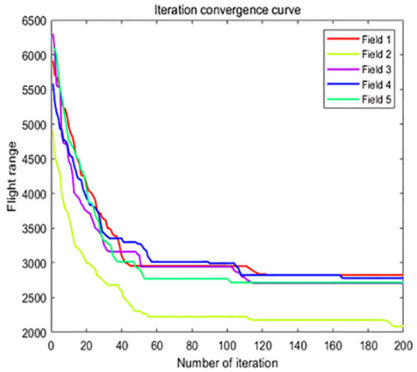
mitigating potential UAV interference. A comparison between the theoretical and actual flight paths reveals deviations due to factors such as wind speed, terrain gradient, and positioning stability, with the actual flight distance exceeding theoretical estimates. Table 5 enumerates the total real flight, operational, and dispatch ranges planned by the three methods, highlighting the superiority of this study's algorithm in these metrics and its ability to effectively minimize operational and dispatch distances. Specifically, compared to manual empirical and brainstorming methods, this algorithm decreases the actual flight's operational distance by 791.9 meters and 359.6 meters, respectively, and reduces the non-operational time share by 4.2% and 6.7%. Notably, the scheduling distance is nearly halved on average, suggesting that the clustering fusion planning algorithm substantially curtails UAVs' non-operational flights. Moreover, full-coverage planning utilizing heuristic algorithms reduces the operational distance by an average of 11.4% and the number of turns by 11, compared to manual empirical planning. This demonstrates its superior applicability for multi-terrain field route scheduling in the hilly, mountainous environment examined in this study.



(a) Iteration Curves for for GA, SA, GASA **(b)** Iteration Curves for HGA, ASA, ACHAGA

FIGURE 15
Comparison of convergence curves for each algorithm pre and post optimization Fusion. **(A)** Iteration Curves for GA, SA, GASA. **(B)** Iteration Curves for HGA, ASA, ACHAGA.

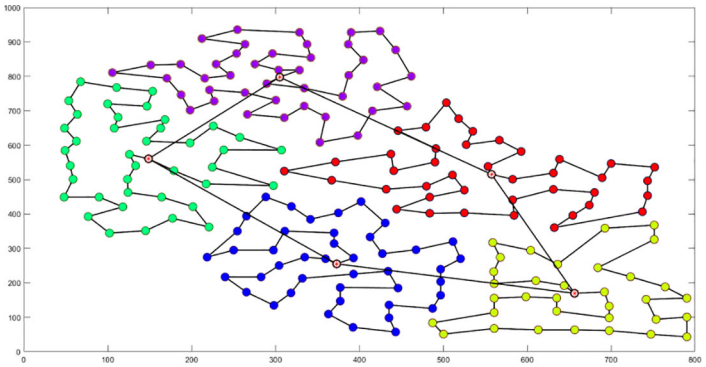
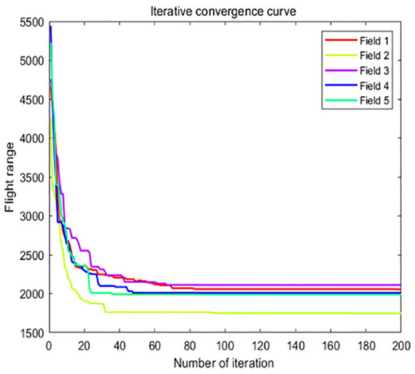
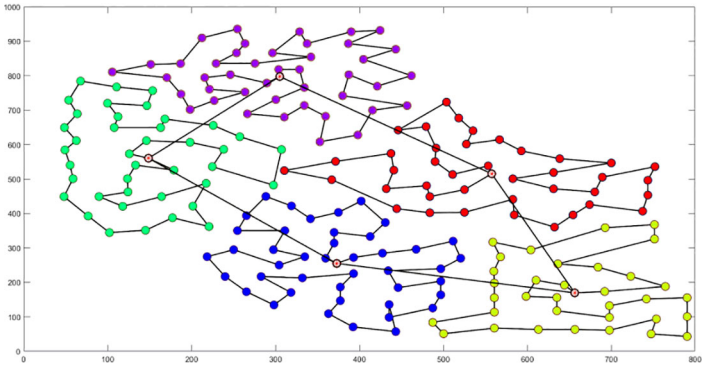
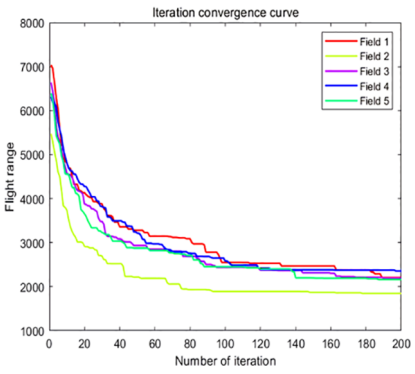
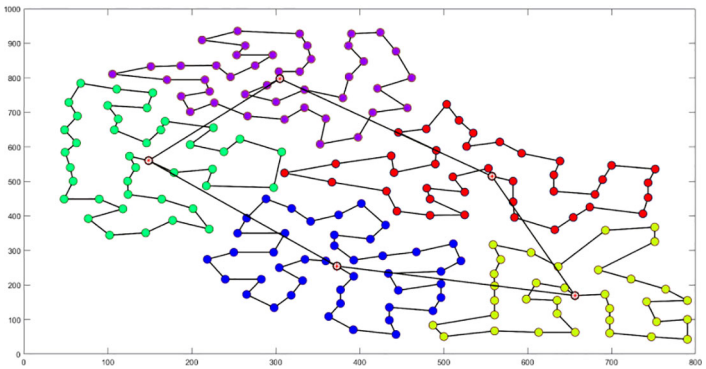
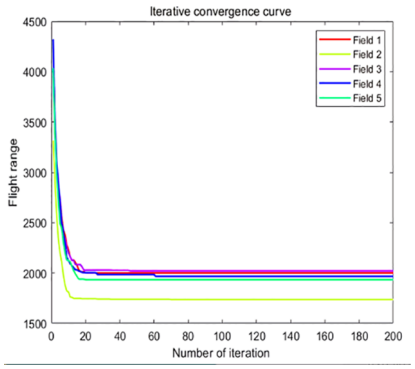
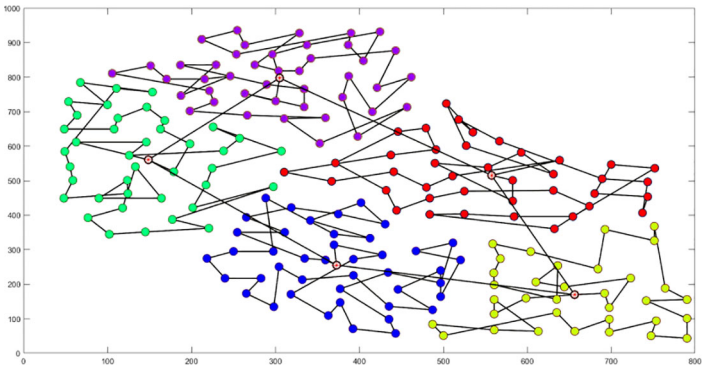
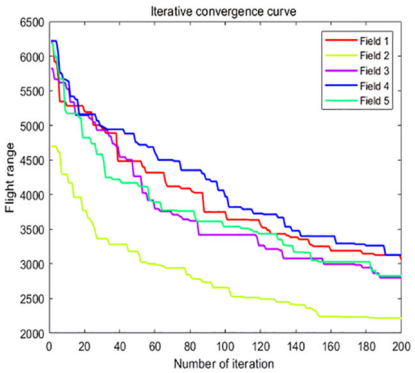
TABLE 6 Optimal routes and convergence curves comparison for clustering partitioning.

Algorithm	Optimal route range (m)	Average number of iterations
AFSA	 12480.2	 137.6

(Continued)

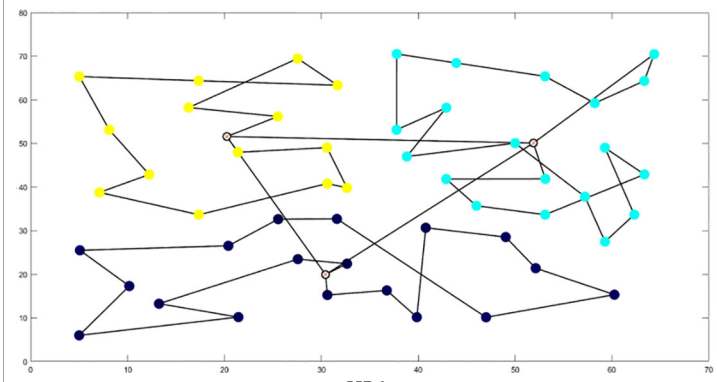
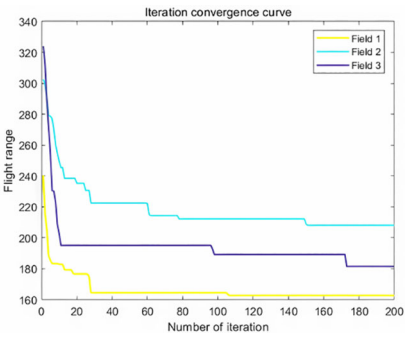
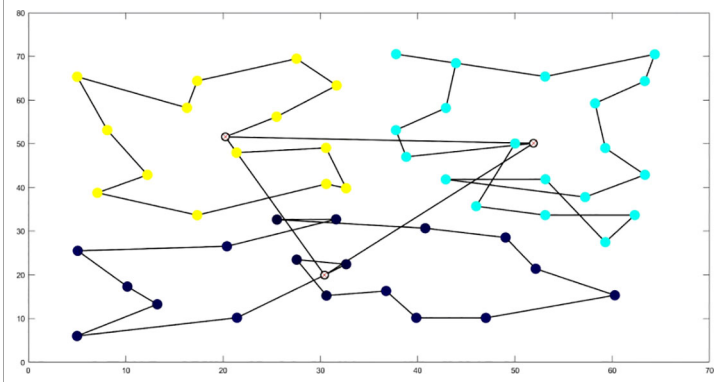
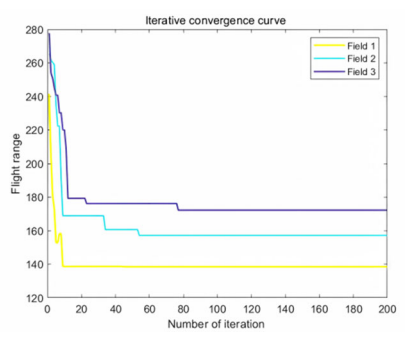
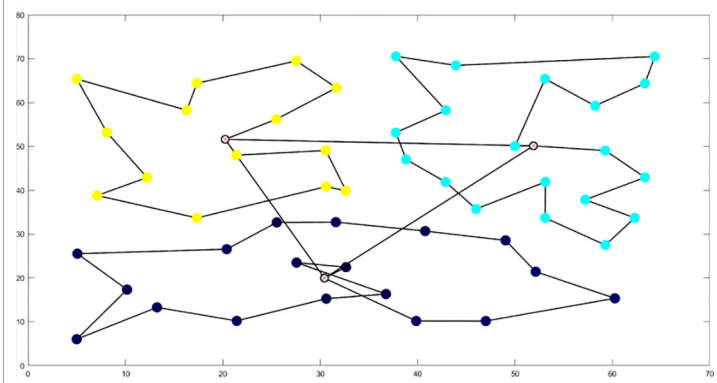
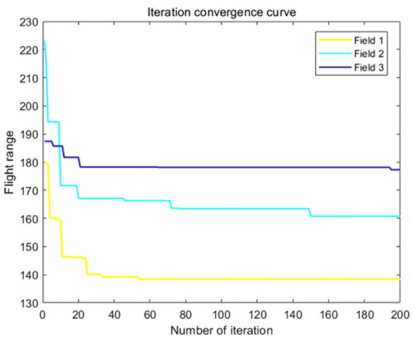
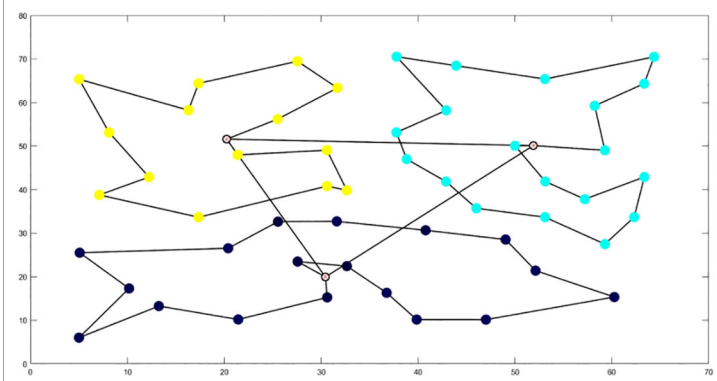
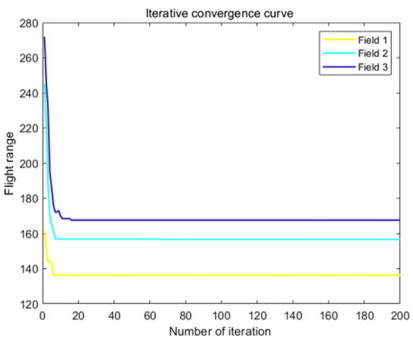
10.3389/fpls.2024.1440234

TABLE 6 Continued

Algorithm	Optimal route range (m)	Average number of iterations
BSO	 10035.9	 66.2
GA-ACO	 10256.2	 185.6
ACHAGA	 9659.7	 31.8
GA	 13114.2	 191.1

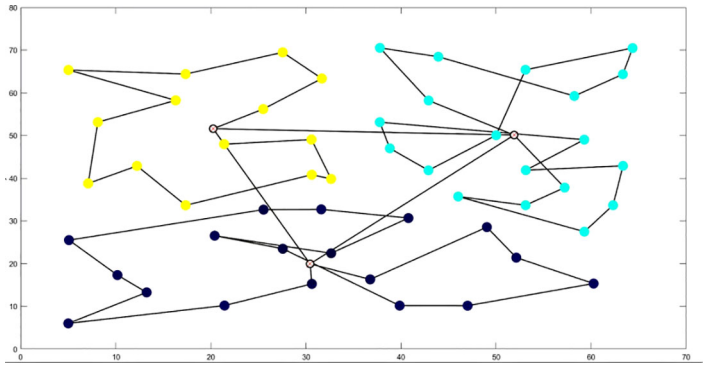
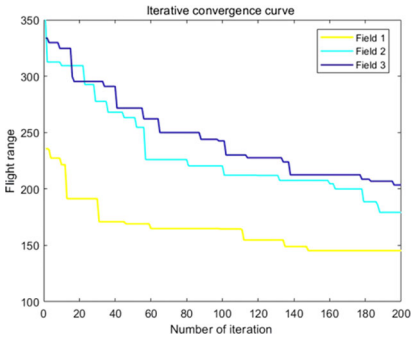
10.3389/fpls.2024.1440234

TABLE 7 Optimal routes and convergence curves comparison for various algorithms' clustering partitions.

Algorithm	Optimal route range (m)	Average number of iterations
AFSA	 557.6	 141.7
BSO	 541.6	 46.7
GA-ACO	 507.3	 133.4
ACHAGA	 460.5	 11.7

(Continued)

TABLE 7 Continued

Algorithm	Optimal route range (m)	Average number of iterations
GA	 533.7	 177.3

4.3 Research limitations

Despite the considerable advancements of the ACHAGA framework in UAV-based precision plant protection, several limitations warrant attention. Firstly, potential biases inherent in simulation studies may undermine the applicability of the findings to real-world scenarios, as the simulations did not account for unforeseen weather changes, equipment malfunctions, or human interventions. Secondly, the scope of field tests was geographically constrained and specific to certain types of tea fields, potentially limiting the generalizability of the results to other terrains and crop types. To address these limitations, we need to expand field tests to encompass a broader range of geographic regions and crop varieties to enhance the robustness of the algorithm's performance. Additionally, developing methods for dynamic resupply point management based on real-time data could further optimize UAV

TABLE 8 Simulation outcomes for different multi-objective problems.

Problem	Algorithm	Average Value(m)	Range(m)	CV	Confidence Interval
Multi-tea Field Problem Model	GA	14598.09	1819.64	0.02797023	[14390.09, 14806.09]
	GA-ACO	10619.34	777.67	0.01629405	[10534.79,10703.89]
	AFSA	13497.23	585.71	0.01340668	[13430.37,13564.086]
	BSO	10493.38	358.70	0.00930587	[10444.34, 10542.42]
	ACHAGA	9693.27	56.29	0.00406241	[9677.34, 9709.20]
Eil51	GA	563.66	32.62	0.01904130	[560.17, 567.15]
	GA-ACO	512.34	11.07	0.00754431	[510.40, 514.28]
	AFSA	577.61	29.86	0.01807257	[572.86, 582.36]
	BSO	550.43	15.53	0.01157147	[546.52, 554.34]
	ACHAGA	462.26	4.56	0.00326200	[461.67, 462.85]

TABLE 9 Comparison of significance level results between ACHAGA and each algorithm.

Comparison of algorithms	*p*-Value		*PCCs*	
	Multi-tea Field Problem	Eil51	Multi-tea Field Problem	Eil51
ACHAGA *vs.*GA	2.24E-37	1.47E-14	-0.015273	-0.513957
ACHAGA *vs.*GA-ACO	3.94E-24	1.67E-12	0.374056	-0.150358
ACHAGA *vs.*BSO	4.77E-15	3.26E-14	-0.275725	0.057610
ACHAGA *vs.*AFSA	3.09E-46	8.49E-17	-0.118178	0.386316

operations. Integrating advanced sensing technologies, such as LiDAR and multispectral imaging, could significantly improve terrain mapping and environmental monitoring, thereby refining route planning precision. Addressing these limitations through rigorous real-world testing and technological enhancements will be crucial for the broader adoption and efficacy of the ACHAGA framework in diverse agricultural contexts.

5 Conclusion

This study introduces the Adaptive Clustering Hyperbolic Annealing Genetic Algorithm (ACHAGA), a hybrid optimization algorithm that integrates the principles of hyperbolic genetic algorithms and simulated annealing. The primary objective is to enhance the precision of UAV-based plant protection scheduling in complex multi-terrain environments. The algorithm leverages efficient crossover and mutation operations inherent to hyperbolic genetic algorithms to achieve rapid convergence and maintain spatial diversity. High-quality solutions derived from this process serve as initial inputs for the simulated annealing algorithm, which, through its jump mutation properties and innovative temperature control mechanism, further refines solution quality and robustness. By employing cluster analysis, multi-tea field regions are segmented, and operational supply centers and flight schedules are meticulously designed to ensure UAVs adhere to the shortest feasible routes within their endurance limits in designated plant protection areas. ACHAGA demonstrates superior performance in both standard traveler's problem scenarios and multi-tea field applications, swiftly identifying optimal solutions and exhibiting robust global search capabilities and high stability. Future research should prioritize incorporating real-time environmental data, such as weather conditions, soil moisture levels, and pest distribution patterns, to refine route planning further. Evaluating the algorithm's scalability and adaptability across robustness crop types and terrains, and developing strategies to optimize UAV battery usage and recharge cycles for more sustainable operations, is also crucial. These research directions aim to build on the current study's foundation, advancing the capabilities of UAV-based precision agriculture. This work establishes a theoretical and technical foundation for the efficient management and operation of extensive tea plantations in hilly and mountainous regions, offering new perspectives on precision planning and management of UAV plant protection operations in complex settings. Beyond the immediate scope of this study, the findings have broader implications for the field of agricultural technology. The integration of real-time environmental data into UAV route planning could lead to more responsive and adaptive agricultural management systems, capable of adjusting to changing conditions such as sudden weather changes or pest outbreaks. This adaptability is crucial for improving crop yields and reducing losses, thereby supporting food security initiatives. The scalability of the ACHAGA algorithm also suggests its potential use in larger and more diverse agricultural settings.

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.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

PZ: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Writing – original draft, Writing – review & editing. YL: Software, Supervision, Visualization, Writing – review & editing. HD: Data curation, Funding acquisition, Investigation, 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 grants from the Tarim University Pres-ident's Fund (Grant Nos. TDZKSS202429, TDZKCX202406).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1440234/full#supplementary-material

References

Ait Saadi, A., Soukane, A., Meraihi, Y., Benmessaoud Gabis, A., Mirjalili, S., and Ramdane-Cherif, A. (2022). UAV path planning using optimization approaches: A survey. *Arch. Comput. Methods Eng.* 29, 4233–4284. doi: 10.1007/s11831-022-09742-7

Asim, M., and Abd El-Latif, A. A. (2023). Intelligent computational methods for multi-unmanned aerial vehicle-enabled autonomous mobile edge computing systems. *J. ISA Trans.* 132, 5–15. doi: 10.1016/j.isatra.2021.11.021

Bao, W., Zhu, Z., Hu, G., Zhou, X., Zhang, D., and Yang, X. (2023). UAV remote sensing detection of tea leaf blight based on DDMA-YOLO. *Comput. Electron. Agric.* 205, 107637. doi: 10.1016/j.compag.2023.107637

Bian, Q., Xu, D., Kang, K., and Liu, J. (2022). UAV base station site selection based on spiral algorithm in complex environment. *J. Physics: Conf. Ser.* 212, 012080. doi: 10.1088/1742-6596/2185/1/012080

Das, J. N., Tiwari, M. K., Sinha, A. K., and Khanzode, V. (2023). Integrated warehouse assignment and carton configuration optimization using deep clustering-based evolutionary algorithms. *Expert Syst. Appl.* 212, 118680. doi: 10.1016/j.eswa.2022.118680

Famili, A., and Stavrou, A. (2022). "Eternal flying: Optimal placement of wireless chargers for nonstop drone flights," in *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. (Kyiv, Ukraine: IEEE). 1–6.

Fesenko, H., Kliushnikov, I., Kharchenko, V., Rudakov, S., and Odarushchenko, E. (2020). "Routing an unmanned aerial vehicle during npp monitoring in the presence of an automatic battery replacement aerial system," in *2020 IEEE 11th international conference on dependable systems, services and technologies (DESSERT)*. (Kyiv, Ukraine: IEEE). 34–39.

He, Y., Hou, T., and Wang, M. (2024). A new method for unmanned aerial vehicle path planning in complex environments. *Sci. Rep.* 14, 9257. doi: 10.1038/s41598-024-60051-4

Kuo, R. J., Kuo, P. H., Chen, Y. R., and Zulvia, F. E. (2016). Application of metaheuristics-based clustering algorithm to item assignment in a synchronized zone order picking system. *Appl. Soft Computing* 46, 143–150. doi: 10.1016/j.asoc.2016.03.012

Kochengin, A., Chrysostomou, G., and Shikhin, V. (2019). "Performance ofnonparametric wilcoxon test with reference to the samples with singularities." in 2019 III International conference on control in technical systems (CTS). (St. Petersburg, Russia: IEEE), 265–268.

Li, K., Han, Y., and Yan, X. (2022a). Distributed multi-UAV cooperation for dynamic target tracking optimized by an SAQPSO algorithm. *J. ISA Trans.* 129, 230–242. doi: 10.1016/j.isatra.2021.12.014

Li, X., Shen, Z., and Wei, W. (2019). "An improved firefly algorithm used to location selection of distribution center for logistics UA V," in *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*. (San Diego, CA, USA: IEEE). 1–8

Li, Y., Cao, G., Chen, C., and Liu, D. (2022b). Planning algorithm for route and task allocation of plant protection UAVs in multiple operating areas. *J. Math. Problems Eng.* 2022, 1–9. doi: 10.1155/2022/7599021

Li, Y., Xu, Y., Xue, X., Liu, X., and Liu, X. (2022c). Optimal spraying task assignment problem in crop protection with multi-UAV systems and its order irrelevant enumeration solution. *J. Biosyst. Eng.* 214, 177–192. doi: 10.1016/j.biosystemseng.2021.12.018

Liu, Y., Zhang, P., Ru, Y., Wu, D., Wang, S., Yin, N., et al. (2022). A scheduling route planning algorithm based on the dynamic genetic algorithm with ant colony binary iterative optimization for unmanned aerial vehicle spraying in multiple tea fields. *J. Front. Plant Sci.* 13, 998962. doi: 10.3389/fpls.2022.998962

Marjit, S., Bhattacharyya, T., Chatterjee, B., and Sarkar, R. (2023). Simulated annealing aided genetic algorithm for gene selection from microarray data. *J. Comput. Biol. Med.* 158, 106854. doi: 10.1016/j.compbiomed.2023.106854

Neshat, M., Sepidnam, G., Sargolzaei, M., and Toosi, A. N. (2014). Artificial fish swarm algorithm: a survey of the state-of-the-art, hybridization, combinatorial and indicative applications. *J. Artif. Intell. Rev.* 42, 965–997. doi: 10.1007/s10462-012-9342-2

Ozkan, O. (2021). Optimization of the distance-constrained multi-based multi-UAV routing problem with simulated annealing and local search-based matheuristic to detect forest fires: The case of Turkey. *Appl. Soft Computing* 113, 108015. doi: 10.1016/j.asoc.2021.108015

Paraforos, D. S., Sharipov, G. M., Heiß, A., and Griepentrog, H. W. (2022). Position accuracy assessment of a UAV-mounted sequoia+ Multispectral camera using a robotic total station. *Agriculture* 12, 885. doi: 10.3390/agriculture12060885

Qiu, H., Duan, H., and Shi, Y. (2015). A decoupling receding horizon search approach to agent routing and optical sensor tasking based on brain storm optimization. *Optik* 126, 690–696. doi: 10.1016/j.ijleo.2015.02.004

Saavedra, P., Pérez Franco, A., and Guerrero, W. J. (2021). "Location-Routing for a UAV-Based Recognition System in Humanitarian Logistics: Case Study of Rapid Mapping," in *Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future: Proceedings of SOHOMA LATIN AMERICA*, vol. 2021. (Cham, Switzerland: Springer International Publishing), 197–207.

Sajid, M., Mittal, H., Pare, S., and Prasad, M. (2022). Routing and scheduling optimization for UAV assisted delivery system: A hybrid approach. *Appl. Soft Computing* 126, 109225. doi: 10.1016/j.asoc.2022.109225

Santé, I., Rivera, F. F., Crecente, R., Boullón, M., Suárez, M., Porta, J., et al. (2016). A simulated annealing algorithm for zoning in planning using parallel computing. *Computers Environ. Urban Syst.* 59, 95–106. doi: 10.1016/j.compenvurbsys.2016.05.005

Sun, F., Wang, X., and Zhang, R. (2020). Task scheduling system for UAV operations in agricultural plant protection environment. *J. Ambient Intell. Humanized Computing*, 1–15. doi: 10.1007/s12652-020-01969-1

Tian, H., Mo, Z., Ma, C., Xiao, J., Jia, R., Lan, Y., et al. (2023). Design and validation of a multi-objective waypoint planning algorithm for UAV spraying in orchards based on improved ant colony algorithm. *J. Front. Plant Sci.* 14, 1101828. doi: 10.3389/fpls.2023.1101828

Verma, S. C., Li, S., and Savkin, A. V. (2023). A hybrid global/reactive algorithm for collision-free UAV navigation in 3D environments with steady and moving obstacles. *Drones* 7, 675. doi: 10.3390/drones7110675

Xu, Y., Sun, Z., Xue, X., Gu, W., and Peng, B. (2020). A hybrid algorithm based on MOSFLA and GA for multi-UAVs plant protection task assignment and sequencing optimization. *J. Appl. Soft Computing* 96, 106623. doi: 10.1016/j.asoc.2020.106623

Yan, F., Chu, J., Hu, J., and Zhu, X. (2024). Cooperative task allocation with simultaneous arrival and resource constraint for multi-UAV using a genetic algorithm. *J. Expert Syst. Appl.* 245, 123023. doi: 10.1016/j.eswa.2023.123023

Zhang, X., Zhang, F., Tang, Z., and Chen, X. (2023). A MILP model on coordinated coverage path planning system for UAV-ship hybrid team scheduling software. *J. Syst. Software* 206, 111854. doi: 10.1016/j.jss.2023.111854

Zhang, Z., and Zhu, L. (2023). A review on unmanned aerial vehicle remote sensing: platforms, sensors, data processing methods, and applications. *Drones* 7, 398. doi: 10.3390/drones7060398

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 October 2024
DOI 10.3389/fpls.2024.1450831

Check for updates

OPEN ACCESS

EDITED BY
Jun Ni,
Nanjing Agricultural University, China

REVIEWED BY
Sunghwan Jung,
Cornell University, United States
Stephen Wilson,
University of Strathclyde, United Kingdom
Tristan Gilet,
University of Liège, Belgium

*CORRESPONDENCE
Xiao-Ya Dong
dongxiaoya@ujs.edu.cn
Bai-Jing Qiu
qbj@ujs.edu.cn

RECEIVED 18 June 2024
ACCEPTED 30 September 2024
PUBLISHED 15 October 2024

CITATION
Gao Z-M, Hu W, Dong X-Y, Zhao X-Y, Wang S, Chen J and Qiu B-J (2024) Motion behavior of droplets on curved leaf surfaces driven by airflow. *Front. Plant Sci.* 15:1450831. doi: 10.3389/fpls.2024.1450831

COPYRIGHT
© 2024 Gao, Hu, Dong, Zhao, Wang, Chen and Qiu. 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.

Motion behavior of droplets on curved leaf surfaces driven by airflow

Zhou-Ming Gao, Wei Hu, Xiao-Ya Dong*, Xiao-Yuan Zhao, Song Wang, Jian Chen and Bai-Jing Qiu*

Key Laboratory of Modern Agricultural Equipment and Technology, Ministry of Education, Jiangsu University, Zhenjiang, Jiangsu, China

In air-assisted spraying, pesticide droplet retention on crop leaves is key to evaluating spray effectiveness. However, airflow can deform leaves, reducing droplet retention and affecting spray performance. This study used wind tunnels and high-speed cameras to capture leaf deformation at different airflow speeds and the motion of droplets on curved leaf surfaces. The results showed that leaf curvature during bending deformation is generally less than 0.05 mm^{-1}. Critical wind speed for droplet movement is negatively correlated with droplet size and leaf curvature, with a 24.8% difference between different leaf curvatures and a 17.5% difference between droplet sizes. The droplet's dimensionless shape variable is positively correlated with both droplet size and leaf curvature. The maximum shape variable on curved leaves reaches 0.24, with acceleration differences of about 30%, while droplets of different sizes show a maximum shape variable of 0.18 and an acceleration difference of up to 68%. These findings enhance understanding of droplet-leaf interactions and provide insights for improving pesticide efficiency.

KEYWORDS

airflow, droplet, curved leaf surface, critical wind speed, motion behavior

1 Introduction

The movement of droplets on the surface of leaves is a complex process influenced by the chemical composition of the droplets and the specific characteristics of the leaf surface (Gaskin et al., 2005). The ultimate goal is to maximize the retention of spray droplets and evenly distribute them on crop leaves to enhance the efficacy of pesticides. Currently, research on the interaction between spray droplets and leaves primarily focuses on the behavior of droplets upon impact and their performance after successfully being retained on the leaves. However, it remains unclear whether droplets can successfully remain on leaves under the influence of airflow. This study aims to examine the ability of droplets to resist external wind forces during air-assisted spraying and to reveal the movement behavior of droplets on leaves.

Existing studies have revealed various kinetic behaviors of droplets upon impact with a leaf surface, including adhesion, fragmentation, splashing, and bouncing (Zabkiewicz, 2007). Dorr et al. (2015) investigated the effects of spray Equation and leaf wetting characteristics on droplet behavior. Their results showed that droplet fragmentation increased with size and impact velocity, when droplets hit hydrophilic leaf surfaces, they are more likely to stick or fragment. In contrast, droplets tend to bounce off when they impact hydrophobic surfaces. Wang et al. (2021) investigated the deposition preferences of pesticide droplets on leaves using a boom mounted with three flat fan-shaped hydraulic nozzles positioned 1 m above the leaves, moving at a speed of 1 m/s. The droplets impacted the leaves at a velocity of approximately 2.8 m/s. The study found that the deposition of pesticide droplets on the leaf was primarily influenced by the leaf's elasticity. Tredenick et al. (2021) further investigated the evaporation behavior of surfactant droplets on plant leaves and synthetic surfaces at tilt angles of 0°, 45°, and 90° using an automatic tilt platform and high-speed camera. The results revealed that the microstructure and chemical properties of plant leaves significantly differ from synthetic surfaces in terms of evaporation behavior. While the tilt angle of the surface does influence evaporation, the concentration of surfactants has a more pronounced effect on droplet evaporation behavior. All of these studies used the method of cutting leaves and affixing them to solid planes, which is a justified approach. However, their treatments of the leaves primarily considered flat or inclined surfaces, overlooking curved surfaces. Natural plant leaves are highly sensitive to minor disturbances such as wind and raindrops (Gart et al., 2015; de Langre, 2019; Ma et al., 2023), which can cause bending and deformation. Therefore, the behavior of droplets on curved leaf surfaces cannot be ignored.

Some researchers have focused on the effect of the curved structure of the leaf on droplet motion. Wang et al. (2005) investigated the motion of droplets on curved surfaces such as leaves, and further achieved an accurate simulation of droplet motion by using a virtual surface and a dynamic contact angle model to generate a stable contact angle. Mayo et al. (2015) investigated the movement of droplets on leaf surfaces with curvature gradients by placing droplets of various sizes near the peak of the leaf surface and observing their movement paths. They discovered that the motion of 0.3 mm diameter spray droplets was primarily driven by the curvature gradient of the leaf surface. However, as the droplet diameter increased to 5 mm, gravity became the dominant force influencing droplet movement. Zheng et al. (2023) investigated the sedimentation behavior of droplets impacting curved rice leaves. They characterized the natural bending of these leaves during growth using the curvature radius. Their research revealed that the curved surface structure of the leaves significantly affects droplet deposition behavior. These studies above examined the curvature gradient and natural bending of the leaf surface, which are influenced by the structural properties and growth mode of the leaf. However, when subjected to external forces like wind, the leaf experiences significant deformations, and the behavior of droplets on such extensively deformed leaves remains unclear.

Current research on the dynamic behavior of fluid ridges on solid substrates under airflow provides valuable insights into the retention of pesticide droplets on plant leaves. Most of these studies rely on theoretical modeling and numerical simulations. For example, Sullivan et al. (2008, 2012) examined the effect of shear stress on the morphology and stability of fluid ridges on inclined and vertical plane substrates. By solving a mathematical model based on the Navier-Stokes equation, they analyzed the effects of shear stress, initial fluid ridge morphology, and surface tension. Their results indicated that strong shear stress can cause fluid ridges to rupture, with stability also influenced by surface tension and initial morphology. Wilson et al. (2011) investigated the dynamic behavior of fluid films and ridges under shear stress on vertical plane substrates, focusing on energy conversion during liquid film rupture. They found that shear stress significantly affects the rupture location, timing, and energy threshold. Paterson et al. (2015) explored the coupling between fluid ridge shape and airflow velocity distribution, revealing that higher airflow velocities have a notable impact on fluid ridge morphology, while ridge stability depends on airflow velocity, surface tension, and initial morphology. Building on their work with planar substrates, Paterson et al. (2014) further investigated the effect of uniform shear stress on the flow of fluid ridges over a horizontal cylindrical surface, finding that the flow pattern depends on both the shear stress and the cylinder's curvature. Mitchell et al. (2022) explored the unsteady behavior of coating flow on a rotating cylinder under non-cyclonic and cyclic flow conditions, discovering that the interaction between airflow circulation and cylinder rotation speed can lead to complex flow patterns, which in turn affect the final distribution of the coating. These studies collectively demonstrate that airflow exerts a significant influence on the morphology and stability of fluid ridges, making it a crucial factor to consider.

In addition, existing research has shown that aerodynamics significantly impacts the dynamic behavior of droplets on solid substrates (Ding and Spelt 2008; Ding et al., 2010) investigated the critical conditions for three-dimensional droplets to initiate movement on solid surfaces under shear flow, as well as their sliding, detachment, and fracture behaviors. They found that the critical condition for droplet sliding is closely related to the balance between applied shear stress, surface tension, and resistance from contact angle hysteresis. Droplet size and contact angle hysteresis were also found to influence the critical conditions for sliding and eventual detachment. Wilson et al. (2023) studied droplet morphology, movement paths, and detachment processes under varying airflow velocities and fiber spacings using wind tunnel experiments and high-speed imaging. Their results showed that droplet motion and deformation are affected by airflow velocity and interactions between adjacent droplets. Smaller fiber spacings enhance droplet interactions, leading to more complex motion patterns. While the above research primarily focuses on artificial surfaces, plant leaf surfaces naturally possess more complex structures. Therefore, studying the dynamic behavior of droplets on natural plant leaf surfaces can deepen our understanding of the interactions between droplets and leaves.

Oqielat et al. (2011) used the Frangipani leaf surface as the basis for droplet motion, modeling the leaf surface as a triangular element composed of multiple inclined planes, guiding droplet motion through the steepest path. It was found that the small structures and angle changes on the surface of the leaves have a significant impact on the movement path of water droplets. Building on this foundation, Veremieiev et al. (2014) further investigated the flow of biopesticide droplets on inclined plane leaves, focusing on the diffusion, coalescence, and evaporation processes through experiments and numerical simulations. Their research demonstrated that factors such as the leaf microstructure and chemical properties, droplet size, environmental temperature and humidity, and wind speed significantly influence the flow behavior of droplets. The above studies primarily examined the dynamics of droplets on inclined leaves influenced by gravity, yet practical engineering applications must also consider airflow forces. Particularly in airflow-assisted spraying, deposited droplets on the leaf may migrate due to airflow, merging with others along their path to form larger droplets. These larger droplets can be lost in runoff along leaf edges, leading to pesticide wastage, reduced efficacy, and environmental pollution.

In this study, a wind tunnel and high-speed camera system were employed to observe the motion of droplets on curved capsicum leaves under airflow. The critical wind speeds necessary for initiating droplet movement on these curved leaves were determined. A mechanical analysis model considering droplet size and leaf curvature was developed, and the influence of these factors on droplet behavior was analyzed. The ultimate goal is to offer new insights for enhancing pesticide utilization and optimizing decisions related to airflow-assisted spraying.

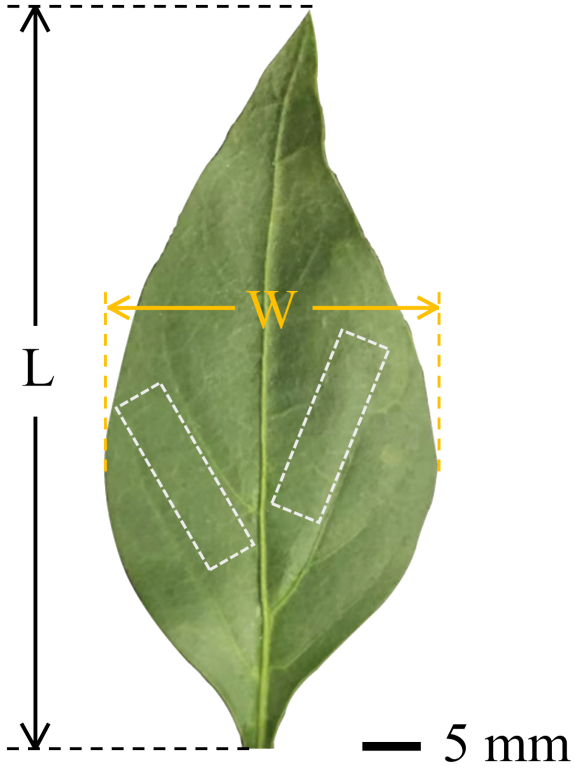


FIGURE 1
Capsicum leaf morphological and geometric parameters. (The white area is the clipping area).

2 Materials and methods

2.1 Plants cultivation and sample preparation

Capsicum plants (Capsicum annuum L., Tianshuai101) were selected as test samples. Leaves from capsicum plants with a growth cycle of about 3 months were used, with the average leaf area measured by ImageJ software being approximately 1461 mm^2. The marginal profile of the capsicum leaves was considered elliptical, with their geometry described by leaf length (L) and width (W), as illustrated in Figure 1. The leaf size and physical parameters are detailed in Table 1.

Healthy, undamaged, and fully saturated leaves were selected as experimental samples. A sharp blade was used to cut the petiole 10 - 15 mm from the base of the leaf, and the remaining petiole was clamped with a sleeve clamp. The intact leaves were placed vertically at the center of the wind tunnel testing section, with the front of the leaf facing the wind. The movement of the petiole was fully restricted, allowing only the bending deformation of the leaf under airflow to be considered. After successfully capturing the deformation data, the adjacent secondary veins of the capsicum leaves were cut to obtain rectangular leaf sections measuring 6 × 18 mm (marked in white in Figure 1). Double-sided tape was used to attach the cut leaf samples smoothly onto a custom-made curved substrate. ensuring the leaves were tightly secured without wrinkles to maintain the accuracy and consistency of subsequent experiments. The entire process was completed within one minute to minimize changes in leaf moisture content due to evaporation caused by the cuts.

2.2 Experiment setup

2.2.1 The overall configuration of experiment instruments

The overall experimental setup is illustrated in Figure 2. It includes a droplet physical characteristic detection platform and a test platform for observing the motion behavior of leaves and droplets under airflow. The droplet physical property detection

TABLE 1 Sample size and physical parameters.

Parameter	Unit	Value
Leaf length (L)	mm	58 ± 7
Leaf width (W)	mm	26 ± 5
Leaf area	mm^2	1461 ± 305
Leaf midrib modulus of elasticity	MPa	8 ± 2
Leaf mesophyll modulus of elasticity	MPa	3 ± 1

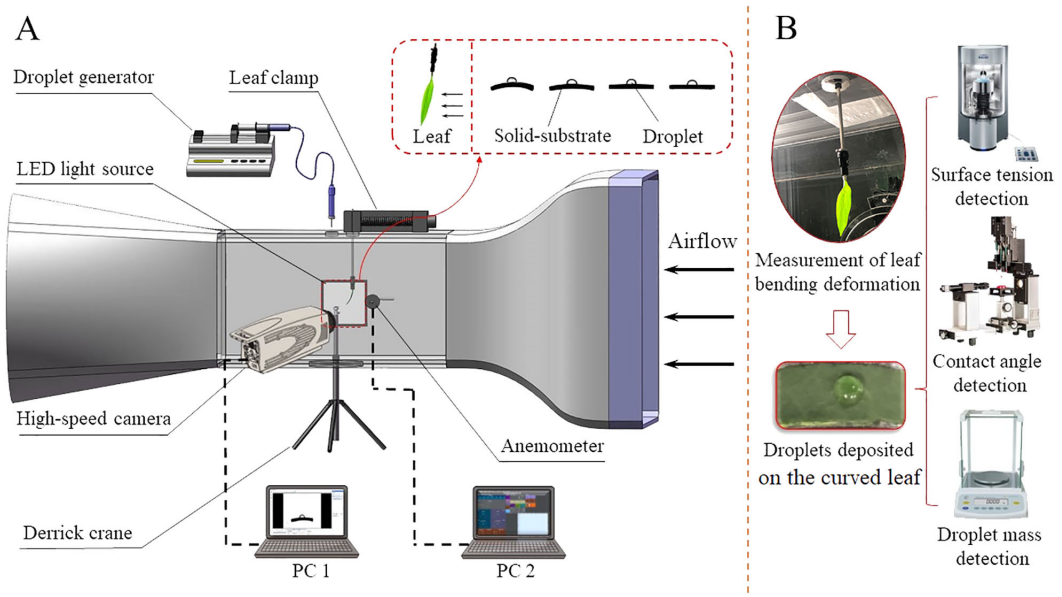


FIGURE 2
Experimental schematic diagram (arrow indicating airflow direction): **(A)** Experimental platform for leaf and droplet motion behavior; **(B)** Droplet physical property detection platform.

platform comprises a contact angle measuring instrument (OCA 25, Dataphysics, Germany), a surface tension instrument (Kruss K100, Klux, Germany), and a precision analytical balance (BT125D, Kunshan Norbond Electronic Technology Co., Ltd., Suzhou, China). The test platform for leaf and droplet motion behavior under airflow includes a small open wind tunnel, a pitot tube anemometer, a high-speed camera (SH6-109, Shenzhen Shenshi Intelligent Technology Co., Ltd., China), PC (Legion Y7000P, Lenovo Co., Ltd., Beijing, China), a micro injection pump (MDG-100, TSI, USA), a tripod, an LED light source, and homemade leaf clamp and curved substrates with different curvatures. (For detailed information about the leaf clamp, please refer to Ma et al., 2023).

2.2.2 Measurement experiment of droplet physical properties

The physical properties of the droplets examined in this study include surface tension, contact angle, and droplet particle size, which were measured as follows:

Measure the surface tension coefficient (σ_{la}) of selected deionized water using the "Wilhelmy plate method". In addition, the contact angle measuring instrument was used to measure the contact angle of droplets on curved capsicum leaves. For static contact angle (θ_s) measurement, a 3 μL droplet was placed on the leaf surface for 5 s, and its image was captured to determine the angle. For dynamic contact angle measurements, the advancing contact angle (θ_A) was obtained by placing a needle near the leaf surface and injecting liquid, while the receding contact angle (θ_R) was measured by withdrawing liquid from the static droplet. To calibrate droplet size, a micro-injection pump generated stable thrust on the syringe at a speed of 50 mL/h, producing droplets. The droplet size was controlled by using different types of needles, and each droplet was weighed with a precision analytical balance. Each experiment was repeated 3 times, and the average value was calculated. Finally, the droplet volume was determined using the relationship between mass and density ($m = \rho \cdot V$).

2.2.3 Measurement experiment on leaf bending deformation and substrate preparation

Before the experiment, the camera was calibrated. The leaf was then fixed at the center position of the wind tunnel test section using a sleeve fixture, as shown in Figure 2. Field measurements indicated that the wind speed reaching the leaf during long-distance air supply and spraying ranges from approximately 2.4 to 11.3 m/s. For detailed field measurement procedures and results, refer to Supplementary File 1. Consequently, the wind speed range selected for this experiment was 3.0 to 10.0 m/s, with increments of 2.0 m/s. The wind tunnel generated the required wind speeds during the experiment, and the overall structure of the wind tunnel system is illustrated in Figure 3.

By setting the desired wind speed through the speed controller, the airflow uniformly passes through the test section. This section is a rectangular channel made of transparent acrylic, measuring 1150 mm in length with a cross-sectional area of 436 × 436 mm^2. A pitot tube was placed at the center of the test section to measure the airflow velocity in real-time with an accuracy of ± 0.1 m/s. To avoid the influence of instantaneous wind speed, the wind speed was stabilized for 5 s before collecting data on leaf motion under airflow. It was found that, although leaves might sway laterally under airflow, the primary deformation observed was bending and oscillation. This was consistent with Shao et al. (2012), as confirmed in over 80% of the test leaves. Therefore, this study focuses on leaf bending deformation as a representative measure of overall deformation, temporarily excluding the oscillation form.

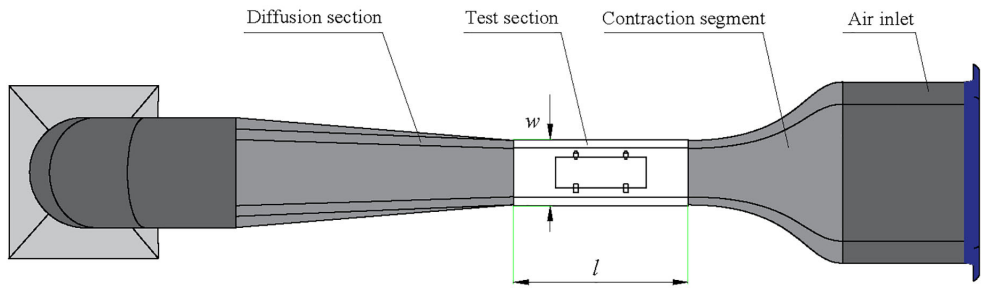


FIGURE 3
Schematic diagram of the experimental wind tunnel structure. (*l* and *w* are the length and width of the test section).

For statistical purposes, the leaves need to be simplified. According to Moulia et al. (1994), the midrib of a leaf plays a crucial role in the bending deformation of the entire leaf, accounting for more than 87% of the leaf's rigidity. Therefore, this paper uses the bending of the midrib to represent the overall leaf bending. The image digitizer software Engage Digitizer 4.1 was used to characterize the bending shape of leaves under airflow, with a set of points representing the contour of the midrib. Each midrib was marked with 50 to 70 points, at a density of 1.7 to 2.4 points/mm, with higher point density in areas of high curvature. After measuring the leaf bending deformation, the experimental data was imported into Origin 2022 and fitted using the least squares method to obtain the polynomial $f(x)$ that characterizes leaf bending. According to the curvature definition of the curve in the plane, the curvature κ of each point (x_i, y_i) along the midrib on the leaf can be further obtained as:

$$\kappa = \frac{\left| d^2 f(x_i)/dx^2 \right|}{\left\{ 1 + [df(x_i)/dx]^2 \right\}^{3/2}} \quad (1)$$

According to Equation 1, the curvature at each local position of the leaf under airflow can be determined. Given the droplet's small size relative to the leaf, it's assumed that the contact area between the leaf and the droplet forms a cylindrical surface. To study droplet motion on capsicum leaves with various bending deformations, curved substrates with different curvature were 3D printed to secure the leaves. The substrate is made of nylon fiber material, and the prepared substrate surface is smooth. The solid substrate size (L × W × Th) selected by the rectangular leaf size to be cut is 50 × 20 × 2 mm.

2.2.4 Measurement experiment on droplet motion behavior on leaves

To study the bending characteristics of capsicum leaves without compromising their structural integrity, we employed straightforward leaf treatments. Firstly, we precisely cut between adjacent secondary veins of capsicum leaves to obtain rectangular samples measuring 6 × 18 mm. These cut leaf samples were then smoothly affixed onto custom-made curved substrates using double-sided tape. Each leaf sample was used once and replaced with new ones for subsequent experiments. The curved substrates were securely fixed on tripods using fixtures, with adjustments made to tripod height and positioning to center the leaves on the curved bases within the wind tunnel test section. Considering the relatively small size of the curved base compared to the wind tunnel test section, its impact on airflow was negligible. The syringe needle was fixed 5 mm vertically above the center of the leaf surface and connected to a micro injection pump to generate droplets of varying sizes. Droplets were gently deposited onto the leaves. During the experiment, a backlight method was employed for photography and recording. The height of the high-speed camera was adjusted to ensure the droplet was centered in the image, with the light source aligned accordingly. The camera settings were adjusted to capture images at a speed of 120 fps.

The required wind speed for this experiment was generated by the wind tunnel. By adjusting the frequency controller of the axial flow fan, we achieved a stable increase in wind speed up to the maximum value. Throughout this process, the motion of droplets was captured using a high-speed camera system. For this experiment, the frequency of the axial fan was set to 45 Hz, and the variation of wind speed from 0 to 45 Hz over time is depicted in Figure 4. At the start of each experiment, the fan and high-speed camera were synchronized to record real-time data on wind speed growth and images of droplet motion. The experiment continued until the droplet reached the edge of the substrate, with a maximum

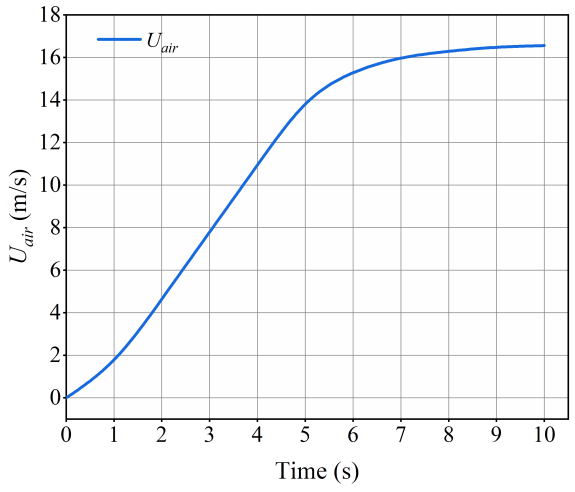


FIGURE 4
The variation of wind speed over time.

10.3389/fpls.2024.1450831

duration of 10 s. Each set of experiments was conducted three times under identical conditions.

To gather comprehensive insights into droplet motion on curved leaves, extracting droplet contact line information from captured images is crucial. AutoCAD 2018 drawing software is utilized for accurately outlining the droplet edge lines visible in the images. By analyzing the positions of the droplet contact lines at different points in time, key parameters that characterize droplet motion are obtained (Hu et al., 2013), such as droplet position, droplet velocity, and droplet shape variables.

3 Results and discussion

3.1 Bending deformation of leaves under airflow

To explore how the curved surface structure of the leaf affects the retention capacity of droplets, this study conducted bending deformation tests using a wind tunnel system under varying airflow velocities. The curvature was introduced to precisely digitize the leaf's bending deformation under airflow. The coordinates of any point on the leaf's midrib were set as the x-axis and y-axis, with the position near the petiole serving as the origin O. Figure 5A illustrates the bending deformation results of the leaves under airflow speeds of 3.0 m/s, 5.0 m/s, and 7.0 m/s, represented by color variations from dark to light. The airflow direction aligns with the positive x-axis.

The bending deformation of the leaf is characterized by fitting the digitized point set on the leaf midrib into a polynomial $f(x)$. Therefore, the accuracy of the polynomial fitting becomes crucial. To determine the optimal order of the fitting polynomial, residual analysis serves as a key indicator of the fit quality. This involves generating data of the form $[x_i, f(x_i) + e(x_i)]$, where x_i represents the true abscissa of the leaf midrib, $f(x_i)$ denotes the polynomial estimate of the midrib ordinate, and $e(x_i)$ represents the residual between the curve fitting result and the sampled point set, as illustrated in Tables in Figures 5B–D. The error in the sampled point set for the leaf midrib primarily arises from the thickness variations of the midrib itself, which were measured to range from 0.82 to 2.96 mm in the experimental capsicum leaves. The residuals obtained from fitting the leaf midrib curve all fall within ± 0.05 mm, indicating compliance with fitting requirements.

Figures 5B–D correspond to different bending deformation states of the leaves shown in Figure 5A under varying wind speeds, and they depict the curvature κ at different positions along the leaf midrib as calculated from Equation 1. In Figures 5B, at an airflow speed of 3.0 m/s, the overall curvature of the leaf ranges from 0 to 0.03 mm^{-1}, with larger values observed in the range of x = 11 to 23 mm. This indicates that at lower wind speeds, bending deformation primarily occurs near the leaf tip and base, with lesser deformation in the middle of the leaf. This is due to the thinner and softer nature of the leaf near these areas, making them more susceptible to bending under the influence of wind. Conversely, the middle section of the leaf, being thicker and sturdier, offers more resistance to wind, resulting in less

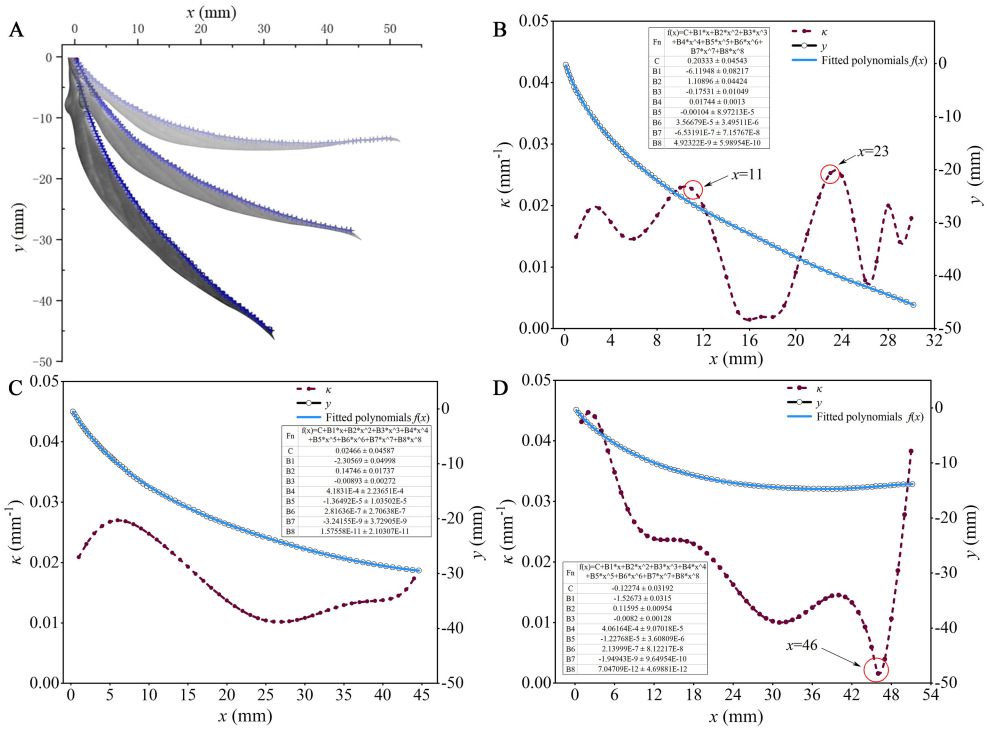


FIGURE 5
Bending deformation of leaves under different wind speeds: **(A)** Information extraction of leaf bending deformation; **(B)** A wind speed of 3.0 m/s; **(C)** A wind speed of 5.0 m/s; **(D)** A wind speed of 7.0 m/s.

deformation. In Figure 5C, at an airflow speed of 5.0 m/s, the overall curvature of the leaf ranges from 0.01 to 0.03 mm^{-1}. This suggests that the degree of bending deformation along different positions of the leaf midrib is relatively uniform. Figure 5D shows the results at an airflow speed of 7.0 m/s, the overall curvature of the leaf ranges from 0 to 0.05 mm^{-1}, where the minimum value observed in x = 46 mm. This indicates that under high wind speeds, bending deformation predominantly occurs away from the leaf tip. As wind speed increases, the distribution of wind pressure on the leaf surface becomes more uniform, affecting all parts of the leaf significantly. However, due to the larger surface area near the middle and base of the leaf, greater deformation is observed in these regions. Overall, the curvature during leaf bending deformation mostly remains below 0.05 mm^{-1}. Therefore, this study selected curved substrates with curvature (κ) of 0.01, 0.02, and 0.04 mm^{-1} to represent different bending deformations of the leaves, alongside a planar substrate (κ = 0) for comparison. This setup aims to explore the motion behavior of droplets on leaves with varying degrees of bending deformation.

3.2 Wetting characteristics of droplets on curved leaf surfaces

The wetting characteristics of droplets on a surface directly influence their ability to adhere to that surface. The static contact angle describes the angle formed by a droplet in its equilibrium state on a surface, while the dynamic contact angle refers to the angle when the droplet is in a non-equilibrium state. To explore the retention capability of droplets on curved leaf surfaces, experiments were conducted to measure the contact angles on capsicum leaves with varying curvature radii. These experiments serve as a foundational step for subsequent investigations into the motion behavior of droplets driven by airflow on curved leaves. The experimental results presented in Table 2. The contact angle measurement is shown in Figure 6.

The experimental results showed that as the curvature of the leaves increases, the static contact angle (θ_s) of droplets on capsicum leaves tends to decrease. In other words, the greater the bending deformation of the leaf, the smaller the static contact angle between the droplet and the leaf. This finding differs from previous research on small droplets on cylindrical fibers. Brendan (1984) demonstrated that the curvature of cylindrical fibers significantly impacts droplet shape and stability. Smaller diameter cylindrical fibers lead to higher curvature effects, potentially causing droplets to become unstable or adopt non-axisymmetric shapes. In contrast, larger diameter cylindrical fibers, with their smaller curvature effect, result in more stable equilibrium shapes for droplets. McHale et al. (1999) studied the wetting behavior of small droplets on fibers with diameters ranging from 40 to 120 μm and found that the small diameter and high curvature of the fibers make droplets more likely to form wrapping structures, resulting in larger contact angles. However, for the bending deformation of leaves, the curvature of the leaves is much smaller than that of cylindrical fibers, and droplets do not form wrapping structures on the leaf surface. Additionally, an increase in leaf curvature means that droplets are more likely to contact the leaves, increasing the surface area of contact. This makes it easier for the gravity and surface tension of the droplets to balance, thereby forming a smaller contact angle.

TABLE 2 Contact angle of droplets on differently curved capsicum leaves.

κ/(mm^{-1})	θ_s/(°)	θ_A/(°)	θ_R/(°)	$\Delta\theta_{CAH}$/(°)
0.041	90 ± 4	105 ± 3	65 ± 3	40 ± 6
0.027	95 ± 3	108 ± 3	69 ± 4	39 ± 7
0.013	99 ± 3	110 ± 3	73 ± 2	37 ± 5
0	104 ± 3	114 ± 3	78 ± 3	36 ± 6

3.3 The motion of droplets on curved leaf surfaces

3.3.1 The critical wind speed at the beginning of droplet movement

To provide a clearer and more visual representation of droplet motion on the leaves, large droplets were selected for this experiment. The dynamic similarity between the experimental model and actual spray droplets was established using a dimensionless similarity principle, as shown in Equation 2. In field spraying, pesticides are often mixed with surfactants to lower the liquid's surface tension, enhancing spreading and adhesion on plant leaves. The typical surface tension of such liquids ranges between 20 and 40 mN/m, with droplet diameters between 147 and 1138 μm (Butts et al., 2024). In this study, deionized water was used as the droplet liquid, with a surface tension of 72 mN/m. The Weber number (We) was kept constant during testing to maintain similarity with field conditions. To ensure clarity and visibility of droplet movement on the leaf, larger droplets were selected. The initial radii of the droplets were set at 1.419 mm, 1.571 mm, 1.748 mm, 1.878 mm, and 2.006 mm, respectively.

$$\begin{cases} We = \frac{\rho U_a^2 R_0}{\sigma_{la}} \\ Oh = \frac{\mu}{\sqrt{\rho \sigma_{la} R_0}} \\ Bo_g = \frac{\rho R_0^2 g}{\sigma_{la}}, Bo_i = \frac{\rho R_0^2 L \omega^2}{2\sigma_{la}} \end{cases} \quad (2)$$

Where, ρ is the liquid density, Kg/m^3; U_a is wind speed, m/s; R_0 is the droplet radius, m; σ_{la} is the surface tension, N/m; μ is the dynamic viscosity of the fluid, Pa·s; g is gravitational acceleration, m/s^2; L is the length of the leaves, m; ω is the oscillation angular frequency of the leaf, rad/s, ω=2πf; f is the oscillation frequency of the leaf, Hz.

The small droplets in actual spray are more sensitive to viscous dissipation, while the large droplets selected for this study are more affected by gravity. Therefore, it is essential to clarify the roles of gravity and viscosity on the dynamic behavior of the droplets. Based on the definition of the dimensionless Bond number (Bo_g) considering the influence of gravity, the maximum Bo_g number corresponding to the droplet radius chosen in this study, $Bo_{g\text{-}max}$ =0.547< 1, is obtained.

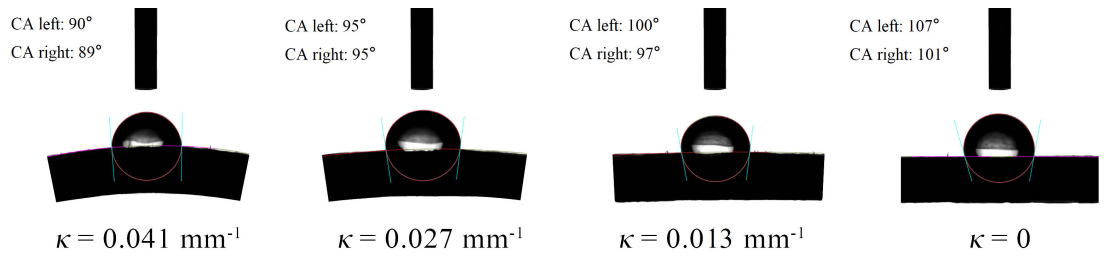


FIGURE 6
Measurement of contact angle of droplets on different curved leaves.

This indicates that the dynamic behavior of droplets is mainly influenced by surface tension, and the effects of droplet gravity is relatively small. In addition, the leaves experience some oscillation under the influence of airflow. According to the research method of Zhang et al. (2022), the oscillation frequency (f) of the leaves in this study is below 10 Hz. As shown in Table 1, the maximum leaf length (L) is below 0.065 m. Based on the definition of the dimensionless Bond number (Bo_i) considering the influence of inertial forces, the solution is $Bo_{i\text{-}max}$ =7.149 > 1. This indicates that droplet behavior is influenced not only by surface tension but also by the inertial forces from leaf oscillation. This study focused on droplet size and leaf curvature, temporarily overlooking the effects of leaf oscillation. Future research will investigate how leaf oscillation influences droplet dynamics and its interaction with surface tension, enhancing our understanding of droplet-leaf interactions in pesticide spraying.

Additionally, it must be acknowledged that perfect geometric similarity between the model and reality was not fully achieved. While size similarity was matched by adjusting surface tension, this altered the contact angle. The similarity in the ratio between droplet size and leaf curvature was also not considered. Given the significant variations in contact angles and curvatures across different leaf types during field pesticide spraying, these differences were temporarily overlooked in the model. Achieving true equivalence in droplet viscosity would require significant changes in fluid properties. The dynamic similarity in this study demands that both the dimensionless *We* number and *Oh* number be equivalent. According to Equation 2, to achieve this similarity for the radius-length ratio between the selected experimental droplet and the actual spray droplet, a fluid with a kinematic viscosity approximately 10 times that of water would be ideal. However, water was used for its practicality and widespread application, leading to a deviation from the ideal *Oh* number similarity. Nonetheless, the *We* number remains constant as the primary parameter, and using water still produces meaningful results, especially for practical pesticide spraying scenarios.

Before the droplets begin to move on curved capsicum leaves, they undergo deformation and oscillation driven by airflow. Figure 7 shows the motion behavior of a droplet with R_0 = 2.006 mm on different curved leaf surfaces. As the wind speed continues to increase and exceeds a critical value, the contact line between the droplet and the leaf starts to move. This airflow velocity is defined as the critical wind speed (Wang et al., 2022). As shown in Figure 8A, the critical wind speed corresponds to the airflow velocity at which the first non-zero value of x_a or x_r occurs.

The motion of droplets on curved leaf surfaces primarily depends on the aerodynamic force (F_d) and hysteresis tension (F_σ) applied to the droplets (Mahato and Mandal, 2021; Mortazavi and Jung, 2023). The hysteresis tension (F_σ) is related to the forward contact angle (θ_A) and the backward contact angle (θ_R). To theoretically analyze the force situation of droplet motion on curved leaf surfaces, we assume that the geometric shape of the droplets on curved leaves is spherical and that the droplet shape does not change due to airflow. The only deformation considered is the local change in contact angle around the contact line. Given the

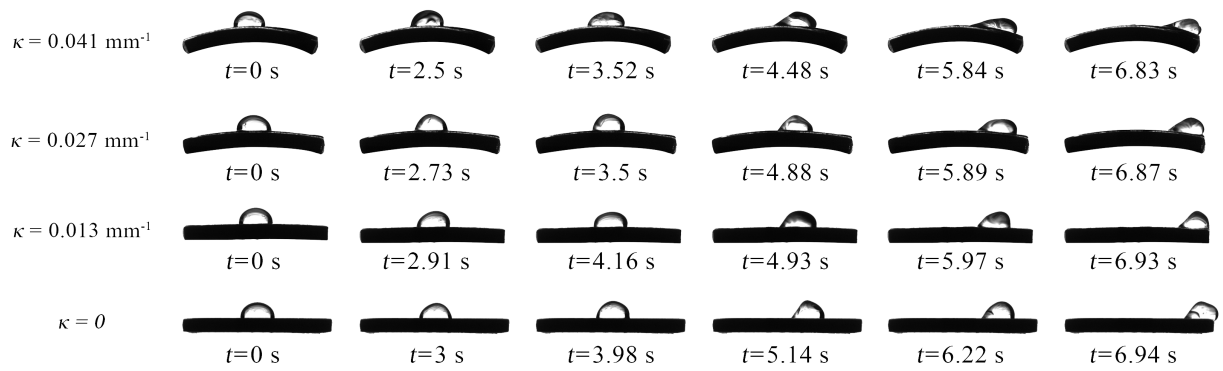


FIGURE 7
Motion process of droplets with R_0 = 2.006 mm on different curved leaves.

10.3389/fpls.2024.1450831

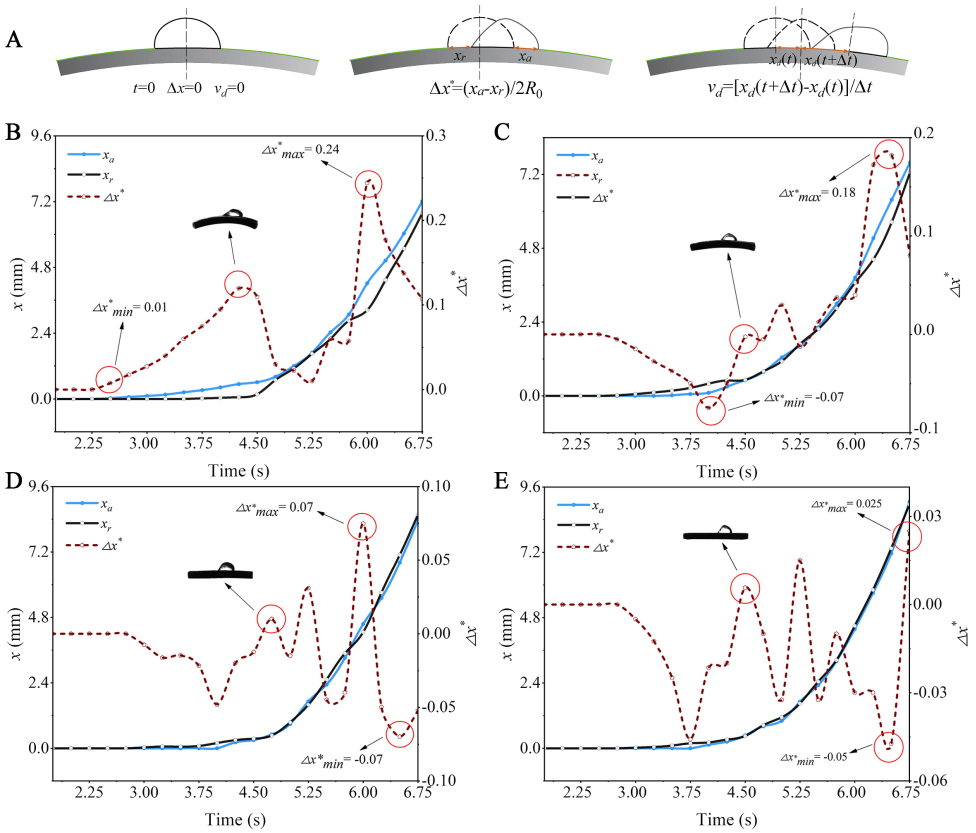


FIGURE 8
The variation of the contact line position of a droplet with R_0 = 2.006 mm on different leaves over time: **(A)** Measurement methods for key characteristic parameters of droplets; **(B)** κ =0.041 mm^{-1}; **(C)** κ =0.027 mm^{-1}; **(D)** κ =0.013 mm^{-1}; **(E)** κ =0.

small size of the droplets relative to the entire leaf, it is assumed that the contact surface between the leaf and the droplets is a cylindrical surface. The center position of the droplet is set as the origin *O* of the coordinate system, with the *y*-axis aligned with the wind direction, the *z*-axis perpendicular to the base normal, and the *x*-axis determined by the right-hand rule, as shown in Figure 9.

For incompressible pipeline flow with a zero pressure gradient in the vertical direction, a Reynolds number *Re*< 2100 indicates laminar flow, while *Re* > 4000 indicates turbulent flow. The range 2100< *Re*< 4000 corresponds to a transitional flow state. For both laminar and turbulent boundary layer flows along a flat channel, the boundary layer thickness (δ) can be approximated as follows:

$$\begin{cases} \delta_{la} \approx \frac{5x}{\sqrt{\mathrm{Re}}} \\ \delta_{tu} \approx \frac{0.37x}{\sqrt[5]{\mathrm{Re}}} \\ \mathrm{Re} = \frac{U_a \cdot x}{\upsilon} \end{cases} \tag{3}$$

Where, *Re* is the Reynolds number; *x* is the distance downstream from the starting point of the boundary layer, m; υ is the kinematic viscosity of the fluid, $\upsilon = \mu/\rho_a$, m^2/s.

The airflow velocity driving droplet motion in this experiment ranges from 7 to 12 m/s, and the distance *x* between the entrance

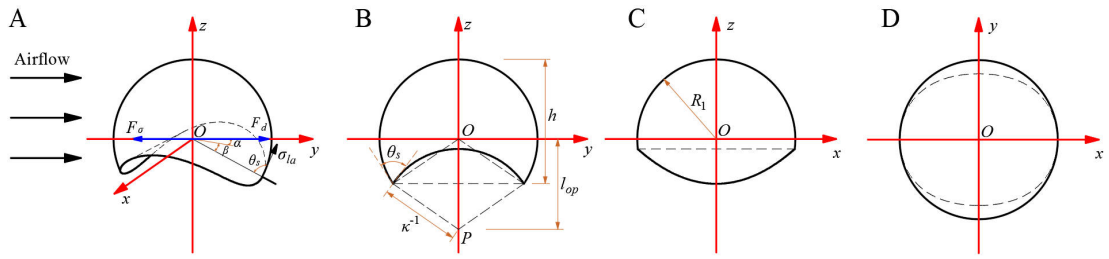


FIGURE 9
Analysis of droplet forces and its projection diagram: **(A)** Droplet force analysis; **(B)** *Oyz* plane projection; **(C)** *Oxz* plane projection; **(D)** *Oxy* plane projection.

of the wind tunnel test section and the droplet's center is approximately 0.6 m. According to Equation 3, the Reynolds number at the highest wind speed is 4.93×10^5, while at the lowest wind speed, it is 2.88×10^5, indicating that the airflow in the wind tunnel for this experiment is turbulent. The corresponding boundary layer thickness (δ_{tu}) at the highest wind speed is 16×10^{-3} m, and at the lowest wind speed, it is 18×10^{-3} m. Additionally, according to Equation 10, the maximum height (h) of the droplets placed on the leaf surface is 2.4×10^{-3} m, showing that all droplets lie within the boundary layer and are influenced by the shear flow within this region.

According to Sugiyama and Sbragaglia (2008), the shear aerodynamic force (F_d) acting on a hemisphere on a solid substrate is:

$$F_d = k \cdot (4.3\pi R_1 \cdot \mu \cdot U_a) \tag{4}$$

Where, k is a wall correction factor accounting for the proximity of the wall; R_1 is the diameter of the fixed droplet obtained based on volume equilibrium, m.

Shear stress in fluids (τ_s) is:

$$\tau_s = \frac{\mu \cdot U_a}{R_1} \tag{5}$$

Numerous experiments and theoretical studies have demonstrated that, for flow inside a pipe, the frictional force (F_s) generated by the fluid is proportional to the kinetic energy term ($\rho_a {U_a}^2$) of the fluid and the contact area (A) between the fluid and the wall. This relationship can be expressed as:

$$F_s = f \cdot \frac{\rho_a \cdot U_a^2}{2} \cdot A \tag{6}$$

Furthermore, the friction coefficient (f) can be obtained as:

$$\begin{cases} f = \frac{2\tau_s}{\rho_a \cdot U_a^2} = \begin{cases} 16 \cdot \mathrm{Re}^{-\frac{1}{2}} & \mathrm{Re} < 500 \\ 0.0792 \cdot \mathrm{Re}^{-\frac{1}{4}} & \mathrm{Re} \geq 500 \end{cases} \\ \mathrm{Re} = \frac{\rho_a U_a D}{\mu} \end{cases} \tag{7}$$

Where, ρ_a is the fluid density, Kg/m^3; D is the effective diameter of the wind tunnel test section, $D = 4A_0/S$, m; A_0 is the cross-sectional area of the wind tunnel test section, m^2; S is the circumference of the wind tunnel test section, m.

Substituting Equation 7 into Equation 5, the shear stress obtained is:

$$\tau_s = \frac{0.0792}{2} \cdot \left(\frac{\rho_a U_a D}{\mu}\right)^{-\frac{1}{4}} \cdot \rho_a \cdot U_a^2 \tag{8}$$

By combining the above formulas and substituting them into Equation 4, the aerodynamic force (F_d) acting on the droplet can be derived as follows:

$$F_d = \frac{0.535k \cdot {R_1}^2 \cdot {\rho_a}^{\frac{3}{4}} \cdot {U_a}^{\frac{7}{4}} \cdot \mu^{\frac{1}{4}}}{D^{\frac{1}{4}}} \tag{9}$$

According to the geometric relationship in Figure 9, it can be concluded that:

$$\begin{cases} R_1 = \left({R_0}^3 + \frac{3}{\pi} \cdot \int_0^{\left({R_1}^2 - \left(\frac{1}{\kappa} - l_{op}\right)^2\right)^{\frac{1}{2}}} dx \int_0^{\left(\frac{1}{\kappa^2} - \left(\left(x^2 - {R_1}^2 + \frac{1}{\kappa^2} + l_{op}^2\right)/2l_{op}\right)^2\right)^{\frac{1}{2}}} dy \int_{-({R_1}^2 - x^2 - y^2)^{\frac{1}{2}}}^{\left(\frac{1}{\kappa^2} - y^2\right)^{\frac{1}{2}} - l_{op}} dz \right)^{\frac{1}{3}} \\ A_P = \pi {R_1}^2 \cdot \left(1 - \left(\arccos\left(\left(l_{op} - \frac{1}{\kappa}\right)/R_1\right)/180\right)\right) + \sqrt{{R_1}^2 - \left(\frac{1}{\kappa} - l_{op}\right)^2} \cdot \left(l_{op} - \frac{1}{\kappa}\right) \\ \quad + 2 \cdot \int_0^{\sqrt{{R_1}^2 - \left(\frac{1}{\kappa} - l\right)^2}} l_{op} - \frac{1}{\kappa} - y(x) dx \\ h = R_1 + R_1 \cdot \left(1 - \left(\frac{\sin^2 \theta_s}{\kappa^2 {R_1}^2 + 1 - 2R_1 \cdot \kappa \cos \theta_s}\right)\right)^{\frac{1}{2}} \\ l_{op} = \left({R_1}^2 + \frac{1}{\kappa^2} - \frac{2R_1 \cos \theta_s}{\kappa}\right)^{\frac{1}{2}} \\ y(x) = \left(\frac{1}{\kappa^2} - \left(\frac{\left(x^2 - {R_1}^2 + \frac{1}{\kappa^2} + {l_{op}}^2\right)}{2l_{op}}\right)^2\right)^{\frac{1}{2}} \end{cases} \tag{10}$$

Where, h is the height of the droplet derived from the geometric relationship in Figure 9B, m; κ is curvature of curved leaves, m^{-1}; θ_s is static contact angle between droplets and leaf surface, °; A_P is the windward area of the droplet is derived based on the geometric relationship in Figure 9C, m^2; l_{op} is the distance between the center of the circle on the surface of a leaf and the center of mass of a spherical droplet, m; $y(x)$ is the projection curve of the spatial curve where the droplet intersects with the leaf surface in the Oxy plane as shown in Figure 9D.

Furthermore, the functional expressions for spherical droplets and cylindrical leaf surfaces are:

$$\begin{cases} x^2 + y^2 + z^2 = {R_1}^2 \\ y^2 + (z + l_{op})^2 = \frac{1}{\kappa^2} \end{cases} \tag{11}$$

The hysteretic tension (F_σ) is calculated by considering the force on the contact line (Fan et al., 2011). As shown in Figure 9A, this force is calculated by integrating the force generated by surface tension projected onto the Oxy plane and around the spatial curve where the droplet intersects the surface of the cylindrical leaf:

$$F_\sigma = \oint_L \sigma_{la} \cdot \cos \theta(\alpha, \beta) \cdot R_1 \cdot \cos \beta \cdot \cos \alpha ds \tag{12}$$

Where, L is the spatial curve where the droplet intersects with the leaf surface; $\theta(\alpha, \beta)$ is the contact angle that constantly changes along the position of the spatial curve L, °; α is the angle between the projection of the line connecting the point on the spatial curve L and the origin O on the Oxy plane and the x-axis, °; β is the angle between the line connecting the point on the spatial curve L and the origin O and the Oxy projection surface, °.

Based on Equations 9, 12, it can be concluded that as the droplet radius increases, both the aerodynamic force and hysteresis tension exhibit an upward trend. However, the increase in hysteresis tension is not as pronounced as that of the aerodynamic force. This disparity arises because hysteresis tension is mainly influenced by adhesive forces and contact angle hysteresis between the droplet and surface. While a larger radius increases the contact area, this effect is limited as hysteresis tension depends more on surface characteristics. In contrast, aerodynamic force is proportional to the upwind area, which scales with the square of the droplet radius, making even a small increase in radius significantly boost the aerodynamic force. Based on this analysis, larger droplets are

10.3389/fpls.2024.1450831

more prone to sliding or detaching from the leaf surface. Additionally, as increased curvature reduces the contact area between the droplet and surface, weakening the hysteresis tension. In contrast, aerodynamic force is less dependent on contact area and more influenced by the upwind area of the droplet, making curvature have a smaller impact. As a result, higher curvature makes it easier for droplets to slide or detach.

Furthermore, to verify the conclusions drawn from the above theory, the critical wind speed values at which droplets begin to move on various curved leaf surfaces were measured experimentally, as shown in Figure 10. The error bars in the figure represent the variation in results due to repeated experiments.

As shown in Figure 10, the critical airflow velocity on the leaf decreases as the initial droplet radius increases, consistent with the findings of Milne and Amirfazli (2009) and Barwari et al. (2019). Experiments have shown that droplets with an initial radius of 2.006 mm require about 13.6~17.5% lower critical airflow velocity to start moving compared to droplets with an initial radius of 1.419 mm. This is because larger droplets have greater initial heights and larger front-end areas, requiring lower wind speeds to initiate movement. Equation 10 also indicates a negative correlation between the windward area (A_p) and the critical wind speed. The critical wind speed not only reflects the droplets' ability to withstand the maximum wind load on the leaf but also measures the maximum retention capacity of droplets on the leaf's surface.

For a constant size droplet, as the curvature κ of the leaf increases from 0.013 to 0.041 mm^{-1}, the critical wind speed at which the droplet begins to move gradually increases, indicating that the greater the leaf bending deformation, the easier it is for the droplet to move. Specifically, the critical airflow velocity for droplets on leaves with a curvature of 0.041 mm^{-1} is about 18.5~24.8% lower than that on flat leaves. This airflow velocity results from the synergistic effect of the static contact angle and contact angle hysteresis. The static contact angle determines the original shape of the droplets fixed on curved capsicum leaves: as the static contact angle increases, the contact area between the droplets and the leaves decreases, and the droplet height increases. The contact angle hysteresis determines the critical shape of the droplet at the critical airflow velocity. For the capsicum leaves used in this experiment, the static contact angle (θ_s)and contact angle hysteresis ($\Delta\theta_{CAH}$) exhibit the following relationships: θ_s (κ = 0.041 mm^{-1})< θ_s (κ = 0.027 mm^{-1})< θ_s (κ = 0.013 mm^{-1})< θ_s (κ = 0) and $\Delta\theta_{CAH}$ (κ = 0.041 mm^{-1}) > $\Delta\theta_{CAH}$ (κ = 0.027 mm^{-1}) > $\Delta\theta_{CAH}$ (κ = 0.013 mm^{-1}) > $\Delta\theta_{CAH}$ (κ = 0). Therefore, the critical airflow velocity at the beginning of the droplet movement is highest on flat leaves. It can be inferred that the critical airflow velocity at which a droplet begins to move on a leaf is positively correlated with the static contact angle and negatively correlated with the hysteresis of the contact angle. This is exactly opposite to the conclusion drawn by Wang et al. (2022) on a planar substrate. The analysis suggests that there is a significant difference between droplets fixed on curved leaves and those on flat surfaces before they begin to move. Before a droplet fixed on a curved leaf begins to move, its backward contact line undergoes an "*uphill*" motion, while its forward contact line undergoes a "*downhill*" motion. This behavior is not observed in droplets moving on a flat surface. When a droplet on a curved leaf is subjected to airflow, it oscillates, and its center of mass will not remain directly above the leaf. As the droplet's center of gravity shifts, gravity will positively influence its movement, thus affecting the critical wind speed at which the droplet begins to move.

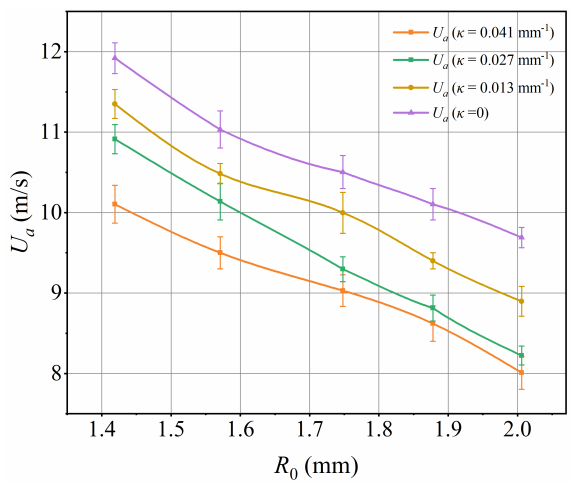


FIGURE 10
Measurement of critical wind speed at the onset of droplet motion.

3.3.2 The motion behavior of droplets on different curved leaf surfaces

The motion of droplets can be quantified by their position (x_{drop}), and velocity (v_{drop}). The value of x_{drop}(t) is determined by the midpoint position of the arc length between the forward and backward contact lines. The velocity (v_{drop}), is obtained by numerically differentiating x_{drop}(t) over time (v_{drop} = dx_{drop}/dt), using a time interval of 0.25 s. When the droplet begins to move forward, the speed of the forward and backward contact lines cannot be consistent, resulting in different degrees of elongation of the droplet on leaves with different curvature. Figure 8 illustrates the motion behavior of a droplet with an initial radius R_0 = 2.006 mm on different curved capsicum leaves. x_a represents the displacement of the droplet forward contact line relative to the initial position, while x_r represents the displacement of the droplet backward contact line relative to the initial position, both expressed in arc length. (x_a - x_r)/2R_0 is defined as the dimensionless shape variable of the droplet. The measurement method for the key characterization parameters of droplet motion is shown in Figure 8A.

Figure 8B shows that when a droplet with an initial radius R_0 = 2.006 mm moves on a leaf with a curvature κ = 0.041 mm^{-1}, the droplet dimensionless shape variable (Δx^*) ranges from 0.01 to 0.24. When a droplet moves on a leaf with this curvature, it produces the largest shape variable. Overall, the greater the curvature of the leaf, the larger the droplet shape variable. This is because the gravity of the droplet itself generates additional tension, making it easier for the droplet to elongate along the bending direction. Different curved leaf surfaces affect the contact angle distribution of droplets, leading to varying forward and backward contact angles, and thus differing

10.3389/fpls.2024.1450831

degrees of elongation along the curved leaf surface. In the context of plant protection spray processes, greater bending deformation of the leaf results in more significant droplet elongation. This can cause multiple droplets fixed on the leaf to converge into larger droplets, increasing the risk of droplet loss. When the curvature of the leaf is $\kappa \leq 0.027$ mm^{-1}, the minimum droplet dimensionless shape variable shows a negative value. This is because the droplet backward contact line is more likely to move on a leaf with a smaller curvature.

3.3.3 The motion behavior of droplets of different sizes on leaves

The size of droplets affects the balance between droplet hysteresis tension and aerodynamic forces. Large droplets have a greater mass and are more challenging to move under aerodynamic driving; however, due to their larger windward area, they are subjected to greater aerodynamic forces. This equilibrium relationship impacts the shape and velocity of droplets on the surface. According to Equation 10, it is confirmed that the larger

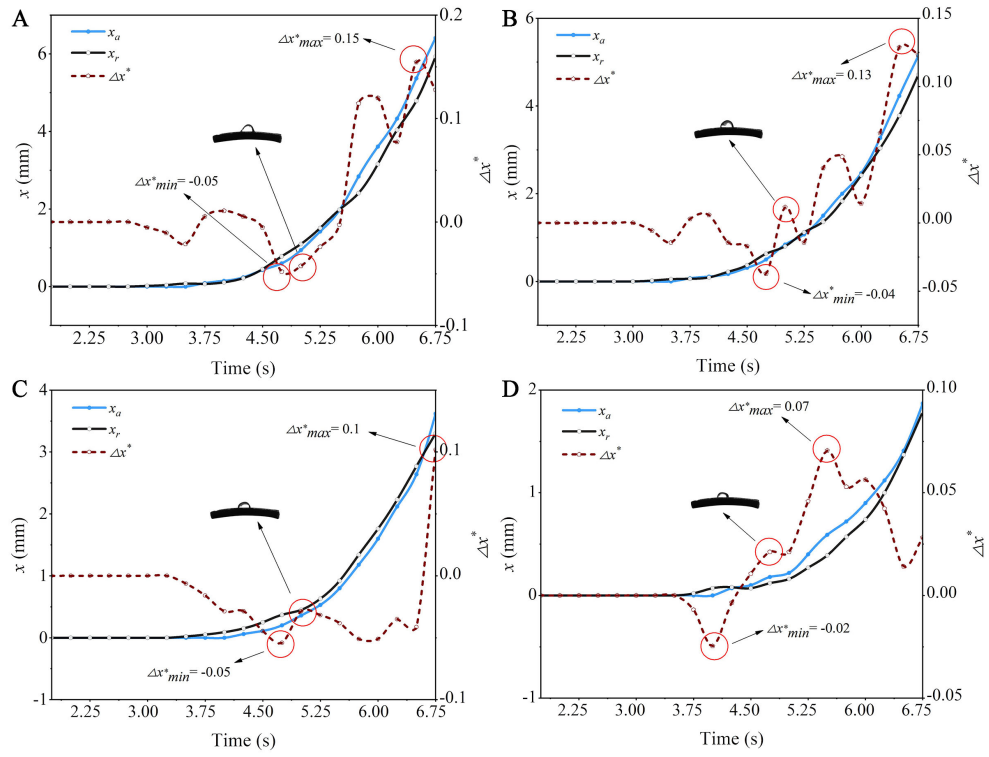


FIGURE 11
The variation of contact line positions of droplets of different sizes on a leaf with κ = 0.02 mm^{-1} over time: **(A)** R_0 = 1.878 mm; **(B)** R_0 = 1.748 mm; **(C)** R_0 = 1.571 mm; **(D)** R_0 = 1.419 mm.

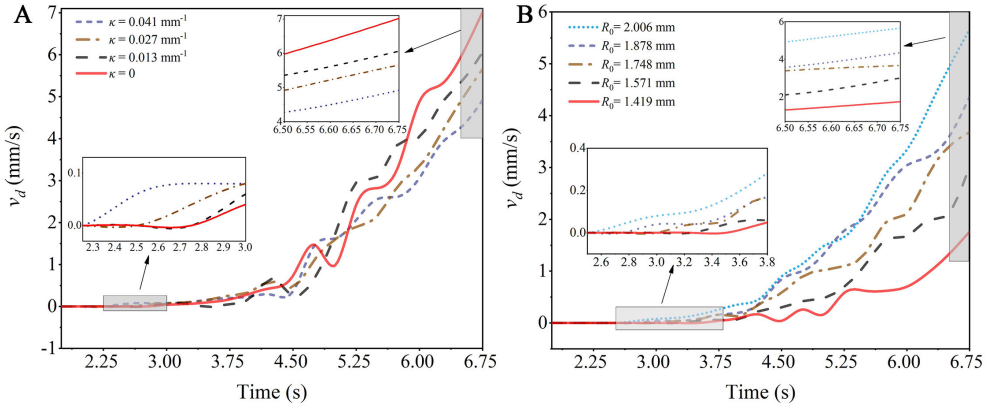


FIGURE 12
Schematic diagram of the velocity changes of droplets over time: **(A)** Leaves with different curvatures; **(B)** Droplets of different sizes.

the droplet radius (R_1), the smaller the corresponding hysteresis tension, and consequently, the lower the wind speed required to initiate droplet movement. This relationship is further validated and quantified through experiments. When the curvature of the leaf is constant, the motion behavior of droplets on the curved capsicum leaf, as the initial radius of the droplet increases, is shown in Figure 11.

Figure 8C shows the motion process of a droplet with an initial radius R_0 = 2.006 mm on a leaf with a curvature κ =0.027 mm^{-1}, with the droplet dimensionless shape variable (Δx^*) ranging from -0.07 to 0.18. Comparison of Figure 11 reveals that droplets with an initial radius R_0 = 2.006 mm exhibit the largest deformation when moving on leaves with a curvature κ =0.027 mm^{-1}, indicating that larger droplet sizes correspond to greater deformations. The analysis suggests that the larger the droplet, the more its contact line with the leaf surface expands, increasing the contact area between the droplet and the leaf, thereby raising the droplet shape variable. Overall, for droplets of different sizes, the initial droplet dimensionless shape variables are all negative, which results from the droplet backward contact line starting to move first.

3.3.4 The velocity changes of droplets over time

Figure 12A illustrates the velocity changes of a droplet with an initial radius R_0 = 2.006 mm moving on capsicum leaves with varying degrees of curvature over the test time span of 0 to 6.75 s. It is evident that the greater the curvature of the leaf, the earlier the droplet begins to move, reinforcing the conclusion that the curvature of the leaf is negatively correlated with the critical wind speed of the droplet. In general, the droplet moves with near-uniform acceleration on the curved leaf, Overall consider the accelerated motion of the droplet on curved leaves as uniformly accelerated motion, and reaches the velocity maximum at 6.75 s, the acceleration of the droplets on the leaf with curvature of 0.041 mm^{-1} is about 30% lower than that on the flat leaf. This indicates that while droplets are more prone to movement on leaves with greater bending deformation, their movement speed is relatively low, and the acceleration is also relatively modest.

Figure 12B shows the velocity changes of droplets of different sizes on a leaf with a curvature κ =0.027 mm^{-1} over the test time span of 0 to 6.75 s. It can be observed that larger droplets begin to move earlier, confirming the conclusion that droplet size is negatively correlated with the critical wind speed of the droplet. Overall, droplets of different sizes accelerate on the leaf, reaching their maximum velocity at 6.75 s. The acceleration of the droplets with an initial radius R_0 = 1.419 mm is about 68.9% lower than that of droplets with an initial radius R_0 = 2.006 mm. This indicates that larger droplets are more prone to movement on the leaves, and their movement speed is relatively faster with a more significant increase.

4 Conclusion

This study utilized wind tunnels and high-speed camera systems to obtain the bending deformation of capsicum leaves under airflow, quantified the bending deformation of the leaves through curvature κ. The experimental results indicate that at low wind speeds, bending deformation primarily occurs near the petiole and the tip of the leaf. At high wind speeds, the deformation shifts to areas farther from the leaf tip. Overall, the curvature κ during leaf bending deformation is predominantly below 0.05 mm^{-1}.

Studied droplet behavior on curved capsicum leaves and developed a mechanical analysis model for droplet motion, which proved that the difficulty of droplet motion is negatively correlated with particle size and leaf curvature. The results of the critical wind speed measurement test show that the difference in critical wind speed between different particle sizes can reach 17.5%, and between droplets on various curved leaves varies by about 24.8%. Considering spray droplet parameters and bending deformation of the leaf aids in understanding maximum droplet retention capacity during air-assisted spraying.

During droplet movement, both the curvature of the leaf and the droplet size significantly influence the droplet shape variable. Within the test duration, the maximum dimensionless shape variable on various curved leaves reaches 0.24, and the acceleration can differ by about 30%. For droplets of different sizes, the maximum dimensionless shape variable is 0.18, the acceleration can differ by about 68%. This distinct motion behavior of droplets on different curved leaves helps to understand droplet-leaf interactions and provides new insights for optimizing pesticide application.

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 in the article/Supplementary Material.

Author contributions

Z-MG: Data curation, Software, Writing – original draft, Writing – review & editing, Conceptualization, Investigation, Methodology, Validation. WH: Methodology, Writing – review & editing, Investigation, Software, Validation. X-YD: Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. X-YZ: Data curation, Investigation, Methodology, Software, Validation, Writing – review & editing. SW: Conceptualization, Methodology, Writing – review & editing, Investigation, Software, Validation. JC: Investigation, Software, Validation, Writing – review & editing. B-JQ: Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The present study was supported by grants from the National Science Foundation of China (31971790), Primary Research & Development Plan of Jiangsu Province (BE2020328) and Priority

Academic Program Development of Jiangsu Higher Education Institutions (PAPD-2018-87).

Conflict of interest

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

Generative AI statement

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1450831/full#supplementary-material

References

Barwari, B., Burgmann, S., Bechtold, A., Rohde, M., and Janoske, U. (2019). Experimental study of the onset of downstream motion of adhering droplets in turbulent shear flows. *Exp. Thermal Fluid Sci.* 109, 109843. doi: 10.1016/j.expthermflusci.2019.109843

Brendan, J. C. (1984). The equilibrium of liquid drops on smooth and rough circular cylinders. *J. Colloid Interface Sci.* 97, 195–200. doi: 10.1016/0021-9797(84)90286-8

Butts, T. R., Virk, S. S., and Kouame, K. B. (2024). Droplet size, velocity, and spray coverage from a magnetic-assisted sprayer. *Biosyst. Eng.* 245, 1–11. doi: 10.1016/j.biosystemseng.2024.06.010

de Langre, E. (2019). Plant vibrations at all scales: a review. *J. Exp. Bot.* 70, 3521–3531. doi: 10.1093/jxb/erz209

Ding, H., Gilani, M. N. H., and Spelt, P. D. M. (2010). Sliding, pinch-off and detachment of a droplet on a wall in shear flow. *J. Fluid Mechanics* 644, 217–244. doi: 10.1017/S0022112009992217

Ding, H., and Spelt, P. D. M. (2008). Onset of motion of a three-dimensional droplet on a wall in shear flow at moderate reynolds numbers. *J. Fluid Mechanics.* 599, 341–362. doi: 10.1017/S0022112008000190

Dorr, G. J., Wang, S. S., Mayo, L. C., McCue, S. W., Alison Forster, W., Hanan, J., et al. (2015). Impaction of spray droplets on leaves: influence of formulation and leaf character on shatter, bounce and adhesion. *Exp. Fluids* 56, 143. doi: 10.1007/s00348-015-2012-9

Fan, J., Wilson, M. C. T., and Kapur, N. (2011). Displacement of liquid droplets on a surface by a shearing air flow. *J. Colloid Interface Sci.* 356, 286–292. doi: 10.1016/j.jcis.2010.12.087

Gart, S., Mates, J. E., Megaridis, C. M., and Jung, S. (2015). Droplet impacting a cantilever: A leaf-raindrop system. *Phys. Rev. Appl.* 3, 44019. doi: 10.1103/PhysRevApplied.3.044019

Gaskin, R. E., Steele, K. D., and Forster, W. A. (2005). Characterising plant surfaces for spray adhesion and retrntion. *Plant Protect* 58, 179–183. doi: 10.30843/nzpp.2005.58.4244

Hu, H., Huang, S. H., and Chen, L. B. (2013). Displacement of liquid droplets on micro-grooved surfaces with air flow. *Exp. Thermal Fluid Sci.* 49, 86–93. doi: 10.1016/j.expthermflusci.2013.04.005

Ma, J., Liu, K., Dong, X. Y., Huang, X., Ahmad, F., and Qiu, B. J. (2023). Force and motion behaviour of crop leaves during spraying. *Biosyst. Eng.* 235, 83–99. doi: 10.1016/j.biosystemseng.2023.09.012

Mahato, L. K., and Mandal, D. K. (2021). Role of resisting shear on the shedding drops. *Colloid Interface Sci. Commun.* 41, 100388. doi: 10.1016/j.colcom.2021.100388

Mayo, L. C., McCue, S. W., Moroney, T. J., Alison Forster, W., Kempthorne, D. M., Belward, J. A., et al. (2015). Simulating droplet motion on virtual leaf surfaces. *R. Soc. Open Sci.* 2, 140528. doi: 10.1098/rsos.140528

McHale, G., Rowan, S. M., Newton, M. I., and Käb, N. A. (1999). Estimation of contact angles on fibers. *J. Adhesion Sci. Technol.* 13, 1457–1469. doi: 10.1163/156856199X00587

Milne, A. J., and Amirfazli, A. (2009). Drop shedding by shear flow for hydrophilic to superhydrophobic surfaces. *Langmuir* 25, 14155–14164. doi: 10.1021/la901737y

Mitchell, A. J., Duffy, B. R., and Wilson, S. K. (2022). Unsteady coating flow on a rotating cylinder in the presence of an irrotational airflow with circulation. *Phys. Fluids* 34, 043105. doi: 10.1063/5.0084472

Mortazavi, M., and Jung, S. Y. (2023). Role of droplet dynamics on contact line depinning in shearing gas flow. *Langmuir* 39, 10301–10311. doi: 10.1021/acs.langmuir.3c00065

Moulia, B., Fournier, M., and Guitard, D. (1994). Mechanics and form of the maize leaf: in *vivo* qualification of flexural behaviour. *J. Materials Sci.* 29, 2359–2366. doi: 10.1007/BF00363427

Oqielat, M. N., Turner, I. W., Belward, J. A., and McCue, S. W. (2011). Modelling water droplet movement on a leaf surface. *Mathematics Comput. Simulation* 81, 1553–1571. doi: 10.1016/j.matcom.2010.09.003

Paterson, C., Wilson, S. K., and Duffy, B. R. (2014). Rivulet flow round a horizontal cylinder subject to a uniform surface shear stress. *Q. J. Mechanics Appl. Mathematics* 67, 567–597. doi: 10.1093/qjmam/hbu018

Paterson, C., Wilson, S. K., and Duffy, B. R. (2015). Strongly coupled interaction between a ridge of fluid and an inviscid airflow. *Phys. Fluids* 27, 072104. doi: 10.1063/1.4926623

Shao, C. P., Chen, Y. J., and Lin, J. Z. (2012). Wind induced deformation and vibration of a Platanus acerifolia leaf. *Acta Mechanica Sin.* 28, 583–594. doi: 10.1007/s10409-012-0074-y

Sugiyama, K., and Sbragaglia, M. (2008). Linear shear flow past a hemispherical droplet adhering to a solid surface. *J. Eng. mathematics* 62, 35–50. doi: 10.1007/s10665-007-9185-z

Sullivan, J. M., Paterson, C., Wilson, S. K., and Duffy, B. R. (2012). A thin rivulet or ridge subject to a uniform transverse shear stress at its free surface due to an external airflow. *Phys. Fluids* 24, 082109. doi: 10.1063/1.4744980

Sullivan, J. M., Wilson, S. K., and Duffy, B. R. (2008). A thin rivulet of perfectly wetting fluid subject to a longitudinal surface shear stress. *Q. J. Mechanics Appl. Mathematics* 61, 25–61. doi: 10.1093/qjmam/hbm023

Tredenick, E. C., Forster, W. A., Pethiyagoda, R., van Leeuwen, R. M., and McCue, S. W. (2021). Evaporating droplets on inclined plant leaves and synthetic surfaces: Experiments and mathematical models. *J. colloid Interface Sci.* 592, 329–341. doi: 10.1016/j.jcis.2021.01.070

Veremieiev, S., Brown, A., Gaskell, P. H., Glass, C. R., Kapur, N., and Thompson, H. M. (2014). Modelling the flow of droplets of bio-pesticide on foliage. *Interfacial Phenomena Heat Transfer* 2, 1–14. doi: 10.1615/InterfacPhenomHeatTransf

Wang, S. S., Chang, S. N., Zhao, H. Y., and Yang, C. (2022). Dynamic behaviors of water droplet moving on surfaces with different wettability driven by airflow. *Int. J. Multiphase Flow* 154, 104127. doi: 10.1016/j.ijmultiphaseflow.2022.104127

Wang, H., Mucha, P. J., and Turk, G. (2005). Water drops on surfaces. *ACM Trans. Graph.* 24, 921–929. doi: 10.1145/1073204.1073284

Wang, Z. C., Zheng, C., Li, T., and He, X. K. (2021). Analysing the preference for pesticide spray to be deposited at leaf-tips. *Biosyst. Eng.* 204, 247–256. doi: 10.1016/j.biosystemseng.2021.01.012

Wilson, J. L., Pahlavan, A. A., Erinin, M. A., Duprat, C., Deike, L., and Stone, H. A. (2023). Aerodynamic interactions of drops on parallel fibres. *Nat. Phys.* 19, 1667–1672. doi: 10.1038/s41567-023-02159-4

Wilson, S. K., Sullivan, J. M., and Duffy, B. R. (2011). The energetics of the breakup of a sheet and of a rivulet on a vertical substrate in the presence of a uniform surface shear stress. *J. Fluid Mechanics* 674, 281–306. doi: 10.1017/S0022112010006518

Zabkiewicz, J. A. (2007). Spray Equationtion efficacydholistic and futuristic perspectives. *Crop Prot.* 26, 312–319. doi: 10.1016/j.cropro.2005.08.019

Zhang, C., Zhou, H., Xu, L., Ru, Y., Ju, H., and Chen, Q. (2022). Measurement of morphological changes of pear leaves in airflow based on high-speed photography. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.900427

Zheng, H., Sun, H., Cao, Y. B., Lv, X. L., Wang, C. X., Chen, Y. F., et al. (2023). Computational fluid dynamics simulation analysis of the effect of curved rice leaves on the deposition behaviour of droplets. *Plant Methods* 19, 116. doi: 10.1186/s13007-023-01082-2

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 24 October 2024
DOI 10.3389/fpls.2024.1435016

Check for updates

OPEN ACCESS

EDITED BY
Jun Ni,
Nanjing Agricultural University, China

REVIEWED BY
Lingxian Zhang,
China Agricultural University, China
José Luis Hernández-Hernández,
Chilpancingo Institute of Technology, Mexico
Ke Xu,
Anhui Polytechnic University, China

*CORRESPONDENCE
Hongyan Zhu
hyzhu-zju@foxmail.com
Jun-Li Xu
junli.xu@ucd.ie
Yong He
yhe@zju.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 19 May 2024
ACCEPTED 30 September 2024
PUBLISHED 24 October 2024

CITATION
Zhu H, Lin C, Liu G, Wang D, Qin S, Li A, Xu J-L and He Y (2024) Intelligent agriculture: deep learning in UAV-based remote sensing imagery for crop diseases and pests detection.
Front. Plant Sci. 15:1435016.
doi: 10.3389/fpls.2024.1435016

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

Intelligent agriculture: deep learning in UAV-based remote sensing imagery for crop diseases and pests detection

Hongyan Zhu[1,2*†], Chengzhi Lin[1,2†], Gengqi Liu[1,2], Dani Wang[1,2], Shuai Qin[1,2], Anjie Li[1,2], Jun-Li Xu[3*] and Yong He[4*]

[1]Guangxi Key Laboratory of Brain-inspired Computing and Intelligent Chips, School of Electronic and Information Engineering, Guangxi Normal University, Guilin, China, [2]Key Laboratory of Integrated Circuits and Microsystems (Guangxi Normal University), Education Department of Guangxi Zhuang Autonomous Region, Guilin, China, [3]School of Biosystems and Food Engineering, University College Dublin, Dublin, Ireland, [4]College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China

Controlling crop diseases and pests is essential for intelligent agriculture (IA) due to the significant reduction in crop yield and quality caused by these problems. In recent years, the remote sensing (RS) areas has been prevailed over by unmanned aerial vehicle (UAV)-based applications. Herein, by using methods such as keyword co-contribution analysis and author co-occurrence analysis in bibliometrics, we found out the hot-spots of this field. UAV platforms equipped with various types of cameras and other advanced sensors, combined with artificial intelligence (AI) algorithms, especially for deep learning (DL) were reviewed. Acknowledging the critical role of comprehending crop diseases and pests, along with their defining traits, we provided a concise overview as indispensable foundational knowledge. Additionally, some widely used traditional machine learning (ML) algorithms were presented and the performance results were tabulated to form a comparison. Furthermore, we summarized crop diseases and pests monitoring techniques using DL and introduced the application for prediction and classification. Take it a step further, the newest and the most concerned applications of large language model (LLM) and large vision model (LVM) in agriculture were also mentioned herein. At the end of this review, we comprehensively discussed some deficiencies in the existing research and some challenges to be solved, as well as some practical solutions and suggestions in the near future.

KEYWORDS

intelligent agriculture (IA), deep learning (DL), crop diseases and pests, remote sensing (RS), unmanned aerial vehicle (UAV)

1 Introduction

Crop diseases and pests are the major natural disasters affecting agricultural production. They are the main factors restricting high yield, high quality, high efficiency, ecology, and agriculture safety. The food and agriculture organization (FAO) of the United Nations reports that pests account for an annual loss of approximately 10-15% in global crop yields, totaling over 300 million tons (Qi et al., 2019). Crop diseases and pests have the characteristics of many kinds, significant influence, and break out easily, which will cause the decline of crop quality and yield. Therefore, in actual agricultural production, the rapid acquisition of crop pest information and early detection of pests and diseases is critical for achieving crop yield increase and reducing disease losses (Thangaraj et al., 2021). It is also an essential basis for the implementation of intelligent agriculture (IA).

Traditional detection methods of crop diseases and pests mainly rely on artificial visual evaluation, serology, and molecular biology-based technical means. Examples of these technical means include flow cytometry, enzyme-linked immunosorbent assay (ELISA), immunofluorescence (IF), polymerase chain reaction (PCR), and fluorescence *in situ* hybridization (FISH) (Carvajal-Yepes et al., 2019). Although these crop disease detection technologies can accurately diagnose crop diseases, they have some drawbacks. These drawbacks include being time-consuming, inefficient, and destructive, requiring detailed sampling and processing procedures, requiring technicians with strong professional knowledge and skills, and being greatly affected by human factors. It is challenging to detect large-scale full-coverage field crop diseases and pests effectively, which limits the development of IA and crop breeding. Therefore, how to quickly, accurately, and efficiently monitor the occurrence of crop diseases and pests in a wide range and timely prevent and control early conditions has become a significant problem in crop production. In addition, how to minimize the losses caused by pests and diseases to crops is also an essential issue that needs to be addressed.

At present, the detection of crop diseases is divided into several types. These types include indoor fine detection based on leaf detection, agricultural machinery as the main platform for a single plant or small-scale crop canopy field fine detection, aircraft monitoring platforms for crop canopy farmland plot scale detection, and satellite images as the data source of regional scale detection (Zhang et al., 2019). Ground-level RS platforms including indoor and field fine detection scales have the characteristics of a controllable environment and high accuracy. However, they cannot obtain crop disease information in an extensive range (Liu et al., 2020). Satellite-level RS has some shortcomings, such as low spatial and temporal resolution, poor timeliness, and low accuracy. The low-altitude RS platform represented by UAVs has become a new means for obtaining crop disease information at the scale of farmland plots under its advantages of flexibility, low cost, and suitability for complex farmland environments (Mukherjee et al., 2019). The UAV platforms can carry a variety of sensors, including RGB cameras, multi-spectral cameras, infrared thermal cameras, hyperspectral cameras, laser radar, etc (Yang et al., 2017). Among these sensors, RGB cameras, multi-spectral cameras, and other spectral imaging sensors are widely used to acquire farmland crop disease information due to their lightweight, low-cost, and easy operation. After diseases attack crops, their color, texture, and spectral characteristics will change to a certain extent, and the effects of different diseases on crops are also disparate. Accordingly, the limitations of traditional detection methods, such as their reliance on manual labor and susceptibility to human error, are significantly addressed by UAV remote sensing technology, which enables rapid, accurate, and scalable monitoring of crop health, thereby fostering new opportunities for intelligent agriculture.

The unmanned aerial systems consist of different integrating sensors (high-resolution RGB, multispectral, hyperspectral, LiDAR, and thermal), Internet connectivity, flight missions, data collection, image processing, and AI algorithms (García et al., 2020). One of the most common applications is the assessment of crop health through RS and image processing (Radoglou-Grammatikis et al., 2020), which is the focus of this paper. UAV is remotely controlled by an operator and can carry a variety of cameras such as multi-spectral and hyper-spectral, thus acquiring aerial images, providing a wealth of information on crop growth, health status, and environmental conditions. However, the full potential of UAV-collected data can only be realized through sophisticated image processing and analysis techniques (Peñuelas et al., 1995). Deep learning, a powerful branch of machine learning, has demonstrated remarkable capabilities in analyzing and interpreting complex image data. By leveraging deep neural networks, deep learning algorithms can extract meaningful features from raw imagery, enabling precise classification and detection of crop diseases and pests.

This paper focuses on UAV-based hardware devices and imagery process methods that are used in crop diseases and pests. The apparatus comprises standard sensor types, featuring AI-driven processing of UAV RS images employing traditional ML algorithms as well as emerging DL algorithms, to achieve larger scale, faster, and higher accurate surveillance and management of crop diseases and pests. To sum up, we made some contributions:

1. Finding the hottest topics in IA related to 'UAV', 'remote sensing' and 'deep learning'. A presentation of keyword co-occurrence analysis, an authors' co-occurrence analysis, and geographical relations.
2. An explanation of crop diseases and pests detection. A synopsis of typical types of UAVs and sensors. As well as the technology road-map of the RS system.
3. A summary of the common vegetation indices and texture characteristics, application of the wavelength selection algorithms which were widely used in features extraction.
4. A review of some examples in crop detection using traditional ML methods and prominent DL methods in recent years. A brief description of these algorithms and an elaborate discussion of several detailed areas in applying UAV-based RS for crop protection.
5. Proposing some deficiencies in the existing research and challenges to be solved. A prospect of the future of deep learning and artificial general intelligence (AGI) in intelligent agriculture.

2 Finding the hot-spots by bibliographic analysis

Using the literature data from the core database of the Web of Science (WOS), we searched the literature on keywords including 'remote sensing' and 'deep learning' from January 1, 2018, to January 1, 2024, to determine the current international mainstream methods for monitoring crop diseases and pests. To avoid duplication and messy data, we manually cleaned up the papers that did not meet the research theme and finally selected 13738 papers for analysis. Herein, we extracted the abstracts, keywords, authors, countries, journal names, and other information, used VOSviewer (version 1.16.18, Leiden University, Leiden, The Netherlands) to build a database of the extracted data, analyzed the article through clustering, principal component analysis (PCA), and different algorithms to present the results in visual charts (Yang et al., 2020).

2.1 Keyword co-occurrence analysis

Firstly, we used the clustering algorithm of VOSviewer to cluster and determine the relationship between the subjects in the selected literature set by the co-occurrence of words or noun phrases in the literature set to obtain the keyword co-occurrence analysis between the core words, including 'UAV (UAVs/UAS/UAV-based platforms/drones) ', 'remote sensing' and 'intelligent agriculture'. As shown in Figure 1, the research hot-spots is 'UAV', 'remote sensing' and 'deep learning' from 2018 to 2024. It can be seen that UAV low-altitude RS has a very close cross-connection with IA, especially in crop diseases and pests. The UAV RS system can effectively use vegetation indices to estimate biophysical parameters and generate water stress detection images of leaf area index, chlorophyll content, photochemical reflectance index, and canopy temperature, which is vital for monitoring and prevention of crop diseases and pests (Ren et al., 2020). We clustered the analysis results and used three colors to represent the three clusters. The blue clusters are UAV-centered for correlation analysis, and similarly, the red ones are for remote sensing and the green ones are for deep learning.

2.2 Author co-occurrence analysis

Subsequently, we focused on the paper's authors and the countries for data statistics and analysis. Statistic methods for instance factor analysis, cluster analysis, and PCA were used to judge the research similarity of the literature of two different authors. It is assumed that more than three authors cite the literature of the same two authors, and the citation frequency is high. In that case, it proves that the academic research relationship between the two authors is relatively close. The authors of the selected kinds of literature were classified through VOSviewer. The visualization method showed the academic relationship between the authors who form a scientific community in 'UAV low-altitude RS ' and 'crop diseases detection'. As illustrated in Figure 2, Guijun Yang (the National Engineering Research Center for Information Technology in Agriculture), Yubin Lan (the College of Engineering, South China Agricultural University/National Center for International Collaboration Research on Precision Agricultural Aviation Pesticides Spraying Technology), Yong He (the College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou), and other authors published many papers on the application of UAV RS in IA from 2018 to 2024 and had close academic relations with many authors.

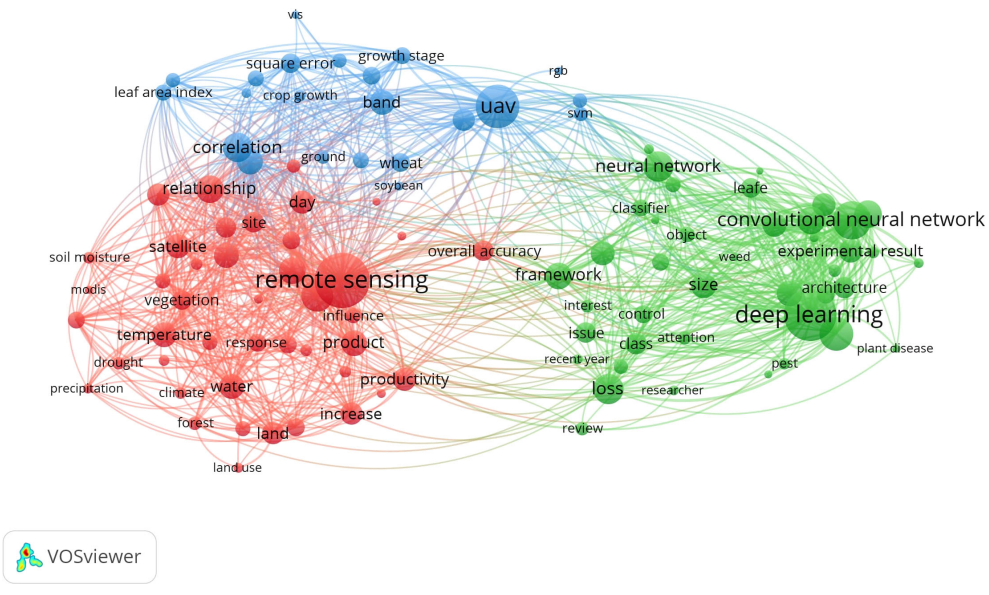


FIGURE 1
The keyword co-occurrence analysis results.

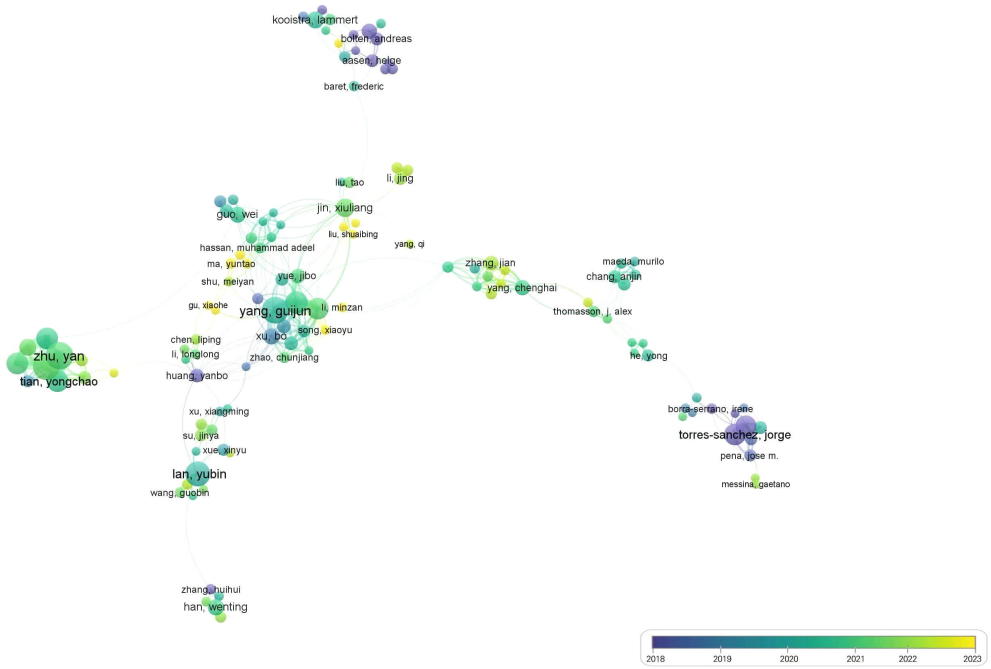


FIGURE 2
The authors' co-occurrence analysis results.

2.3 Analysis of the geographical relationship

After that, we analyzed the geographical relationship of the authors and used VOSviewer to analyze the degree of academic cooperation among countries. Then we got the national cooperation network map, as shown in Figure 3. Moreover, we also statistically analyzed the number of relevant papers published in the past six years. The top ten publishing countries are China, the United States, Australia, Spain, Brazil, Germany, Canada, Italy, the United Kingdom, and Mexico. Two major scientific communities have been formed: a 'cooperative relationship zone' dominated by Sino-US cooperation and radiating to Australia, Canada, Brazil, and other countries. The European academic cooperation circle is led by Spain, Germany, Italy, and the United Kingdom and radiates to Turkey, Iran, Israel, and other countries.

2.4 Publication and citation of articles on crop diseases and pests monitoring by UAV RS

According to the authors' co-occurrence analysis, from the co-occurrence map and national cooperation network formed by VOSviewer, low-altitude RS realized by UAVs has gradually become the most powerful tool in recent years (Abd El-Ghany et al., 2020). The application of UAV RS in IA, especially diseases and pest control, has achieved plentiful good results and has gradually become one of the most promising technologies to support integrated pest management (Xu et al., 2022).

The reasons for the wide use of UAVs are obvious. The bacterium, fungi, actinomyces, and other pathogens usually use soil, water, and wind as a medium to spread their spores to make crops sick. The dormancy of adults often causes infestations during overwintering in the previous year and mass breeding in the suitable season of the following year. After diseases and pests attack the crops, their morphological, color, and spectral characteristics (light absorption, reflection, transmission) will change greatly. Different types and degrees of diseases and pests caused by different crop growth conditions, also have more significant differences. Thus, to a certain extent, crop images and spectral information can reflect the occurrence and severity of crop diseases and pests. It also provides a theoretical basis for acquiring and analyzing information about crop diseases and pests by various imaging sensors on the UAV. Many scholars at home and abroad have used this theoretical basis to conduct experimental analysis and published many high-quality articles on crop diseases and pests with the help of UAV RS.

RS based on the UAV platforms has apparent advantages in obtaining pest information of field crops, for instance, high operation efficiency, high spatial-temporal resolution, synchronous image acquisition, and timely field sampling, and excellent structure and texture information. It can carry out rapid qualitative and quantitative research on much information about crop pests, which has been fully reflected in pests monitoring and classification of rice, wheat, corn, and other crops. The advantages of UAV RS provides a real-time and accurate 'ground-space' integrated platform for crop diseases and pests monitoring (Ahmad et al., 2022).

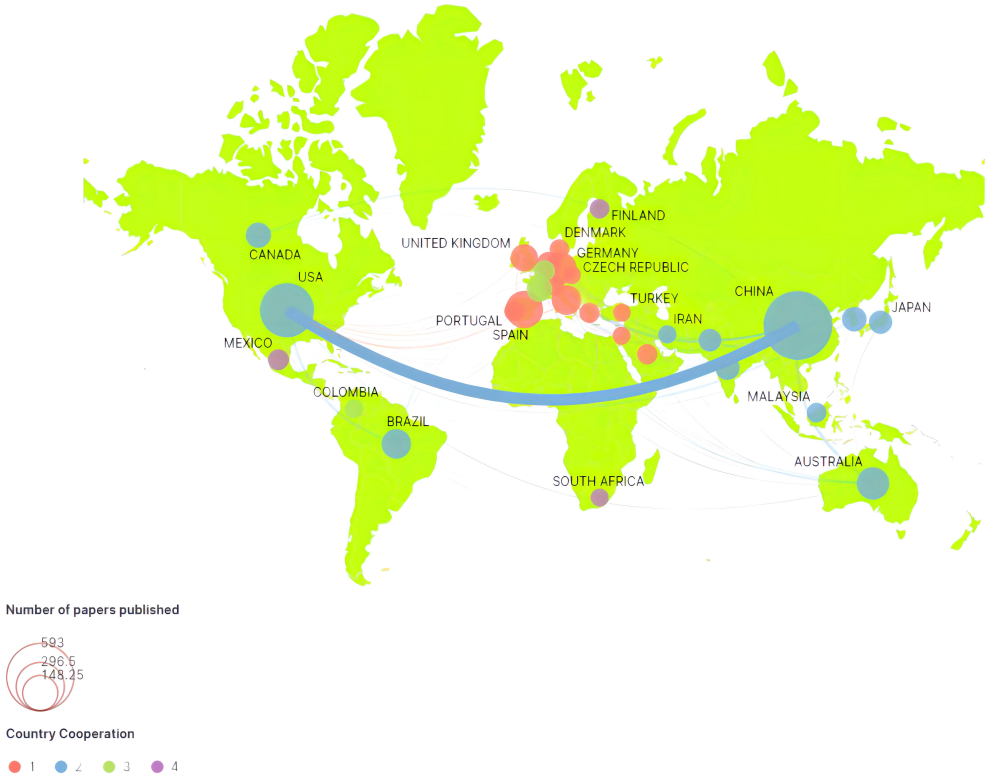


FIGURE 3
The geographical relations analysis results.

3 UAV RS in crop diseases and pests

Crop diseases and pests are one of the major agricultural disasters in China, which are characterized by many types, great influence, and frequent outbreaks, and their occurrence and severity often cause significant losses to our national economy, especially agricultural production. The following types of leaf pests and diseases (Zhou et al., 2023) are common: apple scab, black rot, cedar apple rust, rust, grape black rot, and strawberry leaf scorch. From sowing, and growing to harvesting, crops are often victimized by various pests (plant pathogens, pests, weeds and rodents, etc.), thus affecting the yield and quality of cultivated plants, due to the great variety of pests, their different forms, and their different patterns of occurrence. Therefore, it is very difficult to rely only on human experts to recognize pests and master their habits and characteristics. It is extremely important to rely on the methods of ML models and deep learning models to prevent and control pests.

The principles of disease control are: To depress the pathogen or inhibit its occurrence and spread; To improve the disease resistance of the host plant; To control or modify the environmental conditions so that they are favorable to the host plant and unfavorable to the pathogen, and to inhibit the occurrence and development of the disease. Generally, emphasis is placed on the prevention of plant populations, and integrated control measures are taken by the occurrence and development patterns of crop diseases according to location and time.

3.1 Definition of pest damage and crop diseases

Pest damage is a phenomenon that harmful insects cause damage to plant growth during the growth of a crop. Crop diseases are the stunting, wilting, or death of a plant body, usually caused by bacteria, fungi, viruses, algae, or unsuitable climate and soil, and is a natural disaster.

Diseases are categorized into two main groups: invasive and non-invasive. The classification of invasive diseases caused by pathogenic organisms is: (1) According to the pathogen is divided into fungal, bacterial, viral, and nematode diseases. (2) According to the host plant is divided into crops, vegetables, fruit tree diseases and forest diseases, etc., but also according to the type of crop is divided into wheat, rice, cotton, and other different crop diseases. (3) According to the symptoms, it can be divided into leaf spot disease, rot disease, wilt disease, and so on. (4) According to the site of disease, it can be divided into root disease, stem disease, leaf disease, fruit disease, and so on. (5) According to the mode of transmission, it can be divided into airborne, waterborne, soil-

borne, seedling-borne, insect-mediated transmission, and so on (Chaube and Pundhir, 2005).

3.2 Using UAV RS for efficient crop diseases and pests detection

Currently, the detection methods for plant disease mainly include sensory judgment, physical and chemical inspection, conventional machine vision, and other methods, which have high error rates and are easy to cause waste of pesticide spraying and environmental pollution (Zhao et al., 2013). These methods have a high error rate, and for instance, in large areas of tea plantations, disease identification is time-consuming and costly. Therefore, finding a fast and efficient identification method is of great significance to agricultural plant protection.

Crop diseased leaf image segmentation (Huang et al., 2021) is a difficult problem in the research of crop disease recognition methods based on image analysis and computer vision. It is a method that extracts the significant lesion areas of interest from the original lesion leaf images, eliminates the non-significant and unimportant areas, and highlights the important parts of the lesion images, which is conducive to the detection, diagnosis, and identification of crop diseases in the later stage of the disease.

Annotation methods play a crucial role in image analysis tasks such as classification, detection, and segmentation. Depending on the specific task, annotation methods can be divided into three categories: bounding box annotations for detection tasks, polygon annotations for segmentation tasks, and categorical labels for classification tasks. Bounding box annotations involve drawing rectangles around objects of interest, while polygon annotations outline the precise shape of objects. Categorical labels, on the other hand, assign a class label to each image or region of an image. By understanding the different types of annotations and how they relate to specific image analysis tasks, researchers can choose the most appropriate annotation method for their particular needs. As shown in Figure 4, there is a big difference between the crop images acquired using UAV and the lab images. Most UAV images are canopy images of the plant, containing a variety of data such as leaves, and stalks, and are non-destructive to the crop. Laboratory images require that diseased leaves be removed from the crop and photographed in a laboratory environment. The advantage of laboratory images is their high resolution, which is very useful for pest and disease classification. On the other hand, the resolution of UAV imagery is a big challenge. Imagery with insufficient resolution is difficult to categorize using AI algorithms. Therefore, the choice of the type of cameras and image processing algorithms carried by the UAV is crucial.

3.3 Advantages and application prospects

UAV RS technology in IA can quickly cover large areas of farmland, offering a more efficient and flexible approach compared to traditional ground surveys or manned aircraft remote sensing, which has the advantages of monitoring a wide area, in real-time, objective, high efficiency. It not only can effectively reduce the cost of manpower and material resources, but also facilitates a comprehensive grasp of the overall disaster situation, and puts forward more rapid and effective countermeasures, which in turn reduces the damage caused by pests and diseases to crop resources.

Addressing the UAV imagery resolution challenge, the integration of sensors, including high-resolution cameras and multispectral imaging devices, enhances the capabilities of UAVs. This integration empowers UAVs to capture detailed images with superior resolution, enabling precise identification and localization

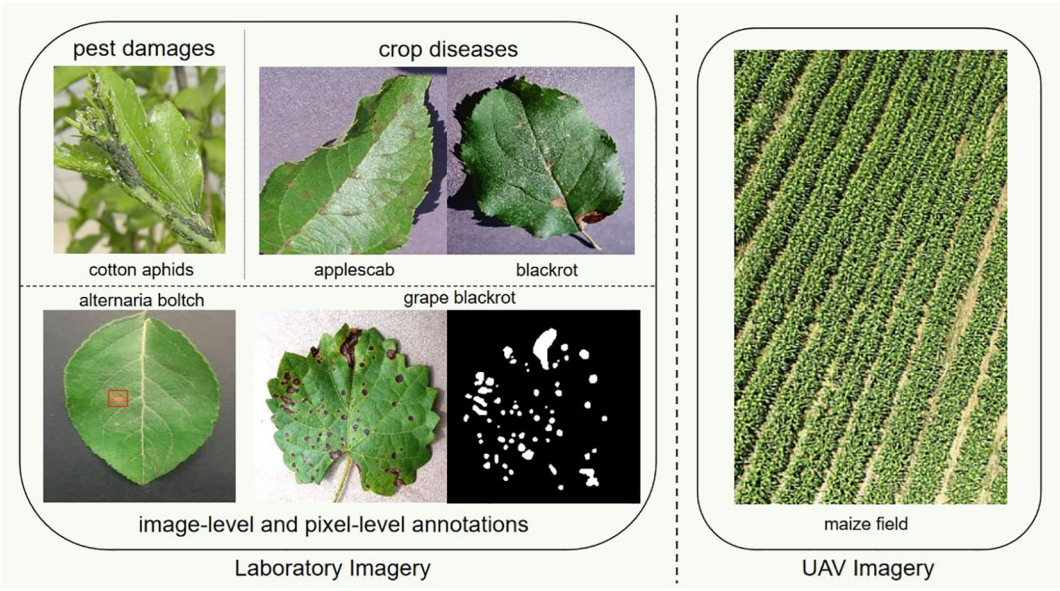


FIGURE 4
Crop diseases and pests imagery.

of crop diseases and pests issues, even in their early stages of infestation. Moreover, using multispectral and hyperspectral sensors, UAVs can gather data across different wavelengths. This helps in more accurately analyzing crop health and identifying specific types of pests or diseases. In addition, UAVs can provide timely information on pests and diseases, enabling farmers to take preventive measures and making agricultural management more intelligent. Preventing widespread infestations while reducing the use of insecticides. In summary, UAV remote sensing technology provides an efficient, cost-effective, timely, and accurate method for pest and disease monitoring, which is important for modern agricultural management and crop protection.

4 UAV RS systems

The above analysis showed that UAV RS has been widely used in various fields of IA, especially in crop diseases and pests monitoring, and the UAV RS system represented by plant protection UAV has been rapidly developed and applied (He, 2018; Yang K. et al., 2020). A complete set of low-altitude UAV RS systems mainly includes a UAV platform, sensors, a ground station system, and a communication data link. Here, we focus on reviewing the typical types of UAVs, as well as the four typical types of sensors.

4.1 Typical types of UAVs

With the rapid development of UAV low-altitude RS, represented by plant protection, UAVs have excellent performance in crop information collection, pest monitoring, spraying, fertilization, and other fields. According to the structure of UAVs, plant protection UAVs are divided into coaxial, single-rotor, and multi-rotor helicopters (Lan et al., 2019). The coaxial plant protection helicopters are generally a hydraulic motor type with solid endurance, sizeable operating area, and other characteristics. But limited by the engine maintenance complexity, as well as the engine life being short and other issues, they are rarely used in practical agricultural operations. Single-rotor plant protection UAVs are mainly the oil-driven type. Compared with coaxial plant protection helicopters, oil-powered single-rotor plant protection drones can effectively reduce maintenance frequency and extend engine life; they have a wide range of crop applications and long working hours. However, because of their high cost and hard to control, they still have certain limitations in large-scale agricultural applications.

Given the limitations of coaxial and single-rotor plant protection drones, electric multi-rotor plant protection drones are widely used in crop disease monitoring. Multi-rotor plant protection UAVs can be divided into four rotors, six rotors, eight rotors, and other unmanned aerial vehicles. The fuselage body is usually made of carbon fiber to reduce its load. With the characteristics of large load, convenient maintenance, sufficient power, etc., they have become the mainstream model of plant protection UAVs (Chen H. et al., 2021; Chen et al., 2017). At present, RS monitoring based on multi-rotor UAVs equipped with visual sensors has been widely used in crop diseases and pests identification (Yinka-Banjo and Ajayi, 2019), growth monitoring, yield estimation (Yang et al., 2015), crop lodging judgment, and other aspects, and RS monitoring provides a new means for crop growth monitoring.

4.2 Typical types of sensors

Given that a variety of sensors are available for UAV systems, it is valuable to provide an overview of the cameras and sensors applicable to UAV systems and their characteristics. Colomina and Molina (2014) within their paper reviewed the specific parameters and applications of several RGB cameras, multi-spectral cameras, hyperspectral cameras, thermal cameras, and laser scanners adapted to UAVs, as can be seen in the literature. In pest and disease detection, especially the first four cameras are more applied in UAV RS, and these images are shown in Figure 5. Therefore, suitable solutions and deployment of UAV platforms can be found according to the research purpose and available budget.

4.2.1 RGB cameras

RGB cameras measure specific intensities of the three-color channels of red (R), green (G), and blue (B) in the visible spectrum (400-780 nm), and the color of each pixel is expressed by the superposition of specific values of the three-color channels. RGB cameras have the advantages of being low-cost, easy to carry, and easy to operate, and has been widely used in plant phenotype research (Su, 2020). RGB images are usually processed and analyzed by ML or DL and have excellent application space in crop recognition, plant defect, and greenness monitoring due to their excellent performance (Dalsass et al., 2016; Darwin et al., 2021; van Iersel et al., 2018). Nevertheless, due to the limitation that the RGB cameras only contain three color channels in the visible light range and the poor spectral resolution, it is usually necessary to cooperate with other cameras to complete the identification and detection task.

4.2.2 Multi-spectral cameras

Multi-spectral cameras are mainly used for the visible/near-infrared (VIS/NIR) region, which can obtain the advantages of both spatial information and spectral information of the detected targets. Rapid acquisition and analysis of crop growth information are achieved by analyzing changes in absorption, transmission, or reflection spectra in the visible (400-780 nm) and near-infrared (780-2500 nm) regions. The visible spectrum primarily conveys color-related information, whereas the near-infrared spectrum is derived from the molecular group's vibrational absorption. When crops are afflicted by pests and diseases, a range of physiological and biochemical reactions take place, altering the composition and distribution of pigments, water, and other tissues in affected crops compared to their healthy counterparts. This leads to changes in the density and vibrational intensity of molecular bonds such as C-H, O-H, N-H, etc., which in turn cause variations in optical properties. It provides a theoretical basis for the wide application of multi-spectral cameras in crop yield, pest monitoring, and other fields (Ali

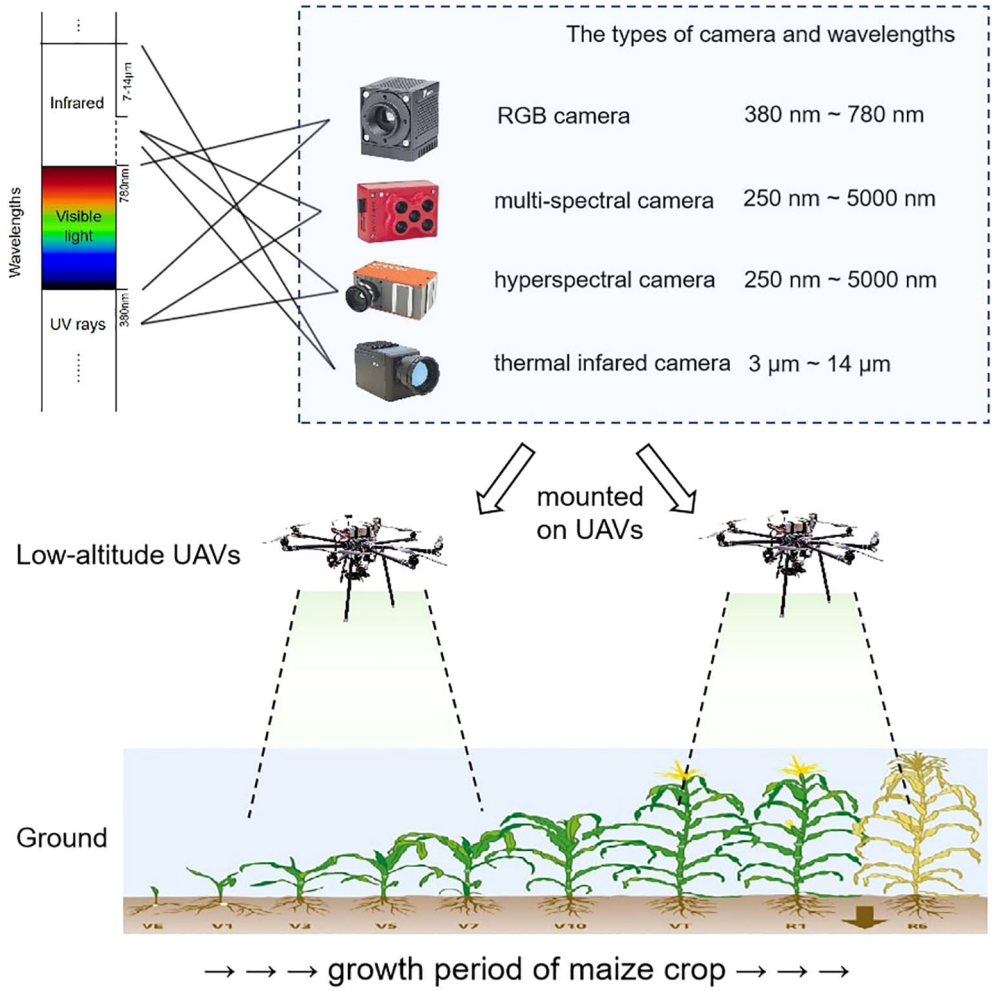


FIGURE 5
The UAV RS system mounted types of cameras.

et al., 2022). With the continuous development of cameras, the development of multi-spectral camera systems has derived many different types: Fourier transform base spectroscopy, wide and narrow band filters, and the like.

4.2.3 Hyperspectral cameras

By capturing the spectral information of a target object at different wavelengths, the hyperspectral camera generates a data cube that contains both spatial and spectral information, allowing the user to analyze the target object from different angles and depths. Compared with the multi-spectral cameras, the imaging systems of the hyperspectral camera can collect more spectral bands, typically up to more than 100 spectral bands, and the hyperspectral cameras usually use one or more continuous wavelength ranges. This provides an incredibly detailed spectral signature for each pixel, enabling the identification of specific materials and chemicals. The use of hyperspectral RS can monitor the growth of crops in the field. By effectively integrating the obtained data, it can provide relevant information promptly, and applying this technology in predicting farmland disasters has a significant effect. Digital processing of images enables segmentation and classification of leaves in the corn fields of the Mexican fields making use of HIS color models (Carranza-Flores et al., 2020; Bravo-Reyna et al., 2020). Applying this mechanism to agricultural management can improve agricultural management and reduce the cost of agricultural management (Lu et al., 2020).

4.2.4 Thermal infrared cameras

Infrared thermal imaging cameras detect infrared radiation emitted by objects, allowing for the measurement of their surface temperatures. They are sensitive to temperature differences and can detect subtle changes that are invisible to the naked eye. Infrared thermal imaging cameras are valuable tools for monitoring water stress in crops. Thermal infrared cameras operate by using the sensors they carry to capture infrared radiation in the range of 0.75 to 1000 μm emitted by the target object and feedback on the temperature of the target object as a digital thermal radiation image (Costa et al., 2013). When pests or pathogens infect crops, physiological states, for instance, transpiration, photosynthesis will change. Thermal infrared cameras can effectively monitor various

characteristics, such as crop growth status and water stress, greatly contributing to IA (Calderón et al., 2013; Gago et al., 2015).

4.3 The technology road-map of RS

From the acquisition of UAV RS images to the identification and location of crop diseases and pests, the whole process involves a series of complex processing processes. The main technology road-map of crop diseases and pests monitoring can be seen in Figure 6. The whole process includes image acquisition, image preprocessing, spectral feature extraction (including vegetation index, effective wavelengths), image feature extraction (texture characteristics, temperature gradient characteristics), modelling, prediction and evaluation using machine learning methods or deep learning methods, and finally using the predicted values to achieve monitoring and treatment of crop pests and diseases. Image preprocessing is necessary to improve image quality, remove noise interference in bad weather, simplify data and increase the accuracy and efficiency of subsequent image analysis and processing tasks. Because of the wide range of applications of spectral image sensors, this paper focuses mainly on spectral images. Traditional machine learning methods need to extract features before building a model and finally get the predicted values, while deep learning methods can build models directly from image inputs. Through qualitative analysis, it is possible to determine whether the crop is

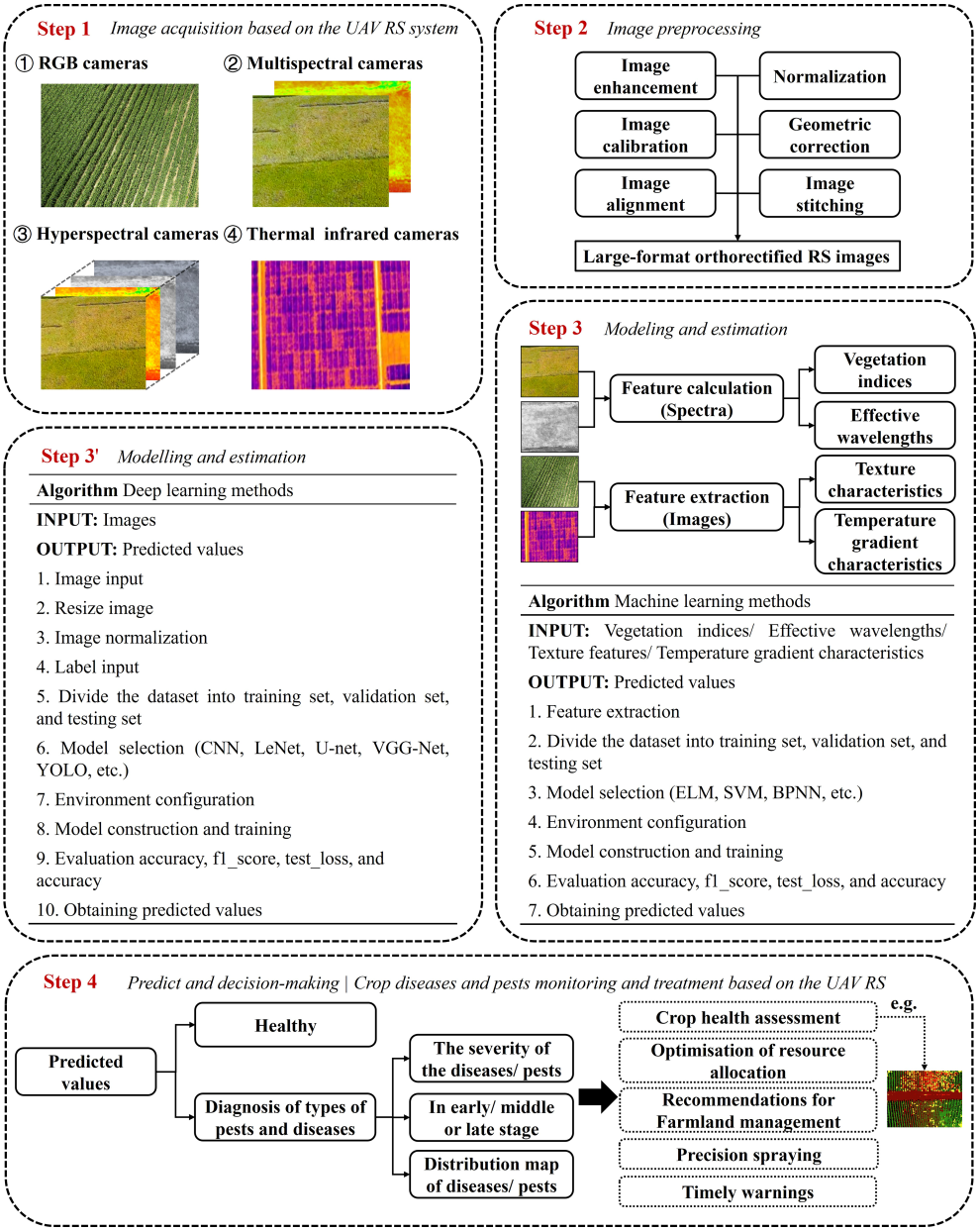


FIGURE 6
The main technology roadmap of crop diseases and pests monitoring.

infected with a disease and what type of disease it is. Through quantitative analysis, it is possible to analyze whether the crop is in the early, middle or late stages of the diseases/pests infestation, the severity of the diseases/pests infestation, and most importantly, construction a map of the regional distribution of the diseases/pests infestation (Kerkech et al., 2020). In this way, we can realize crop diseases and pests monitoring and treatment based on the UAV RS. During the flight of the UAV, it will be affected by many factors, such as light, wind speed, and the state of the crop canopy. The collected RS data often have noise and interference information, which brings adverse effects to constructing the pest detection model (Alvarez-Vanhard et al., 2021).

Although many UAV RS systems are equipped with corresponding image processing software for noise processing, many noises are still in the processed RS data (Bunting, 2017). To eliminate the noises and interference information in RS data, improve the stability and signal-to-noise ratio of data, and obtain more helpful information for the detection tasks, savitizky-golay convolution smoothing (SG), variable normalization (SNV), multivariate scatter correction (MSC), wavelet transform (WT), and other predictive processing methods are often used to process the collected RS data (Yang et al., 2022). Nighttime images captured in hazy conditions often suffer from glow effects, poor lighting, reduced visibility, and considerable noise, leading to a significant decline in image quality. These compromised images can negatively impact subsequent computer vision processes. As a result, effective dehazing of nighttime images is crucial to enhance the clarity of such images and facilitate outdoor computer vision applications. To address this, a novel dehazing algorithm called the 'multi-purpose oriented single nighttime image haze removal based on unified variational retinex model' was introduced (Liu et al., 2023). This advanced unified variational retinex model processes a pre-treated nighttime hazy image by separating it into three components: reflectance, illumination, and noise. The method then individually enhances the reflectance and illumination components through dehazing and gradient domain enhancement techniques. Some crop diseases and pests detection categorize and methods are listed in Supplementary Materials Table A1.

5 Traditional machine learning in intelligent agriculture

Currently, the primary data analysis methods of UAV RS monitoring crop diseases and pests stress focus on the modeling analysis of traditional ML algorithms. The color and texture features of the diseased plants are obtained by using various sensors on the drone (gray level histogram, gray level co-occurrence matrix, wavelet transform, etc.). Temperature and humidity parameters, effective wavelengths, vegetation indices formed by fusing multiple spectral characteristics, and other parameters are used as model input variables. A qualitative or quantitative pest detection model is established by combining ML algorithms such as a support vector machine, a clustering algorithm, a random forest algorithm, a Bayesian algorithm, a least square method, etc. Ultimately, it can realize early discovery, species identification and classification of agricultural pests, and the grading of pest stress degree. It can provide efficient decision information for early prevention and control of crop diseases and pests.

Before using machine learning algorithms for simulation modeling, a feature extraction step is required. For different images, there are different processing methods. Spectral images are processed by feature calculation to get information such as vegetation indices and effective wavelengths. The processing method for RGB images as well as thermal imaging images is image feature extraction to obtain texture characteristics information along with temperature gradient characteristics information. Unlike deep learning methods, feature extraction is an indispensable step for traditional machine learning. Deep learning can directly input the image, but traditional machines need to input the information extracted from the features.

5.1 Features extraction methods

Feature extraction is one of the key steps in UAV remote sensing monitoring of pests and diseases, which involves extracting useful information related to pests and diseases from remotely sensed images. In general, feature extraction can include effective wavelength selection, vegetation index, and texture features. Effective wavelength selection algorithms refine fundamental indices from spectral data and exclude irrelevant spectral data collected by UAV RS systems, thereby improving model performance and simplifying calculations. The vegetation index is a parameter that can reflect the growth condition of vegetation, calculated by RS technology, especially multispectral and hyperspectral data. It can effectively synthesize the relevant spectral signals, enhance the vegetation information and reduce the interference of non-vegetation information. Changes in the vegetation index can reflect the health and vigor of crops, and is an important indicator for assessing the extent of the impact of pests and diseases. Texture features reflect the surface texture information of objects in UAV remote sensing images, and there are differences in leaf morphology, color and other aspects between damaged and healthy plants, which are manifested as different texture features on remote sensing images.

5.1.1 Effective wavelengths selection

UAV RS systems with multi-spectral or hyperspectral cameras often gather extraneous spectral data, which can undermine the precision and consistency of pest detection models (Chen et al., 2002). Researchers have developed wavelength selection algorithms to distill essential indices from spectral data, enhancing model performance and simplifying computations (Thenkabail et al., 2000). These algorithms, chosen based on crop and pest spectral profiles, are tailored to specific detection contexts. Employing a blend of these algorithms or integrating additional ones could bolster model robustness, decrease errors, and boost prediction accuracy. This strategic application of wavelength algorithms

advances detection capabilities, supporting targeted agricultural practices and crop safeguarding. Popular algorithms such as successive projections algorithm (SPA), genetic algorithm partial least squares (GAPLS), uninformative variable elimination algorithm (UVE), and competitive adaptive reweighted sampling (CARS) are in use (Li et al., 2016), with their effectiveness in detecting crop ailments and pests catalogued in research, exemplified in Table 1. for staples like wheat and soybeans. The t-test is a statistical method used to compare whether two sets of means are significantly different, while random forest (RF) is an integrated learning algorithm that makes predictions by constructing multiple decision trees. RF can be applied to select important wavelengths in spectroscopy by training the model on spectral data, extracting feature importance scores for each wavelength, and then choosing wavelengths with higher importance scores as they contribute more significantly to the model's predictive performance.

5.1.2 Vegetation indices extraction

In the research and practice of UAV RS detection, crop spectrum can reflect the growth status of crops. As a result, crop phenotype information can be obtained based on spectral reflectance, but sometimes the direct use of spectral reflectance cannot reflect the growth of plant canopy well. To solve this problem, many researchers try to optimize the crop-sensitive spectral reflectance to reconstruct the monitoring indices, enhance some characteristics and details of the vegetation, and highlight the difference between the detection targets. Twenty-five common in used extraction methods of vegetation indices from the RGB and multispectral camera, including 14 commonly used color vegetation indices and 11 multispectral narrow band vegetation indices are shown in Appendix Supplementary Table A2. The calculated vegetation indices were averaged over the sampled area. The color vegetation indices are visible band vegetation indices that highlight a particular color, such as the green color of plants, which is more intuitive to humans. Vegetation indices extracted from RGB cameras are sensitive to the greenness of plants, and they have been used to extract green vegetation and to calculate vegetation cover. Some improved vegetation indices such as ExG, Woebbecke's index and some combined vegetation indices have also been used to explore the feasibility of predicting yields. First-order derivative spectroscopy is a technique used to remove background signal or noise and for resolving overlapping spectral feature establishment. It is also effective for enhancing the relationship between spectral data and target parameters. In practical scenarios, high-resolution images captured via UAVs can undergo processing by a color feature extraction algorithm to derive diverse color feature values. Subsequently, machine learning techniques are leveraged to develop models for detecting crop diseases and pests.

In addition, 68 commonly used vegetation indices were selected in Supplementary Materials Table A3 regarding available literature. These vegetation indices can be broadly classified into several categories: characteristic, soil line, and atmospheric adaptive indices. Characteristic indices are mostly two or more band operations in the visible and near-infrared bands, such as simple two-band chlorophyll indices, normalized indices and so on. Soil line and atmospheric adaptation vegetation indices are designed to reduce the effects of environmental factors such as soil and atmosphere.

5.1.3 Texture features extraction

As one of the important characteristic variables to characterize the growth of crops, texture features represent the spatial organization of pixel intensity changes in images and contain much essential information related to the occurrence of crop diseases and pests (Mardanisamani et al., 2019). Supplementary Table A4 in Supplementary materials show 9 common texture characteristic parameter calculation methods. In practical applications, high-resolution images can be obtained by the UAVs and processed by a texture feature extraction algorithm to obtain values of different texture features. Then ML is applied to establish crop diseases and pests detection models. At present, the most commonly used texture features extraction method is to extract four texture features from RS images, namely, contrast CON (Contrast), correlation COR (Correlation), entropy EN (Entropy), and homogeneity HO

TABLE 1 Application of the wavelength selection algorithms.

Sensors	Crop diseases/pests	Wavelength selection algorithms	Selected EWs (nm)	References
Hyperspectral cameras	Wheat powdery mildew	CARS	450, 560, 650, 730, 860	(Song et al., 2022)
Hyperspectral cameras	Aphid density of winter wheat	T-test	491, 617, 750, 1690	(Luo et al., 2013)
Hyperspectral cameras	Rice canopy infested with brown spot disease	GAPLS	822, 738, 793, 402, 570, 753	(Zhao et al., 2012)
Hyperspectral cameras	Strawberry anthracnose and gray mold	Random forest (RF)	822.93, 775.55, 783.26, 807.49, 890.28, 829.55,	(Jiang et al., 2021)
Hyperspectral cameras	The blight diseases of tomato leaves	SPA	442, 508, 573, 696, 715	(Xie et al., 2015)
Hyperspectral cameras	The gray mold disease of tomato leaves	T-test	655, 746, 756	(Xie C. et al., 2017)

(Homogeneity), according to gray level co-occurrence matrix (GLCM) (Wang et al., 2017). Combining texture features with their type feature variables can effectively improve the detection performance of quantitative detection models (Pang et al., 2021). In addition, the DL model also shows excellent performance in texture feature extraction. These depth models can automatically learn complex texture features, so they are widely used in the field of crop pest detection. When applying the depth model, a large amount of mark sample data is needed to train the model, which also becomes a big challenge in texture feature extraction by using the depth model.

5.2 Typical traditional machine learning frameworks

An overview of a typical ML approach system is illustrated in Figure 7. ML tasks start with input training data. Depending on the learning signal of the learning system, training data are typically sorted into two types: labeled and unlabeled. Hence, ML approaches are mainly classified into three categories: supervised learning, semi-supervised learning and unsupervised learning. Supervised learning involves the algorithm being trained on a data set complete with input variables and their corresponding output results, aiming to deduce a rule that consistently links the two. Sometimes, the inputs might be incomplete, or the desired outputs may only be provided as feedback within a changing environment, a scenario known as reinforcement learning. In such supervised scenarios, the model that has been honed is then applied to estimate the unknown outputs or 'labels' for new data. Conversely, unsupervised learning deals with data that lacks predefined labels, merging all data without a clear division into training or test subsets. Here, the algorithm's goal is to sift through inputs and unearth underlying structures or patterns.

5.2.1 Extreme learning machine

The ELM is highly effective in generalizing feed-forward neural networks. It utilizes a single hidden layer neural network that can randomly assign input weights and biases in the hidden layer, thereby addressing common issues such as over-fitting and local minima (Ding et al., 2014). The ELM stands out for its rapid learning pace, robust generalization capability, and ability to find unique, optimal solutions. As a result, it's extensively employed in a variety of classification and regression tasks. In the research of Kouadio et al. (2018), ELM model has the ability to analyze soil fertility properties and to generate an accurate estimation of robusta coffee yield. The performance of 18 different ELM-based models with single and multiple combinations of the predictor variables based on the soil organic matter, available potassium, boron, sulfur, zinc, phosphorus, nitrogen, exchangeable calcium, magnesium, and pH, was evaluated. Khan et al. (2022) applied a multi-level deep entropy-ELM feature selection technique to recognize cucumber leaf diseases, and the best accuracy obtained by training on five different datasets was 98.4%.

5.2.2 Support vector machine

The SVM identifies a hyperplane that maximizes the margin between the closest points of the training dataset. This process is executed through quadratic programming optimization using a radial basis kernel function, making it ideal for tackling small-sample, nonlinear, and high-dimensional datasets in both classification and regression contexts. SVM operates on the principle of structural risk minimization, constructing an optimal hyperplane for ideal classification by balancing training set error and complexity (Mahesh, 2020). SVM employs four primary kernel functions: the linear kernel, polynomial kernel, sigmoid function, and radial basis function kernel. These functions facilitate the transformation of low-dimensional space vectors into a high-dimensional space, aiding in the analysis of sample separability in this expanded space. Both regression and classification tasks in SVM share a common foundation. The choice of kernel function plays a crucial role in influencing the effectiveness of SVM in classification or regression. For classification tasks, SVM outputs a class value, whereas in regression scenarios, the output can be any real number. Kale and Shitole (2021) analyzed a SVM based method

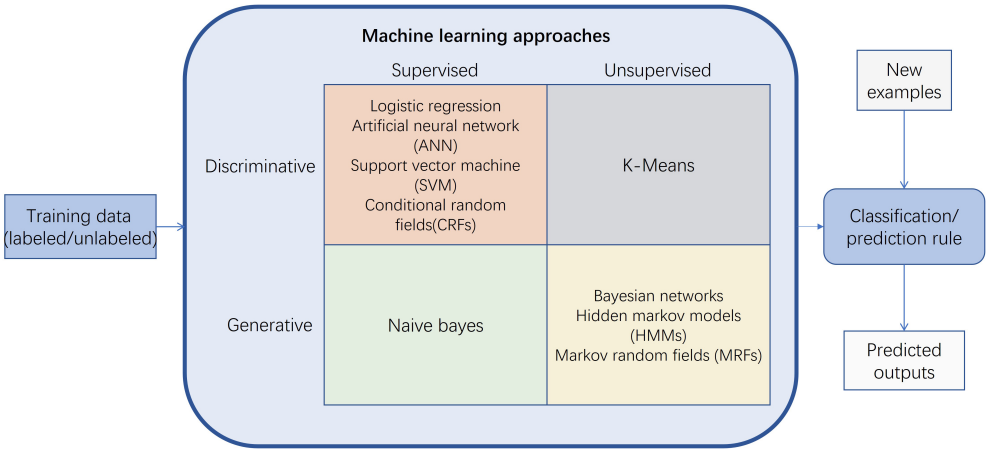


FIGURE 7
A typical machine learning approach system.

for crop disease detection. SVM algorithm is used to classify the extracted features. The data points are classified by finding a maximum spaced hyperplane in N dimensional space and the position of the hyperplane is optimized by means of support vectors in order to minimize the classification error.

5.2.3 Back propagation neural network

The BPNN primarily utilizes an error back-propagation algorithm. This method involves adjusting the network's connection weights after each training cycle based on the error back-propagation, continuing until the discrepancy between the actual and predicted output values is minimized (Hecht-Nielsen, 1992). For a given input sample set in BPNN, it's necessary to define the number of hidden layers and nodes, select a learning algorithm and rate, choose a transfer function, and establish conditions for training termination. The neuron transfer functions in BPNN typically encompass a threshold function, a linear function, and a sigmoid function. The presence of irrelevant information in the data can impact the accuracy of the BPNN algorithm, and the inclusion of large sample data sets can slow down the BPNN's modeling process. Consequently, it's more efficient to build BPNN models using characteristic variables derived from raw data as inputs, which also enhances the computation speed.

6 Deep learning algorithms

In the modeling process of traditional ML algorithms, feature extraction mainly depends on a manually designed feature extractor, which requires solid professional knowledge and the ability to model parameter adjustment. Meanwhile, each algorithm has strong pertinence but poor generalization ability and robustness (Nazarenko et al., 2019). In contrast, deep learning relies on deep neural network (DNN) advantages, for example highly optimized algorithms and multiple unit layer architecture. DNN can extract features based on data-driven, automatically extract deep and dataset-specific feature representation according to the learning of a large number of samples, which is more efficient and accurate for the expression of the dataset. The extracted abstract features are more robust and have better generalization ability, which can realize end-to-end expression (Shrestha and Mahmood, 2019). With the explosive growth of UAV remote sensing image data, the importance of deep learning becomes more important than ML. Some state-of-the-art (SOTA) methods are the followings: convolutional neural network (CNN), VGG-Net (Simonyan and Zisserman, 2014), ResNet (He et al., 2016), ResNeXt (Xie S. et al., 2017), HRNet (Wang et al., 2020), RegNet (Radosavovic et al., 2020), LeNet, U-Net, etc. Therefore, DL algorithms have shown promising results and great potential in the early detection of crop diseases and pests, identification of pest species, and classification of disease severity (Kamilaris and Prenafeta-Boldú, 2018). Many researchers have conducted applied research on DL algorithms and achieved good results, as shown in the appendix Supplementary Table A5. The following describes a variety of DL frameworks commonly used in crop diseases and pests detection.

6.1 Typical deep learning frameworks

With the development of computer technology, advanced learning is represented by CNN (Abdel-Hamid et al., 2014). It extracts features layer by layer through convolution and pooling and has the characteristics of weight sharing and local connection at the same time, which reduces the number of training parameters and makes the model easier to optimize. After the proposal of CNNs, various new deep network structures have been proposed. In this article, the main focus is on three commonly used deep learning network structures found during previous hot-spot analyses.

6.1.1 LeNet network

LeNet is a DL framework proposed by LeCun et al. (1998). LeNet has seven layers of the network, including two convolutional layers, two pooling layers, and three fully connected layers. It uses softmax to process the output data. In addition to the input layer, each layer contains trainable parameters and has multiple feature maps. LeNet is the first successful application of CNN to solve the image classification problem, which is widely used in image recognition and text character recognition tasks. Though the network structure of LeNet is relatively simple, it has great adjustability in changing the size of the full connection layer and the size of the convolutional layer and can adapt to different image classification tasks. The feature map can extract each input feature through a convolutional filter, and its topology is illustrated in Figure 8A. In addition, modern convolutional neural networks are often improved and developed based on the LeNet framework. For example, classical convolutional neural networks such as AlexNet, VGG-Net, GoogLeNet, and ResNet have all proposed more effective network structures and training strategies on the basis of LeNet's work and achieved amazing results in image classification and object recognition. Gayathri et al. (2020) used LeNet to classify and detect various diseases of tea, among which the accuracy rate of tea red spot disease was as high as 94%. Wallelign et al. (2018) obtained 12673 soybean images taken in the natural environment and used the LeNet framework to classify and detect many diseases, such as soybean septicemia and leaf blight, among which the highest accuracy rate reached 99.32%.

6.1.2 U-net network

U-net was proposed by Ronneberger et al. (2015), and its original intention was to solve the problem of medical image segmentation. Later, due to its excellent performance, the U-net architecture became more popular in various fields. The U-net is composed of two main parts: the feature extraction layer and the up-sampling layer. Its name comes from its U-shaped architecture, and its topology is illustrated in Figure 8B. The feature extraction part consists of two 3x3 convolutional layers (ReLU activation function) plus a 2x2 max-pooling layer (maximum pooling layer) to form a down-sampled module right half, which is repeatedly composed of an up-sampled convolutional layer (deconvolution layer) plus feature concatenation concat plus two 3x3 convolutional layers (ReLU). To further improve the performance of U-net, many researchers have improved and extended the U-net network

structure to adapt to more complex scenes. For example, the introduction of deeper network layers, attention mechanisms, multi-scale structure, cross-scale connection, and other technologies can improve the performance and generalization ability of the U-net network to some extent. Because of the excellent performance and flexibility of the U-net algorithm itself, as well as its abundant improvement and extension methods, its application prospect and research value in the field of computer vision are paid more and more attention. Bhujel et al. (2022) used the U-net model to perform image segmentation and quantitative analysis on strawberry gray mold, and the U-net model showed robust performance on a wide range of test images. Chen S. et al. (2021) employed the U-net architecture to subdivide rice bacterial leaf spot disease and evaluate the severity of leaf spot disease. Rao (2021) utilized the improved separable convolution of the U-net model to classify cassava mosaic disease and cassava bacterial wilt, and the accuracy reached 83.9%.

6.1.3 VGG-Net network

VGG-Net is a model proposed by the visual geometry group of Oxford University. It is characterized by small convolution and deep network layers, which proves that increasing network depth can effectively improve performance. According to the size of the convolution kernel and the number of convolution layers, there are six configurations of VGG, namely A, A-LRN, B, C, D, and E, of which D and E are the most commonly used, namely VGG16 and VGG19. VGG16 has sixteen layers, including thirteen convolutional layers and three fully connected layers. The first time, after two convolutions with 64 convolution kernels, one pooling is used. The second time, after two convolutions with 128 convolution kernels,

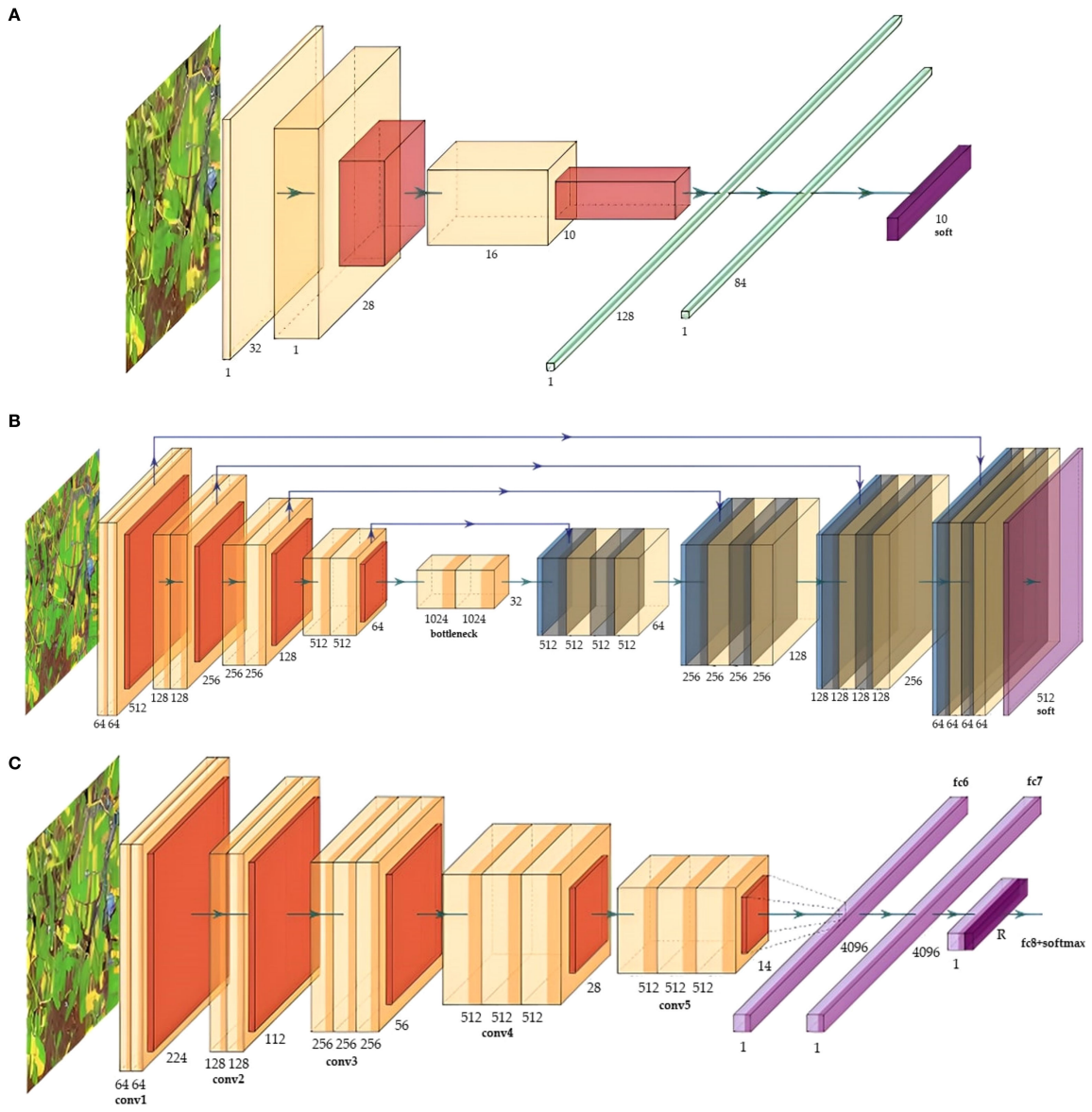


FIGURE 8
The deep learning models. **(A)** LeNet, **(B)** U-net, **(C)** VGG-Net.

another pooling is used. Three 512 convolution kernels are repeated twice, and then pooling is repeated. Finally, three fully connected layers are used. The topology is shown in Figure 8C. Li et al. (2020) used the VGG16 and InceptionV3 to identify different degrees of ginkgo leaf disease. The accuracy of VGG16 in laboratory datasets was 98.5%, and the accuracy rate in the field dataset was 92.19%. Rangarajan et al. (2018) used the pre-trained DL architecture VGG16 net to classify tomato crop diseases. The classification accuracy of the VGG16 network using 13,262 images was 97.29%. Yan et al. (2020) employed an enhanced version of the VGG16 model to classify and identify apple leaf diseases. In this modified model, they replaced the fully connected layer with a global average polarization layer to decrease the number of parameters. Additionally, a batch normalization layer was incorporated to accelerate the convergence speed. The experiments conducted with this model demonstrated an overall accuracy of 99.01% in apple leaf disease classification.

6.2 Latest technology in processing crop diseases and pests imagery

Large language models (LLM) have been widely used in various fields with promising results. However, the IA has yet to fully integrate LLM into its practice due to the dominance of visual images. For this reason, it is worth exploring how LLM can be applied to IA on a large scale. It's worth mentioning that the target detection network——the YOLO (You only look once) network was proposed as a fast-type method since it adopts a single neural network to run all components of the given task. YOLO is widely used for its small size and fast processing speed, and has been upgraded iteratively, the latest version is YOLOv10 (Wang et al., 2024). Qing et al. (2023) made a nice try to simultaneous application of YOLO and large models for pests and disease detection. A novel approach that combines YOLOPC and Gpt-4 was proposed. YOLOPC is a YOLO lightweight variant. Using the ability of YOLOPC, turning the affected leaves images into a natural language description of the disease, which is called image-to-text. After that, these descriptions of the crop diseases are then entered into a chatbot, which uses the reasoning capabilities of GPT-4 as well as the language generation capabilities to generate a disease diagnosis report. The experiments were tested and evaluated using datasets from different sources. Test results show that on the premise of entering text assistant, this network model's induction and reasoning module demonstrates 90% reasoning accuracy in generating agricultural diagnostic reports. A very interesting and novel attempt to use existing models in combination with LLM. In the future, there will be more powerful LLMs such as GPT-5 or GPT-6, and this combination of ideas will still be very useful in the future.

The development of large vision models (LVM) is a direction that AI researchers want to follow to make further progress as they continue to refine LLM. Compared to LLM, visual information is typically 2-dimensional images, 3-dimensional stereo images, or 4-dimensional stereo video information. Therefore, the increase of 1-3 dimensions compared to the processing of linguistic information results in a higher difficulty level. However, the emergence of GPTs that are simultaneously LLM and LVM has changed our perception. It is capable of both processing image data and expressing the results obtained by processing the data in natural language to form an agricultural inspection report. As shown in Figure 9, we tried out the results of GPT-4's reasoning with images of a diseased leaf. The image used was the apple leaves scab fungus disease image. We tested GPT-4 on images of one leaf that was severely diseased, and another leaf that was slightly diseased. The two paragraphs in the picture are ChatGPT's responses to the two pictures I uploaded of diseased leaves. ChatGPT gives its judgement and advice. Notably, the image of the slightly diseased leaf posed challenges for recognition. The outcomes revealed the capability of GPT-4 to successfully detect visibly diseased leaves. However, it fell short in effectively identifying subtle signs of disease in the slightly diseased leaves, providing less satisfactory results in such cases.

6.3 Deep learning in UAV RS - some discussions in detail areas

It is undeniable that DL-based methods are powerful and efficient tools for processing the large amounts of data generated by RS systems. In this section, there are some brief discussions of some detail areas.

6.3.1 Image resolution problem

The input of DL architecture is low-resolution images, and the processing method for high-resolution images is to scale them to the size required for processing. Hence, it is a challenge that maintain the high resolution of images in the process. In this regard, the latest HRNet (Kannojia and Jaiswal, 2018) attempts to maintain high-resolution CNN architecture in deeper layers.

6.3.2 Real-time processing

In the era of IA, crop protection applications can benefit from DL. However, these DL algorithms are highly computer-intensive. Usually, a data center or graphics processing unit (GPU) with strong computing power is required for post-processing. Therefore, the bottleneck of the development of DL in the field of IA is still real-time processing. There are two ways to solve this problem: developing better and faster algorithms and enhancing better GPUs. In terms of algorithms, MobileNets (Howard et al., 2017) with layers with fewer parameters can still maintain prediction performance. In terms of hardware, some platforms such as NVIDIA's Jetson has been developed to better run DL (Mittal, 2019).

6.3.3 Dimensional reduction

Unlike RGB images, the latest hyperspectral images captured by drones are composed of dozens to hundreds of spectral bands, which can help classify and regress the characteristics of leaves. However, high dimensionality brings two problems: First, the high correlated bands. Second, the increasing learning cost. Therefore, hyperspectral data may hinder the accuracy of DL-based methods,

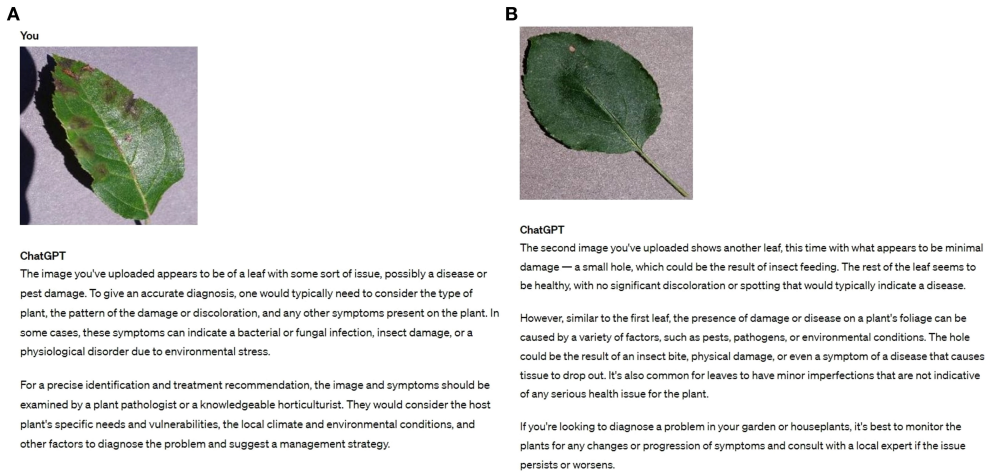


FIGURE 9
The GPT-4 results after entering an image of diseased leaves. **(A)** Severe diseases **(B)** Slight diseases.

which is an important issue to consider in RS practice. At present, the classic method to solve high dimensions is to apply PCA, but today when the amount of data has increased significantly, this method may need to be updated. There, Licciardi et al. (2011) suggested a singular, integrated method within a network engineering framework. This method focuses on identifying the most pertinent frequency band combinations from hyperspectral sensors, directly correlating to the labeled data presented in the input layer at the network's initial phase.

6.3.4 Transfer learning

Transfer learning is the ability to learn to draw inferences from one example and learn new knowledge by using existing knowledge. Its core is to find the similarity between existing knowledge and new knowledge. How to reasonably find the similarities between them and then use this bridge to help learn new knowledge is the core issue of transfer learning. In practical applications of this model, there are frequently spectral shifts observed between the training and testing images. These shifts typically arise from variations in image acquisition, geographical locations, and atmospheric conditions (Tuia et al., 2016). Therefore, how to reduce the impact of the difference between actual indicators on the DL algorithm is also a direction worth studying.

7 Practical application of UAV RS

7.1 Application of the UAVs assembled the RGB cameras

The RGB cameras equipped with the UAVs are often used to obtain morphological indicators of crops, for instance, lodging area, leaf color, plant height, canopy coverage, and panicle number. Morphological indices, as important phenotype parameters to characterize crop growth, are essential in monitoring crop diseases. The RGB images are processed by RS image classification, image feature extraction, and hybrid model recognition, which can realize the rapid acquisition and accurate diagnosis of crop disease information.

Liu et al. (2018) constructed a wheat disease detection model based on a random coefficient regression model based on RGB RS images of UAV and realized an accurate evaluation of wheat powdery mildew disease severity. Oh et al. (2021) extracted crop canopy coverage, canopy volume, and vegetation indexes based on UAV RGB RS images and combined them with various ML algorithms to construct various detection models of corn tar spot disease, realizing accurate detection of corn tar spot disease. Calou et al. (2020) monitored banana yellow leaf disease with an RGB camera mounted on UAV and constructed the banana yellow leaf disease detection model by using ML algorithms including maximum likelihood, minimum distance, SVM, and artificial neural network. Among the above models, the overall accuracy of the SVM model achieved 99.28%. RGB cameras have the advantages of low prices and easy operation. Nevertheless, it is limited by the limited spectral band in RGB images, which cannot reflect more physiological information about crops. Compared with RGB cameras, UAV multi-spectral imaging systems can obtain more diverse spectral characteristics and perform better in biochemical trait estimation due to the contribution of near-infrared spectral information.

7.2 Application of the UAVs mounted the multi-spectral cameras

The multispectral imaging sensors carried by the UAVs obtain the map information of the crop, which can more comprehensively present the spectral characteristics of the crop and realize the *in-situ*, rapid, and efficient monitoring and accurate acquisition of the crop disease information. High spectral-resolution sensors have significant advantages and wide application potential in quantitative RS of IA.

Marin et al. (2021) extracted various vegetation indices including the green ratio vegetation index (GRVI), green

normalized difference vegetation index (GNDVI), and normalized difference vegetation index (NDVI) based on UAV multispectral images. They constructed various coffee leaf rust detection models using four decision tree models (logistic model tree (LMT), reduced-error pruning tree (REPTree), regression tree (RT), and RF), among which the LMT detection model achieved 91.5% accuracy. Lan et al. (2020) used the UAV RS platform to collect multispectral RS images of large citrus orchards, explored the best vegetation index combination for citrus Huanglongbing detection, and constructed various citrus Huanglongbing detection models in combination with various ML algorithms (SVM, k-nearest neighbors (KNN), naive Bayes, logistic regression, ensemble learning). Xiao et al. (2022) constructed various apple fire blight detection models based on UAV multispectral RS images and various ML algorithms (decision tree, RF, and SVM), among which the overall accuracy of the random forest model reached 94.0%. Zheng et al. (2018) designed and proposed a real-time wheat disease detection algorithm called efficient double-flow UNet (DF-UNet) based on UAV multispectral images and a DL algorithm for the diagnosis of wheat yellow rust severity. Compared with the jagged spectral map provided by the multispectral cameras, the hyperspectral cameras can provide a smooth spectrum and higher spectral resolution, which makes up for the defect that the multispectral cameras cannot depict narrow spectral features. In addition, the hyperspectral cameras' imaging speed is faster, making the data acquisition cycle shorter and more efficient.

7.3 Application of the UAVs mounted the hyperspectral cameras

Hyperspectral imaging mainly uses narrow electromagnetic bands to obtain the spectral information of ground objects. Since it came out, it has been widely used in crop diseases and pests. Zhang et al. (2003) identified and detected tomato leaf miners based on hyperspectral fusion, and the recognition rate and overall recognition rate of the GA-BPNN model for samples at all levels reached 93.33%. Based on the hyperspectral RS of UAV, Feng S. et al. (2021) constructed the classification and detection model of leaf blasts and explored the classification method of rice leaf blasts, among which the classification prediction accuracy reached 98.58%. Ma et al. (2021) conducted hyperspectral RS on a chestnut planting area using a UHD185 hyperspectral camera mounted on a UAV (DJI Dajiang M600). They analyzed the spectral characteristics of leaves that were locally infected, unevenly infected, and recovering from infection. Then they determined the relationship between the damage degree of red spider diseases and pests and the changes in 6 spectral characteristics, namely the green peak, red valley, low position, red edge, high position, and high shoulder. Abdulridha et al. (2020) used unmanned aerial vehicles (Matrice 600 Pro, Hexacopter, DJI Inc.) with a hyperspectral camera to detect powdery mildew at different stages in pumpkins. In field conditions, the RBF neural network was used to classify the early and late stages of disease development as 89% and 96%, respectively.

7.4 Application of the UAVs mounted the thermal infrared cameras

The upper limit of the spectral range that the above three sensors can collect is 5000 nm, while the thermal infrared camera can collect spectral information between 3000 nm and 14000 nm. The thermal infrared cameras can collect crop temperature information, photosynthesis, transpiration, and other crop physiological conditions that cannot be obtained by the above cameras, providing another crop pests monitoring method.

The real-time monitoring of crop temperature using thermal infrared cameras mounted on UAVs is crucial for the detection of early pests and diseases with non-visual symptoms, and thermal images from UAVs have greater temporal and geographic resolution than images received from satellites, making them a valuable source of information for agronomic applications (Ribeiro-Gomes et al., 2017). Feng et al. (2022) used a six-rotor UAV (Matrice600PRO DJI) as an RS platform equipped with thermal capture real-time monitoring of wheat canopy temperature information by thermal cameras to classify and detect wheat powdery mildew. Chen A. et al. (2021) used UAV to obtain thermal and visible-light aerial images. Thermal imaging can effectively detect late blight and white mold in field crops.

7.5 Application of multi-source data fusion

Limited by the different band ranges, interference conditions, imaging mode, and mechanism obtained by each sensor, any single data source has limitations and cannot fully reflect the spatial-temporal characteristics of crops. The multi-source RS carried by UAVs acquires RGB images, full-band spectrum or multi-band spectrum information, three-dimensional structure information, and temperature of crops by carrying various crop information monitoring equipment. It realizes the joint application of multi-source monitoring information through data fusion and forms a complete working system of UAV RS monitoring crop diseases and pests.

Because of the complexity and diversity of the actual farmland layout and the seriousness of spectral mixing, it is imperative to improve the spatial resolution of UAV RS. Multi-source data fusion can effectively improve the ability to describe the details of crop spatial distribution, improve the spatial resolution and clarity of the image, and reduce the impact of mixed pixels to a certain extent. Image fusion is the most common multi-source RS data fusion method. It aims at improving the spatial resolution. The synthetic image with new spatial and spectral characteristics is generated by processing and calculating the multi-source RS data according to specific rules. The hue-luminance-saturation (HLS) transform method based on color correlation and the PCA, wavelet transform methods based on statistical methods can effectively improve the spatial resolution of UAV RS. In addition, grain characteristics, spectral characteristics, and biochemical characteristics produced by crops at different stages of suffering from different diseases and pests are significantly different.

Therefore, long-term RS data collection and multi-temporal data fusion can make up for the deficiency in the timing of crop monitoring and greatly improve the accuracy of classification monitoring of crop diseases and pests.

For wheat powdery mildew, Feng Z. et al. (2021) fused the hyperspectral features collected by the spectral radiometer, thermal infrared image data obtained by a thermal infrared camera, and texture features obtained by an RGB camera, which greatly improved the intensive reading of RS monitoring of wheat powdery mildew. DadrasJavan et al. (2019) constructed a citrus Huanglongbing detection model based on the low-altitude RS platform of RGB and multispectral cameras carried by UAV, combined with the SVM algorithm, and achieved 81.75% overall classification effect. In their study, Lei et al. (2021) utilized RGB and multispectral RS images obtained by UAV to extract normalized difference vegetation indices, optimized soil adjusted vegetation index (OSAVI), leaf chlorophyll index (LCI), GNDVI, and normalized difference red edge (NDRE) index. They constructed different detection models for the injury levels of areca yellow leaf disease depended on five ML algorithms (BPNN, decision tree, naive Bayes, SVM, KNN). The results demonstrated that the classification accuracies of the test sets of the BPNN algorithm and SVM algorithm were satisfactory, at 86.57% and 86.30%, respectively. Francesconi et al. (2021) used an unmanned DJI Matrice 600 multi-rotor aircraft equipped with a Zenmuse X5 RGB camera and a Zenmuse XT thermal infrared camera to collect RGB images and thermal infrared images for data fusion, which improved the accuracy and efficiency of classification and detection of wheat fusarium cephalosporins. Based on the fusion of the hyperspectral image and the multispectral image source, Cock et al. (2016) established a monitoring model for corn borer damage, effectively classified areas with different disease severity levels, and facilitated exemplary management of corn planting areas.

8 Conclusion and prospects

8.1 Deficiencies in the existing research and challenges to be solved

In recent years, with the continuous innovation of UAV RS, outstanding platforms have been built for research in crop monitoring. The UAVs have made significant progress in detecting and classifying crop diseases and pests and greatly promoted the development of IA (Hunt and Daughtry, 2018). However, there are still some deficiencies in the existing research, and there are still many problems and challenges to be solved.

8.1.1 Multi-source data fusion

There are a bunch of sources of monitoring data on crops, such as images or digital data obtained by hand-held spectral camera detectors, image data captured by cameras mounted on unmanned vehicles, and images captured by various types of cameras on UAVs. It's worth exploring how to fusion these types of data because it will help us form a better overall system.

On the other aspect, due to the diversity of crop diseases and pests damages, different diseases and pests may appear simultaneously in similar spectral characteristics in the same crop. The same disease and pests may also appear in different spectral characteristics in the same crop at different times: the so-called 'same spectrum foreign matter' and 'same substance different spectrum' phenomenon. At the same time, a single sensor is limited by its spectral band, resolution, and other factors (Gao et al., 2020). The information on crop diseases and pests obtained by it is relatively limited, so it is difficult to solve the above problems. In contrast, multi-source data fusion can fully use different sensors' characteristics and advantages. It can coordinate multiple sensors to jointly monitor crops' growth, analyze crops' growth from multiple levels, and provide a new solution for solving the errors and interference caused by the phenomenon of the same spectrum of foreign matter and the same thing different spectrum.

8.1.2 Combination of multi-scale RS methods

The RS of UAVs is limited by the load capacity and endurance of the UAV itself, so it is challenging to monitor farmland in real time in large areas and long time. Therefore, how to improve the UAV RS to achieve multi-temporal and multi-space crop diseases and pests monitoring has become a momentous research topic. To solve this problem, multi-scale RS monitoring of crop diseases and pests, which combines space RS, aerial RS, ground RS, and UAV RS, is a more prominent solution.

RS monitoring by using sensors carried on satellites can obtain three characteristics spatial resolution, time resolution, and spectral resolution at the same time. It combines the temporal dynamic change of crop diseases and pests process, the spatial range change of pests and diseases occurrence area with spectral characteristics. It has the technical advantages of all-weather, multi-mode, and multi-polarization (Neupane and Baysal-Gurel, 2021). Additionally, aerial RS and ground RS data have the advantages of high spectral resolution and convenient and diverse acquisition methods, which entirely make up for the shortcomings and defects of UAV RS. By utilizing the plurality of RS technical means, the trinity of ground-space satellite and the point-surface combination is realized to monitor and classify crop diseases and pests together. Such a combination allows for multilevel monitoring of crop leaves and canopies. Thus, the method has extremely high practical value.

8.1.3 Improvement of algorithms

In processing and analyzing the acquired UAV RS data using the traditional ML algorithm, tedious feature extraction operations need to be carried out so that real-time processing of the UAV RS data is difficult to realize. Meanwhile, the period of building a model is prolonged. However, using DL for modeling analysis requires much data to train the model. When the training data is insufficient, or the data sample type is unbalanced, problems such as under-fitting or over-fitting are prone to occur. In addition, the generalization ability of the DL model is weak. It is difficult to directly apply the model for specific crop pests to the pest recognition of other crops. Because of the above problems, the traditional ML algorithm is combined with the DL algorithm, and by utilizing the characteristics of the DL

algorithm to automatically learn and extract the deep feature information of the data, the key feature information about the crop diseases and pests in the RS data is efficiently obtained, and the key feature information is taken as the input variable of the traditional ML algorithm so that a disease and pest detection model with stronger universality and stability is constructed.

8.2 Future of deep learning in IA

DL is still considered a 'black box' solution to most problems (LeCun et al., 2015), although new research is minimizing this concept to a considerable extent. Due to the complexity of layer number issues and parameter issues in deep learning, it is difficult for us to clearly understand why solutions that are considered good from the results are good. Regardless, in the field of RS, it has provided important discoveries in most of its implementations. In this paper, our aim is that this review of the literature will act as a comprehensive overview, encapsulating the various applications of UAVs in the realm of DL networks.

Therefore, from the above analysis, we can draw some conclusions:

1. It is necessary for drones to obtain additional labeled publicly available datasets for training and benchmarking networks.
2. Methods such as R-CNN and XGBoost are slightly insufficient in detection accuracy. However, emerging algorithms such as U-Net and PSNet have achieved better accuracy, which means there is still a lot of room for continuous improvement in algorithms.
3. DL can provide fast inference solutions with the assistance of GPU processing. However, further investigation is still needed regarding the real-time processing of UAVs using embedded systems.
4. In the context of UAV RS, whether some promising topics. For example, combining multi-head attention-based mechanisms and multi-task learning to form a new networks. Whether general large model technology can be applied to pest and disease detection in IA under UAV RS is also worth studying.

8.3 Something even better than DL in the near future

Since the AI application AlphaGo (Silver et al., 2017) defeated the world champ Jie Ke in May 2017, AI has become the talk of the town. Furthermore, the release of ChatGPT (Brown et al., 2020) and GPT-4 (Achiam et al., 2023) exacerbated this process. Maybe AGI could be useful in IA. General AI refers to the capability of machines to mimic human cognition and execute a wide range of tasks through transfer learning and other methods. GPT-4 exemplifies this as a multimodal system, combining language and visual recognition, allowing it to understand and process images. It is feasible to use LLM e.g. GPT4 for image recognition, text organization, and inference to generate crop diseases and pests diagnostic reports with suggested solutions (Lu et al., 2023).

The image processing process of traditional AI models is to analyze, simply reason, and draw conclusions. However, AGI models can recognize and analyze the image directly through 'prompts' and begin to reason, search the knowledge base, draw organized conclusions, and give readable scientific advice. This is

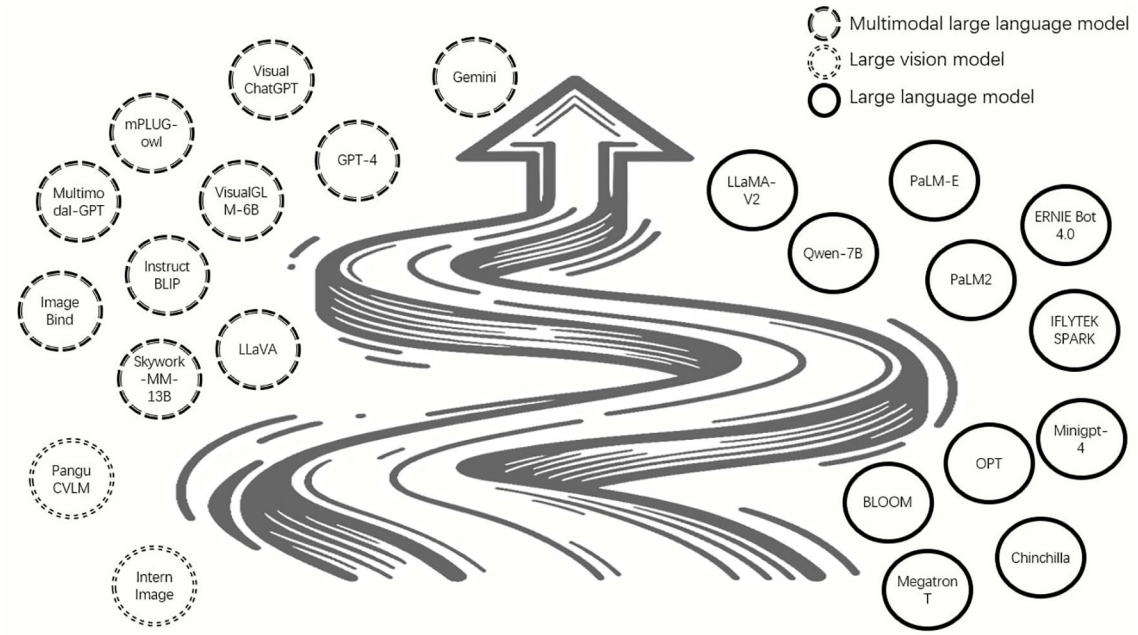


FIGURE 10
LLMs/LVMs/MLLMs that could be used on IA in the future.

the question-answering and dialogue system. Not everyone knows how to use AI models. However, the natural language conversation capability that AGI has allows these individuals to utilize AGI models using a spoken question-and-answer format. Assuming that AGI can be integrated into mobile, it would be a real convenience for a large number of farmers. In addition to this, there are many applications of knowledge graphs to agricultural AGI. For example, a search engine designed for farmers. Agricultural search engines enhance the connection between farmers and experts, providing easier access to necessary information. Additionally, agricultural recommender systems play a pivotal role. These systems aid in distributing reliable information, assisting farmers in making well-informed choices to boost production. This includes recommendations on the best agricultural inputs like seeds, fertilizers, and pesticides to use. By utilizing auxiliary information from the knowledge graph, user preferences can be captured more accurately, enabling more precise recommendations.

AI service integration holds great potential for assistance, documentation, education, interpretation forecasts or data-driven predictions. LLM depict a fundamental step to reduce the gap between AI-driven data analysis and common users. For agriculture, this implies that LLM will improve farmer consultation by providing all necessary information e.g. crop cultivation, breeding, machines, or phytopathology to an advisor or the farmer directly (Kuska et al., 2024).

As shown in Figure 10, we summarized the development process of the big models and selected some models released in 2022 and 2023. The overall trend in large models is from LLM to LVM to multi-modal large language model (MLLM). On one hand, the use of LLM can be combined with other computer vision techniques. On the other hand, the use of LVM could process UAV RS imagery directly. Further on, the MLLM is capable of inputting text, sound, image, and video, and also outputs the multimodal items. The aforementioned LLM, LVM, and MLLM, are all belong to AGI. Considering the generalization of AGI, all these models mentioned in the Figure can be applied to IA in the future. However, for the time being, LLM still has a number of limitations. The outputs generated by LLM sometimes do not provide valid answers due to insufficient information. Often the source of data used to generate the output is not clearly known, affecting the reliability of the output. The performance of multilingual extensions is low, especially in agricultural applications, which are more complex when different languages, dialects and technical terms are involved. Recommendations and conclusions that are valid in one region may not be applicable when translated to another region with different geographical, legal and agricultural conditions.

Author contributions

HZ: Funding acquisition, Methodology, Supervision, Writing – review & editing, Conceptualization, Writing – original draft, Resources. CL: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – original draft, Writing – review & editing. GL: Software, Writing – original draft. DW: Investigation, Writing – review & editing. SQ: Visualization, Writing – review & editing. AL: Software, Writing – review & editing. JX: Writing – review & editing. YH: Funding acquisition, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Our research was supported by the Natural Science Foundation of Guangxi (No. 2024 GXNSFBA010381), the Guangxi Science and Technology Base and Talent Project (Guike AD21075018), the National Natural Science Foundation of China (No. 32071895), and the grant (No. NCOC-24-02) from Key Laboratory of Nonlinear Circuit and Optical Communications (Guangxi Normal University), Education Department of Guangxi Zhuang Autonomous Region.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2024.1435016/full#supplementary-material

References

Abd El-Ghany, N. M., Abd El-Aziz, S. E., and Marei, S. S. (2020). A review: application of remote sensing as a promising strategy for insect pests and diseases management. *Environ. Sci. pollut. Res.* 27, 33503–33515. doi: 10.1007/s11356-020-09517-2

Abdel-Hamid, O., Mohamed, A.-r., Jiang, H., Deng, L., Penn, G., and Yu, D. (2014). Convolutional neural networks for speech recognition. *IEEE/ACM Trans. audio speech Lang. Process.* 22, 1533–1545. doi: 10.1109/TASLP.2014.2339736

Abdulridha, J., Ampatzidis, Y., Roberts, P., and Kakarla, S. C. (2020). Detecting powdery mildew disease in squash at different stages using UAV-based hyperspectral imaging and artificial intelligence. *Biosyst. Eng.* 197, 135–148. doi: 10.1016/j.biosystemseng.2020.07.001

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., et al. (2023). Gpt-4 technical report. *arXiv preprint: arXiv:2303.08774*. doi: 10.48550/arXiv.2303.08774

Ahmad, U., Nasirahmadi, A., Hensel, O., and Marino, S. (2022). Technology and data fusion methods to enhance site-specific crop monitoring. *Agronomy* 12, 555. doi: 10.3390/agronomy12030555

Ali, A. M., Abouelghar, M. A., Belal, A.-A., Saleh, N., Younes, M., Selim, A., et al. (2022). Crop yield prediction using multi sensors remote sensing. *Egyptian J. Remote Sens. Space Sci.* 25, 711–716. doi: 10.1016/j.ejrs.2022.04.006

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

Bhujel, A., Khan, F., Basak, J. K., Jaihuni, M., Sihalath, T., Moon, B.-E., et al. (2022). Detection of gray mold disease and its severity on strawberry using deep learning networks. *J. Plant Dis. Prot.* 129, 579–592. doi: 10.1007/s41348-022-00578-8

Bravo-Reyna, J. L., Montero-Valverde, J. A., Martínez-Arroyo, M., and Hernández-Hernández, J. L. (2020). "Recognition of the damage caused by the cogollero worm to the corn plant, Using artificial vision," in *International Conference on Technologies and Innovation*, Cham. 111–122 (Guayaquil, Ecuador: Springer International Publishing). doi: 10.1007/978-3-030-62015-8

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* 33, 1877–1901.

Bunting, P. (2017). "Pre-processing of remotely sensed imagery," in R. Díaz-Delgado, R. Lucas and C. Hurford eds. *The Roles of Remote Sensing in Nature Conservation*. (Cham: Springer), 39–63. doi: 10.1007/978-3-319-64332-8_3

Calderón, R., Navas-Cortés, J. A., Lucena, C., and Zarco-Tejada, P. J. (2013). High-resolution airborne hyperspectral and thermal imagery for early detection of Verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices. *Remote Sens. Environ.* 139, 231–245. doi: 10.1016/j.rse.2013.07.031

Calou, V. B. C., dos Santos Teixeira, A., Moreira, L. C. J., Lima, C. S., de Oliveira, J. B., and de Oliveira, M. R. R. (2020). The use of UAVs in monitoring yellow sigatoka in banana. *Biosyst. Eng.* 193, 115–125. doi: 10.1016/j.biosystemseng.2020.02.016

Carranza-Flores, J. L., Martínez-Arroyo, M., Montero-Valverde, J. A., and Hernández-Hernández, J. L. (2020). "Search for damage of the citrus miner to the lemon leaf, implementing artificial vision techniques," in *In International Conference on Technologies and Innovation*, Cham. 85–97 (Guayaquil, Ecuador: Springer International Publishing). doi: 10.1007/978-3-030-62015-8_

Carvajal-Yepes, M., Cardwell, K., Nelson, A., Garrett, K. A., Giovani, B., Saunders, D. G., et al. (2019). A global surveillance system for crop diseases. *Science* 364, 1237–1239. doi: 10.1126/science.aaw1572

Chaube, H., and Pundhir, V. (2005). *Crop diseases and their management* (PHI Learning Pvt. Ltd).

Chen, A., Jacob, M., Shoshani, G., Dafny-Yelin, M., Degani, O., and Rabinovitz, O. (2021). "Early detection of soil-borne diseases in field crops via remote sensing," in *Precision agriculture'21* (Wageningen Academic Publishers), 37–44. doi: 10.3920/978-90-8686-916-9_25

Chen, H., Lan, Y., Fritz, B. K., Hoffmann, W. C., and Liu, S. (2021). Review of agricultural spraying technologies for plant protection using unmanned aerial vehicle (UAV). *Int. J. Agric. Biol. Eng.* 14, 38–49. doi: 10.25165/j.ijabe.20211401.5714

Chen, S., Lan, Y., Li, J., Xu, X., Wang, Z., and Peng, B. (2017). Evaluation and test of effective spraying width of aerial spraying on plant protection UAV. *Trans. Chin. Soc. Agric. Eng.* 33, 82–90.

Chen, S., Li, M.-T., Chen, C.-T., Lin, Y.-C., Huang, C.-W., Wu, T.-H., et al. (2002). Remote sensing of crop growth characteristics in greenhouses. *Acta Hortic.* 578. doi: 10.17660/ActaHortic.2002.578.36

Chen, S., Zhang, K., Zhao, Y., Sun, Y., Ban, W., Chen, Y., et al. (2021). An approach for rice bacterial leaf streak disease segmentation and disease severity estimation. *Agriculture* 11, 420. doi: 10.3390/agriculture11050420

Cock, M., Tang Rui, T. R., Liu Zhi, L. Z., Wan HuanHuan, W. H., McGillivray, L., Thomas, S., et al. (2016). The main agricultural insect and disease pests of China and implications for the use of remote sensing for their management. *CABI Rev.* 2016, 1–23. doi: 10.1079/PAVSNNR201611014

Colomina, I., and Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS J. Photogrammetry Remote Sens.* 92, 79–97. doi: 10.1016/j.isprsjprs.2014.02.013

Costa, J. M., Grant, O. M., and Chaves, M. M. (2013). Thermography to explore plant–environment interactions. *J. Exp. Bot.* 64, 3937–3949. doi: 10.1093/jxb/ert029

DadrasJavan, F., Samadzadegan, F., Seyed Pourazar, S. H., and Fazeli, H. (2019). UAV-based multispectral imagery for fast Citrus Greening detection. *J. Plant Dis. Prot.* 126, 307–318. doi: 10.1007/s41348-019-00234-8

Dalsass, M., Scheuerpflug, H., Fecher, F. W., Buerhop-Lutz, C., Camus, C., and Brabec, C. J. (2016). "Correlation between the generated string powers of a photovoltaic: Power plant and module defects detected by aerial thermography," in *2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC)*, Portland, OR, USA. 3113–3118. doi: 10.1109/PVSC.2016.7750239

Darwin, B., Dharmaraj, P., Prince, S., Popescu, D. E., and Hemanth, D. J. (2021). Recognition of bloom/yield in crop images using deep learning models for smart agriculture: a review. *Agronomy* 11, 646. doi: 10.3390/agronomy11040646

Ding, S., Xu, X., and Nie, R. (2014). Extreme learning machine and its applications. *Neural Computing Appl.* 25, 549–556. doi: 10.1007/s00521-013-1522-8

Feng, S., Cao, Y., Xu, T., Yu, F., Zhao, D., and Zhang, G. (2021). Rice leaf blast classification method based on fused features and one-dimensional deep convolutional neural network. *Remote Sens.* 13, 3207. doi: 10.3390/rs13163207

Feng, Z., Song, L., Duan, J., He, L., Zhang, Y., Wei, Y., et al. (2021). Monitoring wheat powdery mildew based on hyperspectral, thermal infrared, and RGB image data fusion. *Sensors* 22, 31. doi: 10.3390/s22010031

Feng, Z., Song, L., Zhang, S., Jing, Y., He, L., Li, F., et al. (2022). Wheat Powdery Mildew monitoring based on information fusion of multi-spectral and thermal infrared images acquired with an unmanned aerial vehicle. *Sci. Agric. Sin.* 55, 890–906. doi: 10.3864/j.issn.0578-1752.2022.05.005

Francesconi, S., Harfouche, A., Maesano, M., and Balestra, G. M. (2021). UAV-based thermal, RGB imaging and gene expression analysis allowed detection of Fusarium head blight and gave new insights into the physiological responses to the disease in durum wheat. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.628575

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

Gao, D., Sun, Q., Hu, B., and Zhang, S. (2020). A framework for agricultural pest and disease monitoring based on internet-of-things and unmanned aerial vehicles. *Sensors* 20, 1487. doi: 10.3390/s20051487

García, L., Parra, L., Jimenez, J. M., Lloret, J., Mauri, P. V., and Lorenz, P. (2020). DronAway: A proposal on the use of remote sensing drones as mobile gateway for WSN in precision agriculture. *Appl. Sci.* 10 (19), 6668. doi: 10.3390/app10196668

Gayathri, S., Wise, D. J. W., Shamini, P. B., and Muthukumaran, N. (2020). "Image analysis and detection of tea leaf disease using deep learning," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India. 398–403. doi: 10.1109/ICESC48915.2020.9155850

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

He, X. (2018). Rapid development of unmanned aerial vehicles (UAV) for plant protection and application technology in China. *Outlooks Pest Manage.* 29, 162–167. doi: 10.1564/v29_aug_04

Hecht-Nielsen, R. (1992). Theory of the backpropagation neural network. In. *Neural Networks Percept.*, 65–93. doi: 10.1016/B978-0-12-741252-8.50010-8

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. doi: 10.48550/arXiv.1704.04861. *arXiv preprint*.

Huang, M., Xu, G., Li, J., and Huang, J. (2021). A method for segmenting disease lesions of maize leaves in real time using attention YOLACT++. *Agriculture* 11, 1216. doi: 10.3390/agriculture11121216

Hunt, E.R. Jr, and Daughtry, C. S. (2018). What good are unmanned aircraft systems for agricultural remote sensing and precision agriculture? *Int. J. Remote Sens.* 39, 5345–5376. doi: 10.1080/01431161.2017.1410300

Jiang, Q., Wu, G., Tian, C., Li, N., Yang, H., Bai, Y., et al. (2021). Hyperspectral imaging for early identification of strawberry leaves diseases with machine learning and spectral fingerprint features. *Infrared Phys. Technol.* 118, 103898. doi: 10.1016/j.infrared.2021.103898

Kale, M. R., and Shitole, M. S. (2021). Analysis of crop disease detection with SVM, KNN and random forest classification. *Inf. Technol. Industry* 9, 364–372.

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

Kannojia, S. P., and Jaiswal, G. (2018). Effects of varying resolution on performance of CNN based image classification: An experimental study. *Int. J. Comput. Sci. Eng.* 6, 451–456. doi: 10.26438/ijcse/v6i9.451456

Kerkech, M., Hafiane, A., and Canals, R. (2020). Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach. *Comput. Electron. Agric.* 174, 105446. doi: 10.1016/j.compag.2020.105446

Khan, M. A., Alqahtani, A., Khan, A., Alsubai, S., Binbusayyis, A., Ch, M. M. I., et al. (2022). Cucumber leaf diseases recognition using multi-level deep entropy-ELM feature Selection. *Appl. Sci.* 12, 593. doi: 10.3390/app12020593

Kouadio, L., Deo, R. C., Byrareddy, V., Adamowski, J. F., and Mushtaq, S. (2018). Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties. *Comput. Electron. Agric.* 155, 324–338. doi: 10.1016/j.compag.2018.10.014

Kuska, M. T., Wahabzada, M., and Paulus, S. (2024). AI for crop production–Where can large language models (LLMs) provide substantial value? *Comput. Electron. Agric.* 221, 108924. doi: 10.1016/j.compag.2024.108924

Lan, Y., Chen, S., Deng, J., Zhou, Z., and Ouyang, F. (2019). Development situation and problem analysis of plant protection unmanned aerial vehicle in China. *J. South China Agric. Univ.* 40, 217–225. doi: 10.7671/j.issn.1001-411X.201905082

Lan, Y., Huang, Z., Deng, X., Zhu, Z., Huang, H., Zheng, Z., et al. (2020). Comparison of machine learning methods for citrus greening detection on UAV multispectral images. *Comput. Electron. Agric.* 171, 105234. doi: 10.1016/j.compag.2020.105234

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791

Lei, S., Luo, J., Tao, X., and Qiu, Z. (2021). Remote sensing detecting of yellow leaf disease of Arecanut based on UAV multisource sensors. *Remote Sens.* 13, 4562. doi: 10.3390/rs13224562

Li, K., Lin, J., Liu, J., and Zhao, Y. (2020). Using deep learning for Image-Based different degrees of ginkgo leaf disease classification. *Information* 11, 95. doi: 10.3390/info11020095

Li, L., Lu, J., Wang, S., Ma, Y., Wei, Q., Li, X., et al. (2016). Methods for estimating leaf nitrogen concentration of winter oilseed rape (Brassica napus L.) using in *situ* leaf spectroscopy. *Ind. Crops Products* 91, 194–204. doi: 10.1016/j.indcrop.2016.07.008

Licciardi, G., Marpu, P. R., Chanussot, J., and Benediktsson, J. A. (2011). Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles. *IEEE Geosci. Remote Sens. Lett.* 9, 447–451. doi: 10.1109/LGRS.2011.2172185

Liu, W., Cao, X., Fan, J., Wang, Z., Yan, Z., Luo, Y., et al. (2018). Detecting wheat powdery mildew and predicting grain yield using unmanned aerial photography. *Plant Dis.* 102, 1981–1988. doi: 10.1094/PDIS-12-17-1893-RE

Liu, Y., Yan, Z., Tan, J., and Li, Y. (2023). Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model. *IEEE Trans. Circuits Syst. Video Technol.* 33, 1643–1657. doi: 10.1109/TCSVT.2022.3214430

Liu, M., Zhang, Z., Liu, X., Yao, J., Du, T., Ma, Y., et al. (2020). Discriminant analysis of the damage degree caused by pine shoot beetle to yunnan pine using UAV-based hyperspectral images. *Forests* 11, 1258. doi: 10.3390/f11121258

Lu, B., Dao, P. D., Liu, J., He, Y., and Shang, J. (2020). Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sens.* 12, 2659. doi: 10.3390/rs12162659

Lu, G., Li, S., Mai, G., Sun, J., Zhu, D., Chai, L., et al. (2023). AGI for agriculture. doi: 10.48550/arXiv.2304.06136. *arXiv preprint*.

Luo, J., Huang, W., Zhao, J., Zhang, J., Zhao, C., and Ma, R. (2013). Detecting aphid density of winter wheat leaf using hyperspectral measurements. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 6, 690–698. doi: 10.1109/JSTARS.2013.2248345

Ma, S., G. Z., and Wang, S. (2021). Hyperspectral remote sensing monitoring of chinese chestnut red mite insect pests in UAV. *Trans. Chin. Soc. Agric. Machinery* 52, 171–180.

Mahesh, B. (2020). Machine learning algorithms-a review. *Int. J. Sci. Res. (IJSR)* 9, 381–386. doi: 10.21275/ART20203995

Mardanisamani, S., Maleki, F., Hosseinzadeh Kassani, S., Rajapaksa, S., Duddu, H., Wang, M., et al. (2019). "Crop lodging prediction from UAV-acquired images of wheat and canola using a DCNN augmented with handcrafted texture features," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*.

Marin, D. B., Santana, L. S., Barbosa, B. D. S., Barata, R. A. P., Osco, L. P., Ramos, A. P. M., et al. (2021). Detecting coffee leaf rust with UAV-based vegetation indices and decision tree machine learning models. *Comput. Electron. Agric.* 190, 106476. doi: 10.1016/j.compag.2021.106476

Mittal, S. (2019). A Survey on optimized implementation of deep learning models on the NVIDIA Jetson platform. *J. Syst. Architecture* 97, 428–442. doi: 10.1016/j.sysarc.2019.01.011

Mukherjee, A., Misra, S., and Raghuwanshi, N. S. (2019). A survey of unmanned aerial sensing solutions in precision agriculture. *J. Network Comput. Appl.* 148, 102461. doi: 10.1016/j.jnca.2019.102461

Nazarenko, E., Varkentin, V., and Polyakova, T. (2019). "Features of application of machine learning methods for classification of network traffic (features, advantages, disadvantages)," in *2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*, Vladivostok, Russia. 1–5. doi: 10.1109/FarEastCon.2019.8934236

Neupane, K., and Baysal-Gurel, F. (2021). Automatic identification and monitoring of plant diseases using unmanned aerial vehicles: A review. *Remote Sens.* 13, 3841. doi: 10.3390/rs13193841

Oh, S., Lee, D.-Y., Gongora-Canul, C., Ashapure, A., Carpenter, J., Cruz, A. P., et al. (2021). Tar spot disease quantification using unmanned aircraft systems (UAS) data. *Remote Sens.* 13, 2567. doi: 10.3390/rs13132567

Pang, T., Rao, L., Chen, X., and Cheng, J. (2021). Impruved prediction of soluble solid content of apple using a combination of spectral and textural features of hyperspectral images. *J. Appl. Spectrosc.* 87, 1196–1205. doi: 10.1007/s10812-021-01129-z

Peñuelas, J., Filella, I., Lloret, P., Mun Oz, F., and Vilajeliu, M. (1995). Reflectance assessment of mite effects on apple trees. *Int. J. Remote Sens.* 16, 2727–2733. doi: 10.1080/01431169508954588

Qi, T., Guo, J., Peng, H., Liu, P., Kang, Z., and Guo, J. (2019). Host-induced gene silencing: A powerful strategy to control diseases of wheat and barley. *Int. J. Mol. Sci.* 20, 206. doi: 10.3390/ijms20010206

Qing, J., Deng, X., Lan, Y., and Li, Z. (2023). GPT-aided diagnosis on agricultural image based on a new light YOLOPC. *Comput. Electron. Agric.* 213, 108168. doi: 10.1016/j.compag.2023.108168

Radoglou-Grammatikis, P., Sarigiannidis, P., Lagkas, T., and Moscholios, I. (2020). A compilation of UAV applications for precision agriculture. *Comput. Networks* 172, 107148. doi: 10.1016/j.comnet.2020.107148

Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., and Dollár, P. (2020). "Designing network design spaces," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10428–10436. doi: 10.48550/arXiv.2003.13678

Rangarajan, A. K., Purushothaman, R., and Ramesh, A. (2018). Tomato crop disease classification using pre-trained deep learning algorithm. *Proc. Comput. Sci.* 133, 1040–1047. doi: 10.1016/j.procs.2018.07.070

Rao, P. K. (2021). Cassava leaf disease classification using separable convolutions UNet. *Turkish J. Comput. Mathematics Educ. (TURCOMAT)* 12, 140–145.

Ren, Q., Zhang, R., Cai, W., Sun, X., and Cao, L. (2020). Application and development of new drones in agriculture. *IOP Conf. series: Earth Environ. Sci.* 440, 52041. doi: 10.1088/1755-1315/440/5/052041

Ribeiro-Gomes, K., Hernández-López, D., Ortega, J. F., Ballesteros, R., Poblete, T., and Moreno, M. A. (2017). Uncooled thermal camera calibration and optimization of the photogrammetry process for UAV applications in agriculture. *Sensors* 17, 2173. doi: 10.3390/s17102173

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in N. Navab, J. Hornegger, W. Wells and A. Frangi eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science*. vol 9351. (Cham: Springer). doi: 10.1007/978-3-319-24574-4_28

Shrestha, A., and Mahmood, A. (2019). Review of deep learning algorithms and architectures. *IEEE Access* 7, 53040–53065. doi: 10.1109/ACCESS.2019.2912200

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., et al. (2017). Mastering the game of go without human knowledge. *Nature* 550, 354–359. doi: 10.1038/nature24270

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. doi: 10.48550/arXiv.1409.1556. *arXiv preprint*.

Song, L., Wang, L., Yang, Z., He, L., Feng, Z., Duan, J., et al. (2022). Comparison of algorithms for monitoring wheat powdery mildew using multi-angular remote sensing data. *Crop J.* 10, 1312–1322. doi: 10.1016/j.cj.2022.07.003

Su, W.-H. (2020). Advanced machine learning in point spectroscopy, RGB-and hyperspectral-imaging for automatic discriminations of crops and weeds: A review. *Smart Cities* 3, 767–792. doi: 10.3390/smartcities3030039

Thangaraj, R., Anandamurugan, S., Pandiyan, P., and Kaliappan, V. K. (2021). Artificial intelligence in tomato leaf disease detection: a comprehensive review and discussion. *J. Plant Dis. Prot.* 129, 469–488. doi: 10.1007/s41348-021-00500-8

Thenkabail, P. S., Smith, R. B., and De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens. Environ.* 71, 158–182. doi: 10.1016/S0034-4257(99)00067-X

Tuia, D., Persello, C., and Bruzzone, L. (2016). Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geosci. Remote Sens. magazine* 4, 41–57. doi: 10.1109/MGRS.2016.2548504

van Iersel, W., Straatsma, M., Addink, E., and Middelkoop, H. (2018). Monitoring height and greenness of non-woody floodplain vegetation with UAV time series. *ISPRS J. Photogrammetry Remote Sens.* 141, 112–123. doi: 10.1016/j.isprsjprs.2018.04.011

Wallelign, S., Polceanu, M., and Buche, C. (2018). "Soybean plant disease identification using convolutional neural network," in *FLAIRS conference*. 146–151.

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., et al. (2024). Yolov10: Real-time end-to-end object detection. arXiv:1704.04861. doi: 10.48550/arXiv.1704.04861

Wang, G.-L., He, Z.-M., Zhu, H.-Y., Gao, Y.-G., Zhao, Y., Yang, H., et al. (2017). Involvement of serotonergic, noradrenergic and dopaminergic systems in the antidepressant-like effect of ginsenoside Rb1, a major active ingredient of Panax ginseng CA Meyer. *J. Ethnopharmacology* 204, 118–124. doi: 10.1016/j.jep.2017.04.009

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., et al. (2020). Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 3349–3364. doi: 10.1109/TPAMI.2020.2983686

Xiao, D., Pan, Y., Feng, J., Yin, J., Liu, Y., and He, L. (2022). Remote sensing detection algorithm for apple fire blight based on UAV multispectral image. *Comput. Electron. Agric.* 199, 107137. doi: 10.1016/j.compag.2022.107137

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1492–1500. doi: 10.48550/arXiv.1611.05431

Xie, C., Shao, Y., Li, X., and He, Y. (2015). Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging. *Sci. Rep.* 5, 1–11. doi: 10.1038/srep16564

Xie, C., Yang, C., and He, Y. (2017). Hyperspectral imaging for classification of healthy and gray mold diseased tomato leaves with different infection severities. *Comput. Electron. Agric.* 135, 154–162. doi: 10.1016/j.compag.2016.12.015

Xu, Z., Zhang, Q., Xiang, S., Li, Y., Huang, X., Zhang, Y., et al. (2022). Monitoring the severity of Pantana phyllostachysae Chao infestation in Moso bamboo forests based on UAV multi-spectral remote sensing feature selection. *Forests* 13, 418. doi: 10.3390/f13030418

Yan, Q., Yang, B., Wang, W., Wang, B., Chen, P., and Zhang, J. (2020). Apple leaf diseases recognition based on an improved convolutional neural network. *Sensors* 20, 3535. doi: 10.3390/s20123535

Yang, G., Li, C., Yu, H., Xu, B., Feng, H., Gao, L., et al. (2015). UAV based multi-load remote sensing technologies for wheat breeding information acquirement. *Trans. Chin. Soc. Agric. Eng.* 31, 184–190.

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

Yang, J., Ouyang, C., Dik, G., Corry, P., and ter Hofstede, A. H. (2022). "Crop harvest forecast via agronomy-informed process modelling and predictive monitoring," in X. Franch, G. Poels, F. Gailly and M. Snoeck Eds. *Advanced Information Systems Engineering. CAiSE 2022. Lecture Notes in Computer Science*. 201–217 (Cham: Springer). doi: 10.1007/978-3-031-07472-1_12

Yang, K., Yang, G. Y., and Fu, S. I. H. (2020). Research of control system for plant protection UAV based on Pixhawk. *Proc. Comput. Sci.* 166, 371–375. doi: 10.1016/j.procs.2020.02.082

Yinka-Banjo, C., and Ajayi, O. (2019). Sky-farmers: Applications of unmanned aerial vehicles (UAV) in agriculture. *Autonomous vehicles*, 107–128.

Zhang, J., Huang, Y., Pu, R., Gonzalez-Moreno, P., Yuan, L., Wu, K., et al. (2019). Monitoring plant diseases and pests through remote sensing technology: A review. *Comput. Electron. Agric.* 165, 104943. doi: 10.1016/j.compag.2019.104943

Zhang, M., Qin, Z., Liu, X., and Ustin, S. L. (2003). Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing. *Int. J. Appl. Earth Observation Geoinformation* 4, 295–310. doi: 10.1016/S0303-2434(03)00008-4

Zhao, J., Peng, Y., Sagar, D., and Zhang, L. (2013). On-line detection of apple surface defect based on image processing method. *Trans. Chin. Soc Agric. Mach.* 44, 260–263.

Zhao, J., Zhang, D., Luo, J., Dong, Y., Yang, H., and Huang, W. (2012). Characterization of the rice canopy infested with brown spot disease using field hyperspectral data. *Wuhan Univ. J. Natural Sci.* 17, 86–92. doi: 10.1007/s11859-012-0809-4

Zheng, Q., Huang, W., Cui, X., Shi, Y., and Liu, L. (2018). New spectral index for detecting wheat yellow rust using Sentinel-2 multispectral imagery. *Sensors* 18, 868. doi: 10.3390/s18030868

Zhou, L., Xiao, Q., Taha, M. F., Xu, C., and Zhang, C. (2023). Phenotypic analysis of diseased plant leaves using supervised and weakly supervised deep learning. *Plant Phenomics* 5, 22. doi: 10.34133/plantphenomics.0022

Abbreviation

IA	intelligent agriculture
UAV	unmanned aerial vehicle
RS	remote sensing
AI	artificial intelligence
DL	deep learning
ML	machine learning
LLM	large language model
LVM	large vision model
FAO	food and agriculture organization
ELISA	enzyme-linked immunosorbent assay
IF	immunofluorescence
PCR	polymerase chain reaction
FISH	fluorescence *in situ* hybridization
AGI	artificial general intelligence
WOS	web of science
PCA	principal component analysis
VIS	visible light
NIR	near-infrared
SG	savitizky-golay convolution smoothing
SNV	standard normal variate
MSC	multivariate scatter correction
WT	wavelet transform
CON	contrast
COR	correlation
EN	entropy
HO	homogeneity
GLCM	gray level co-occurrence matrix
SPA	successive projections algorithm
GAPLS	genetic algorithm partial least squares
UVE	uninformative variable elimination algorithm
CARS	competitive adaptive reweighted sampling
RF	random forest
ELM	extreme learning machine
SVM	support vector machine
BPNN	back propagation neural network
DNN	deep neural network
CNN	convolutional neural network
SOM	self-organizing map
LLM	large language model
YOLO	you-only-look-once
LVM	large vision model
GPU	graphics processing unit
GRVI	green ratio vegetation index
GNDVI	green normalized difference vegetation index
NDVI	normalized difference vegetation index
LMT	logistic model tree
REPTree	reduced-error pruning tree
RT	regression tree
KNN	k-nearest neighbors
DF-UNet	double-flow UNet
HLS	hue-luminance-saturation
OSAVI	optimized soil adjusted vegetation index
LCI	leaf chlorophyll index
NDRE	normalized difference red edge
MLLM	multi-modal large language model.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 December 2024
DOI 10.3389/fpls.2024.1481909

Check for updates

OPEN ACCESS

EDITED BY
Ning Yang,
Jiangsu University, China

REVIEWED BY
Zhang Guozhong,
Huazhong Agricultural University, China
Yinyan Shi,
Nanjing Agricultural University, China
Elio Romano,
Centro di ricerca per l'Ingegneria e le Trasformazioni agroalimentari (CREA-IT), Italy

*CORRESPONDENCE
Ganran Deng
dengganran@163.com
Guojie Li
guojie88@163.com

†These authors share first authorship

RECEIVED 16 August 2024
ACCEPTED 15 November 2024
PUBLISHED 03 December 2024

CITATION
Yan B, Cui Z, Deng G, Li G, Zheng S, He F, Li L, Chen P, Wang X, Zhou S, Dai Y, Qin S and Liu Z (2024) Design and validation of a real-time cassava planter seed quality monitoring system based on optical fiber sensors and rotary encoders. *Front. Plant Sci.* 15:1481909. doi: 10.3389/fpls.2024.1481909

COPYRIGHT
© 2024 Yan, Cui, Deng, Li, Zheng, He, Li, Chen, Wang, Zhou, Dai, Qin 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.

Design and validation of a real-time cassava planter seed quality monitoring system based on optical fiber sensors and rotary encoders

Bin Yan[1,2†], Zhende Cui[1,2†], Ganran Deng[1,2*], Guojie Li[1,2*], Shuang Zheng[1,2], Fengguang He[1,2], Ling Li[1,2], Pinlan Chen[1,2], Xilin Wang[1,2], Sili Zhou[1,2], Ye Dai[1,2], Shuangmei Qin[1,2] and Zehua Liu[1,2]

[1]Agricultural Machinery Research Institute, Chinese Academy of Tropical Agricultural Sciences, Zhanjiang, Guangdong, China, [2]Key Laboratory of Tropical Agricultural Machinery, Ministry of Agriculture and Rural Affairs, Zhanjiang, Guangdong, China

Targeting the issues of seed leakage and cutting segment adhesion due to poor seed feeding and cutting in real-time seed-cutting cassava planters, this study developed a seeding quality monitoring system. Based on the structure and working principle of the seed cutting and discharging device, the installation methods of the matrix fiber optic sensor and rotary encoder were determined. By combining the operational characteristics of the planter's ground wheel drive with seed cutting and seed dropping, a monitoring model correlating the sowing parameters with seed dropping time was established; a monitoring window was created by extracting and processing the rotary encoder pulse signal, and the number of seeds sown after each opposing cutter's operation was calculated based on the pulse width information within the monitoring window. The monitoring system's statistics were compared and analyzed with the manual statistics, and the bench test showed that the monitoring system designed in this study offers high accuracy. When the simulated rotational speed of the opposing cutter ranges from 10 to 30 rpm, the average monitoring error between the monitored and actual seeding quantities for the left and right rows is less than 1.4%. The monitoring system can promptly and accurately activate sound and light alarms for faults, achieving a 100% success rate in alarms and an average fault response time of less than 0.4 seconds. Field tests demonstrate that the average error in seeding volume is 0.91%, and the monitoring system can timely alert to faults occurring in the planter. The system fulfills the requirements for real-time monitoring of cassava seeding volume at various operating speeds in field conditions, and can serve as a reference for monitoring operational parameters in subsequent cassava combine harvesters.

KEYWORDS

cassava combine planter, cassava seed stalk, seeding quality monitoring, fault monitoring, sensor

1 Introduction

Cassava is a crucial crop extensively cultivated in tropical and subtropical regions, playing a vital role in global food security and economic development (He et al., 2023; Chenna et al., 2024). Seeding quality, as a fundamental aspect of cassava production, directly influences both crop yield and health (Cay et al., 2018; Albarenque et al., 2023). Due to its benefits in enhancing sowing accuracy, crop yield, and farmland management (Rossi et al., 2023), precision sowing technology has become integral to modern agricultural practices (Tang et al., 2022; Ling et al., 2024; Jin et al., 2024a). With ongoing advancements in agricultural mechanization, real-time cut-seed sowing technology is increasingly being employed in cassava planting to address various challenges associated with mechanized sowing (Lungkapin et al., 2009; Chen et al., 2023). However, when a cassava planting machine operates in a fully closed state, the operator cannot directly monitor seeding quality (Bai et al., 2022), leading to issues such as seed leakage and cut-section sticking (Xing et al., 2020), which reduce cassava yield (Wang et al., 2020). Consequently, developing an efficient seeding quality monitoring system is essential.

Failures affecting planter seeding can arise from various factors, including the planter, seed stalk, and operator. The length of individual seed poles is non-standardized, making it difficult to regulate the cutting length accurately at the end of the cut; seed stalks can become bent due to typhoons, preventing their timely and efficient transfer to the cutter position; the operator failed to fill the seed rods to the seed release point in a timely manner during the planting process while the planter continued to move forward; all these factors may result in planting faults. Prolonged and high-intensity operation of the planter can lead to a shift in the relative position of the opposing cutter or a reduction in sharpness, resulting in incomplete cuts that cause adhesion. Real-time monitoring of seeding passes and failures, along with timely feedback to operators and farmers, plays a crucial role in ensuring high cassava yields and accurately predicting crop yields (Karimi et al., 2019).

The performance of the seeding quality monitoring system reflects the seeding effectiveness of planters in real time. To achieve effective seeding quality monitoring, the design of the monitoring system and the integration of advanced sensor technologies are essential. Recent advancements in sensor technology have significantly improved the potential for enhancing seeding quality monitoring (Zhao et al., 2023; Jin et al., 2024). Optical sensors, known for their low cost and high accuracy, have become prevalent in precision agriculture (Xie et al., 2021a). For instance, Wang et al. developed a sowing quality monitoring system for small-sized seeds using an opposed matrix fiber optic sensor, which achieved a counting relative error of no more than 4.6% and was capable of issuing alerts when a seed hole was blocked (Wang et al., 2022). Tang et al. employed diffuse reflective optoelectronic sensors and rectangular fiber optic sensors to monitor seed counts during the maize sowing process and observed a decline in monitoring accuracy as the rotational speed of the seed metering disk increased (Tang et al., 2022). Additionally, other researchers have explored various monitoring techniques, including image recognition, piezoelectric, and capacitive methods, to enhance the seeding process (Wang et al., 2019; Bai et al., 2021; Zhang et al., 2022; Xu et al., 2023). However, image recognition methods encounter challenges related to handling large volumes of data, complex structures, and higher costs during application. Meanwhile, the monitoring accuracy of piezoelectric and capacitive sensing technologies is highly susceptible to interference from environmental factors such as vibration, temperature, and humidity, leading to significant fluctuations in accuracy, these methods face practical limitations in field applications (Jiang et al., 2021; Chen et al., 2019).

Examining the current status of existing research on seeding monitoring systems, it is evident that most research focuses on the monitoring of small-sized seeds (Basso et al., 2016; Zhao et al., 2018; Xie et al., 2021b), and the quality of sowing is predominantly assessed by estimating the number of seeds, rather than their exact count. Cassava is an asexually propagated crop (Sheat et al., 2024)with large seed fragments that result in poor mobility, and the uneven terrain in cassava-growing areas complicates the accurate control of the planter's forward speed (Chen et al., 2022; Xie et al., 2023), making it difficult to evaluate the quality of monitoring using the existing system. Furthermore, cassava seeders are prone to two types of operational faults: missed seeding and cut-seed sticking. Existing seeding quality monitoring systems typically detect only two types of faults: missed sowing and re-sowing (Xie et al., 2022), and lack a real-time model and system for monitoring cut-seed sticking. This makes it challenging to effectively meet the real-time monitoring needs of cassava planting.

To address these issues, this study focuses on the accurate monitoring of sowing parameters for cassava planters, and designs a real-time cut-section sowing quality monitoring system incorporating a rotary encoder and a matrix fiber optic sensor. By extracting and processing pulse signals to obtain temporal characteristic parameters within the monitoring window, seeding parameters are accurately calculated, and the system's effectiveness and reliability were evaluated through bench and field tests. The real-time seed cut seeding quality monitoring scheme proposed in this study offers valuable insights for the development and optimization of seeding quality monitoring systems.

2 Materials and methods

2.1 Research carrier

The entire machine structure, as depicted in Figure 1, is designed by the research team as the research carrier for the monitoring system: the starter cassava joint planter. The planter primarily consists of a seed cutting device, a ridging device, a ridging protection device, a mulching device, a fertilizer discharge device, a ground wheel driving device, a transmission device, and a seed frame. The machine employs a wide and narrow double-row planting mode, with row spacing of 50 to 70 cm, a width of 160 to 180 cm, plant spacing of 60 to 80 cm, row height of 25 to 30 cm, and

FIGURE 1
Schematic diagram of cassava combine planter.

sowing depth of 5 to 10 cm. It can complete a one-time cassava sowing operation for 1 or 2 rows. The planting machine performs real-time seed cutting and sowing operations through a three-point suspension device under the tractor's traction. During operation, the cassava seed stems are manually placed into the seeding area. The ridging device creates rows under the drive of the transmission system, the furrow opener forms grooves in the rows, and simultaneously, the ground wheel driving device rotates the seed cutting device to cut the cassava seed stems in real time. These cut stems then fall into the grooves, completing sowing through the seed guide tube. Finally, the mulching device, protected by the ridge guarding device, covers the soil in the grooves to complete mulching. This planting machine can simultaneously perform the functions of ridging, furrowing, fertilizing, and planting, significantly improving planting efficiency to 4-6 mu per hour, and meeting the mechanized planting needs of cassava.

2.2 Design of seeding quality monitoring system

2.2.1 System fundamentals

The working process of the real-time seed cutting and discharging unit can be divided into four stages: seed release, transmission, seed cutting, and seed discharging, as shown in Figure 2. During operation, manually place the seed stems into the seed release hopper. The seed stems are then transported to the seed cutting area via clamping and transmission by the soft cartridges. The opposite cutter, driven by the ground wheel drive device, cuts off the cassava seed stems, and the cassava seeds are discharged by gravity. The real-time seed discharge process involves a simple hopper structure positioned close to the cutter. The seeds, after being cut, complete the discharge process through the hopper, undergoing free-fall movement, which is more stable. Therefore, it is preferable to install a seeding monitoring sensor in the hopper. Fiber optic sensors exhibit high sensitivity, strong anti-interference capabilities, and robust environmental adaptability, enabling stable performance in complex operating environments. This makes them highly effective for real-time, high-speed seed cutting and precise monitoring of cassava seeds. To ensure accurate monitoring of the rotational angle of the opposite cutter, the encoder is directly mounted on the rotary axis of the opposite cutter via a fixed bracket.

During seed rowing, cassava seeds flow through the photoelectric sensor monitoring area, causing variations in the intensity of the infrared beam, which in turn leads to changes in the voltage of the photoresistor at the receiving end, as illustrated in Figure 3A. The microcontroller employs the input capture interrupt to collect the pulse signals generated by cassava seeds flowing through the matrix fiber optic sensor group, thereby obtaining the seed rowing time sequence. The angle-splitting node is determined by monitoring the number of pulses from the rotary encoder. The sowing number monitoring window is established by two neighboring angle-splitting nodes, with each monitoring window corresponding to one sowing number, as illustrated in Figure 3B. The pulse width within the single monitoring window is analyzed to determine the seeding number. If there is a sowing leakage or a fault in the cutting section, the pulse width within the monitoring time window deviates from the theoretical threshold range, triggering the alarm circuit to activate both sound and light alarms (Wang et al., 2020).

2.2.2 Hardware system design

The hardware system primarily consists of a microcontroller, an opposed matrix fiber optic sensor, a fiber optic amplifier, an

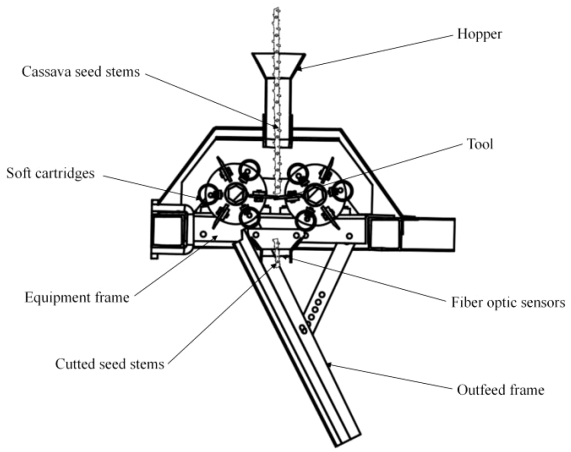


FIGURE 2
Simplified sketch of the arrangement of the seed cutting device and the seed discharge device.

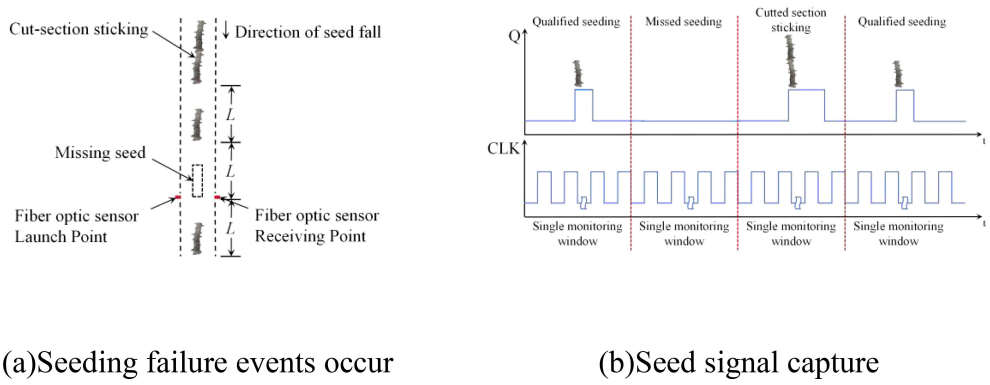


FIGURE 3
Seeding quality monitoring schematic. **(A)** Seeding faults. **(B)** Window monitoring principle.

incremental rotary encoder, an acousto-optic alarm, a keypad, and a digital display module, as illustrated in Figure 4. The microcontroller utilizes the ARM Cortex-M4-based N32G430C8L7 microprocessor as the primary control chip. It operates at a main frequency of 128 MHz and is equipped with 64 KB flash memory, 16 KB SRAM, 8 timers, and communication interfaces including 4 U(S)ART, 2 SPI, 2 IIC, and 1 CAN. This configuration adequately supports the high-speed seeding flow monitoring required for cassava in this study. The matrix fiber optic sensor, FT-70S, equipped with an NPN-type fiber optic amplifier, comprises a matrix sensor group designed to monitor seed flow during leakage from the planting machine. It utilizes infrared light-emitting diodes for emission and infrared photosensitive diodes for reception. The sensor has a minimum monitoring diameter of 0.4 mm and a maximum detection range of 70 mm × 110 cm. It effectively monitors seed flow after each opposing cutter operation, with a response time of less than 50 μs, and can differentiate pulse signals generated by various seeds.

To accurately measure the rotation angle intervals of the planter relative to the cutter shaft, an Omron E6B2-CWZ5B incremental rotary encoder (with a resolution of 1000 pulses per revolution) is employed. Fault alarm is a critical function of the seeding quality monitoring system. Given that cassava cultivation involves a real-time cut-and-mulch planting mode, it is essential to halt the planting operation promptly for inspection and troubleshooting in the event of seed shortages or sticky cutting sections. This study implements an alarm function to address two major faults: seed leakage and cutting section sticking. When a fault is detected by the system, the microcontroller drives the red and blue LED light-emitting modules to flash at varying intervals as part of the acoustic and optical alarms. Simultaneously, the system generates pulse-width modulation (PWM) signals with varying duty cycles to control the speakers, thereby providing acoustic and optical alarms. The components are electrically interconnected via power lines, signal lines, and communication lines to facilitate information acquisition and interaction. Specifically, the power line connects the tractor's onboard 12V DC power supply to each piece of equipment, while the communication line utilizes an RS485 serial communication network.

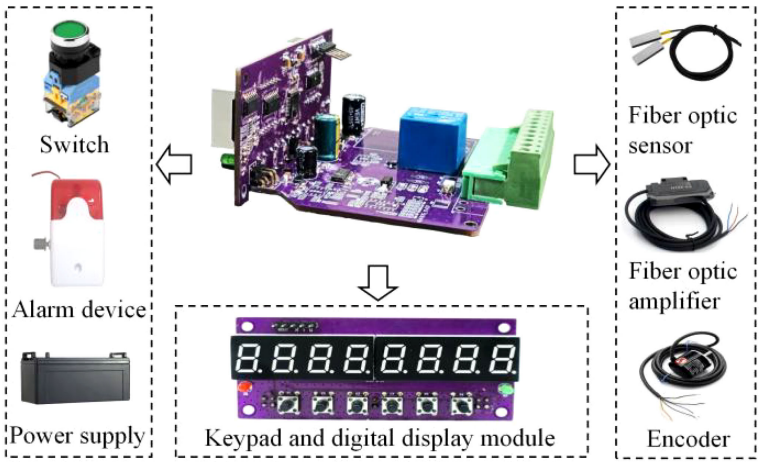


FIGURE 4
Schematic diagram of the hardware of the seeding quality monitoring system.

2.2.3 Seed metering algorithm

The seeding quality monitoring system is predicated on a matrix fiber optic sensor array and an incremental rotary encoder, which monitors the change in signal levels as seeds pass through adjacent cutters. Under normal operating conditions, the matrix fiber optic sensor array outputs a low-level signal to the microcontroller; when the cassava seed passes, the sensor array triggers an interrupt and outputs a high-level signal. After the seed passes, the system outputs a low-level signal again. The microcontroller collects the pulse width via the timer and subsequently calculates the duration of the cassava seed's passage. Based on the duration of the passage time, the microcontroller determines whether there is leakage during seeding or instances of adhesion in the cut sections, and it controls the alarm unit to produce appropriate auditory and visual signals. The maximum forward speed of the cassava planting machine designed by the research team is 1.39 m/s (5 km/h). The minimum spacing between individual seeds is 0.06 m, and the minimum time interval between the drop of two adjacent seeds is

$$T_{min} = \frac{S_{min}}{V_{max}} \tag{1}$$

Where T_{min} is the minimum seeding interval, s; S_{min} is the minimum planting spacing of single grain, m; V_{max} is the maximum forward speed of the planter, m/s.

Calculations indicate that the minimum seeding interval Tmin=0.043s=43ms, which is significantly larger than the duration of one cycle of the microcontroller control program (1–1.5 ms). Therefore, the developed program satisfies the requirements for single-grain seeding intervals.

$$\theta = \frac{M}{R} \times 360^{\circ} \tag{2}$$

The cutter rotary axis of the cassava planter consists of three pairs of cutters; therefore, the rotation angle of each neighboring pair of cutters is set to 120°. Consequently, the monitoring window for the number of seeds sown is defined by this rotation angle, with each monitoring window corresponding to one seed sown. The real-time rotation angle of the planter's cutter shaft is measured using a rotary encoder. During operation, the number of pulses generated by the rotary encoder is recorded in real time through a high-speed counter, enabling the determination of the rotation angle of the cutter shaft based on the encoder's resolution.

where θ is the theoretical angle of rotation of the rotary axis of the grower cutter, °; M is the number of pulses of the encoder, and R is the resolution of the encoder, P/R.

The duration for which a cassava seed blocks light in the sensor monitoring area is related to the height from which the seed drops; this duration is closely associated with sowing quality. The cassava seeds complete the seeding process when they pass through a pair of cutters, with these cutters serving as the initial reference position. The time for the seed's head and tail to reach the sensor monitoring region is calculated as follows

$$h = \frac{1}{2} g t_1^2 \tag{3}$$

$$h + l = \frac{1}{2} g t_2^2 \tag{4}$$

Where h is the height at which the first end of the cassava seed starts to fall to the monitoring area when it leaves the opposite cutter, m; l is the length of the cassava seed cut off, m; g is the acceleration of gravity, m/s^2; t_1 is the time used by the first end of the cassava seed to arrive at the monitoring area, s; and t_2 is the time used by the last end of the cassava seed to arrive at the monitoring area, s.

The time t_s at which the cassava seed as a whole passes through the monitoring area is

$$t_s = t_2 - t_1 = \frac{\sqrt{2(h+l)} - \sqrt{2h}}{\sqrt{g}} \tag{5}$$

Through experimental testing involving the use of a cutting device with three pairs of cutters, the minimum length of the cassava mechanized planting section is determined to be lmin=0.13m and the standard length is l=0.17m. When the seed section departs from the opposite cutter, its leading edge begins to drop to the monitoring area at a height of h=0.05m. Using the gravitational acceleration g=9.8m/s2 and substituting lmin=0.13m into Equation 5, calculated that the minimum seed completely blocked the beam time required for 0.09s. The selected fiber optic sensor has a response time of less than 200 μs, which meets the monitoring requirements for cassava seeding quality. Similarly, substituting l=0.17 m into Equation 5 yields that the time required for the standard-sized seed to completely block the beam is ts=0.11 s.

Combined with the theoretical spacing to determine the planting machine work sowing state (Abdolahzare and Abdanan Mehdizadeh, 2018), according to GB/T6973-2005 "single grain (precision) planter test methods" can be seen, when the actual spacing is less than or equal to 0.5 times the theoretical spacing for resowing; the actual spacing is greater than 1.5 times the theoretical spacing for missing, between the two for qualified, that is, the

Resowing

$$d \le 0.5 d_0 \tag{6}$$

Missing

$$d > 1.5 d_0 \tag{7}$$

Qualifying

$$0.5 d_0 < d \le 1.5 d_0 \tag{8}$$

Where d_0 is the theoretical plant spacing, m; d is the actual plant spacing, m.

In practice, within the rotation angle monitoring window of two adjacent cutters, 0.5 times the time interval t_s is set as the threshold for missed seeding detection, while 1.5 times t_s is used as the threshold for cut section sticking detection. If t ≤ 0.5ts,the planter is judged to have missed sowing; if t>1.5ts,it is judged to have cut section sticking; if 0.5ts<t ≤ 1.5ts,the planter is considered to be performing acceptable sowing. The parameter calculation flow is illustrated in Figure 5.

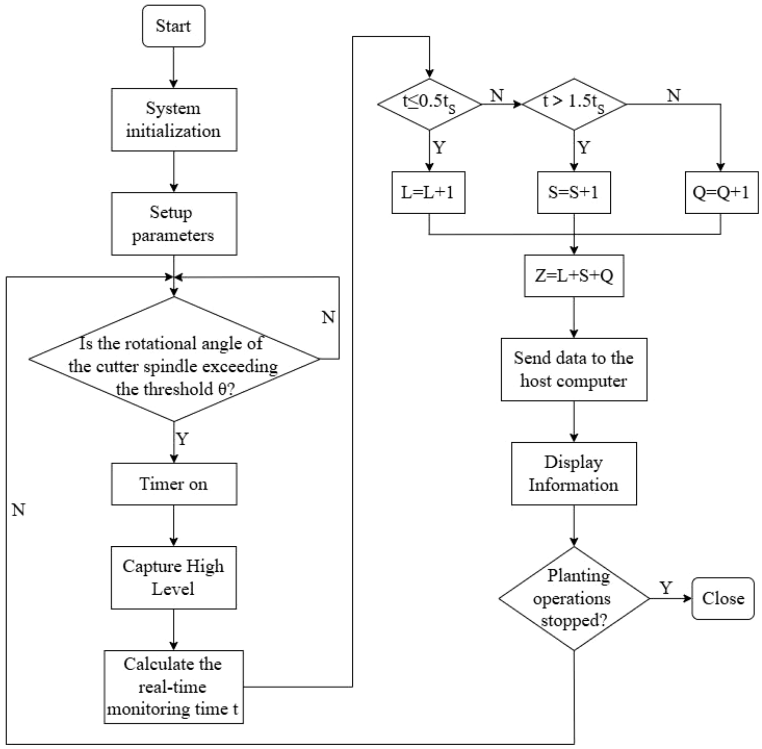


FIGURE 5
Flowchart for calculating seeding parameters. (L is the amount of missed seeding; S is the amount of cut section sticking; Q is the amount of qualified seeding; Z is the total amount of seeding; θ is the rotation angle threshold of the cutter rotary axis, 120°; t is the actual time of the seed section passing through the monitoring area, s; t_s is the theoretical time of the standard seed section passing through the monitoring area, s.).

3 Testing of seeding quality monitoring system

3.1 Monitoring of sowing parameters and comparative trials

To verify the stability and accuracy of the system, a calibration test was conducted at various working speeds of the rotary axis of the opposing cutter. South Plant 199 cassava seed stems were used as test samples. Specifically, one hundred cassava seed stems were randomly selected, with an average moisture content of 37.65%. The length of these seed stems ranged from 1.4 to 1.7 meters, and their diameters ranged from 16 to 28 millimeters. The test equipment included a real-time seed cutting device, a monitoring system, a computer, dynamic torque transducers, a drive motor, and a motor frequency converter, as shown in Figure 6. The test was performed at the Key Laboratory of Agricultural Equipment for Tropical Crops, Ministry of Agriculture and Rural Affairs, Agricultural Machinery Research Institute, Chinese Academy of Tropical Agricultural Sciences.

In the test, according to GB/T6973-2005 "Test Methods for Single Grain (Precision) Seeders," the rotational speed of the rotor shaft of the opposing cutter was chosen as the test factor (Xie et al., 2021c), with speeds of 10 r/min, 15 r/min, 20 r/min, 25 r/min, and 30 r/min. This speed gradient was representative of the actual working conditions for the cassava joint planting machine. The manual method was used to determine the actual number of cassava seed segments cut by the real-time seed cutting device and compare this with the values recorded by the monitoring system. Absolute and relative errors in sowing volume were calculated to assess the stability and accuracy of the monitoring system. Each test involved measuring 200 cassava seed segments, and the procedure was repeated three times.

3.2 Fault alarm response reliability test

The reliability of the fault alarm response was assessed by confirming that the monitoring system could record and generate alarms accurately and in a timely manner when faults were encountered. The test simulated two of the most critical and frequently occurring faults: a lack of cassava seed stems in the seed-cutting device and sticking due to incomplete cuts in the opposite sections of the seed-cutting device. The operation of the seeding quality monitoring system was managed through the start-stop button on the controller. During the seeding process, when cassava seed stems were not manually inserted while the seed-cutting device operated normally, it indicated that no seeds were dropped, simulating a seeding failure of the planter. Additionally, cassava seed stems were manually added to the hopper, the seed-cutting device operated normally, and the relative position and sharpness of the opposite cutter were adjusted to simulate incomplete cuts, representing the sticking fault. The two fault

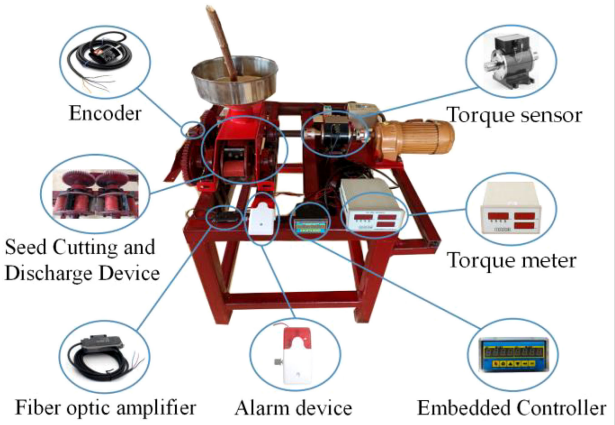


FIGURE 6
Seeding quality monitoring test device.

FIGURE 7
Field test.

simulation tests were repeated 100 times each. During the tests, the accuracy of the fault alarm response was evaluated by observing if the acoustic and visual alarms activated when faults occurred, and whether the missed seeding and cut-section sticking were accurately displayed on the digital module. Simultaneously, a stopwatch was used to record the time elapsed from the occurrence of the fault to the monitoring system's response.

3.3 A trial of real-time online monitoring of seeding parameters for planters in field operations

The seeding quality monitoring test described above was conducted at a fixed rotational speed. During field operations, the rotational speed of the opposing cutter rotor shaft varies in real-time with the ground wheel drive of the combine planter, which results in more complex field operating conditions. To assess the effectiveness of the seeding quality monitoring system during actual field operations, a performance test of the monitoring system was carried out in Zhanjiang City, Guangdong Province, as illustrated in Figure 7. The test equipment and materials included cassava seed stems of South Plant No. 199, a starter cassava combine planter, a Dongfanghong 904 tractor, a monitoring system, a tape measure, a shovel, and other related items.

The influence of lighting conditions on the seeding quality monitoring system was examined prior to field seeding operations. The combine planter was elevated off the ground, and the seed cutting device was operated at 18 r/min under two conditions: no seed pole feeding and normal seed pole feeding. The monitoring system was activated, and tests were conducted under sunlight, artificial lighting, and shaded natural light conditions, respectively. The test results indicated that, under all three conditions, the terminal of the monitoring system recorded a value of zero when there was no seed rod feeding. Furthermore, the average relative error in seeding volume monitoring was 0.73% when the seed rods were fed normally. This suggests that natural light conditions in the field have minimal impact on the operation of the monitoring system.

During the test, in accordance with GB/T6973-2005 Test Methods for Single Grain (Precision) Planters, 20 meters were allocated at both the front and rear ends of the test field for speed adjustment of the planter and tractor. Additionally, 150 meters were designated as the data collection area for the planter's stable operation. Based on the actual operating conditions of the combine planter, the tractor's forward speed was set at three target speeds: 2 km/h, 3 km/h, and 4 km/h. Following sowing, the soil layer was manually removed to determine the total number of seeds sown by the planter. The effectiveness of the monitoring system was evaluated by comparing the system's count with manually obtained results. Each test group was repeated three times, and the average value was calculated.

4 Results and discussion

The sowing quality monitoring system was fully and accurately verified through indoor bench tests and field tests, and the results of each test were analyzed as follows:

4.1 Monitoring of sowing parameters and analysis of results of comparative trials

The results of monitoring the sowing volume of the cassava seed pole cutting device at various rotational speeds of the opposing cutter rotary axis are presented in Table 1 and Figure 8. Throughout the test, the system did not experience any white crash screens or data transmission failures. Based on Table 1 and Figure 8, the average absolute error between the monitored sowing volumes of the left and

TABLE 1 Results of seeding count trials.

Rotational speed of a rotating shaft (r/min)	Number	Left-hand side			Right-hand side		
		Actual number	Number of monitors	Relative error(%)	Actual number	Number of monitors	Relative error(%)
10	1	200	199	0.50	200	199	0.50
	2	200	197	1.50	200	197	1.50
	3	200	198	1.00	200	196	2.00
15	1	200	197	1.50	200	199	0.50
	2	200	198	1.00	200	197	1.50
	3	200	198	1.00	200	198	1.00
20	1	200	198	1.00	200	199	0.50
	2	200	199	0.50	200	200	0
	3	200	199	0.50	200	198	1.00
25	1	200	199	0.50	200	198	1.00
	2	200	198	1.00	200	199	0.50
	3	200	198	1.00	200	197	1.50
30	1	200	200	0	200	197	1.50
	2	200	199	0.50	200	200	0
	3	200	198	1.00	200	199	0.50

right rows and the actual values is ≤ 2.67, while the average relative error is ≤ 1.33%. These results demonstrate that the system exhibits high monitoring accuracy and satisfies the practical requirements.

In order to further investigate the influence of different rotational speeds of the opposing cutter rotor shaft on the monitoring accuracy of the system, the relative error of the sowing volume of the left and right rows was analyzed as a function of the rotational speed of the cutter rotor shaft, using IBM SPSS Statistics software for analysis and Duncan's Multiple Comparison Analysis for data processing. The results are presented in Table 2. The results indicated that the P-value for seeding volume in the left row across five rotational speeds was 0.306, and the P-

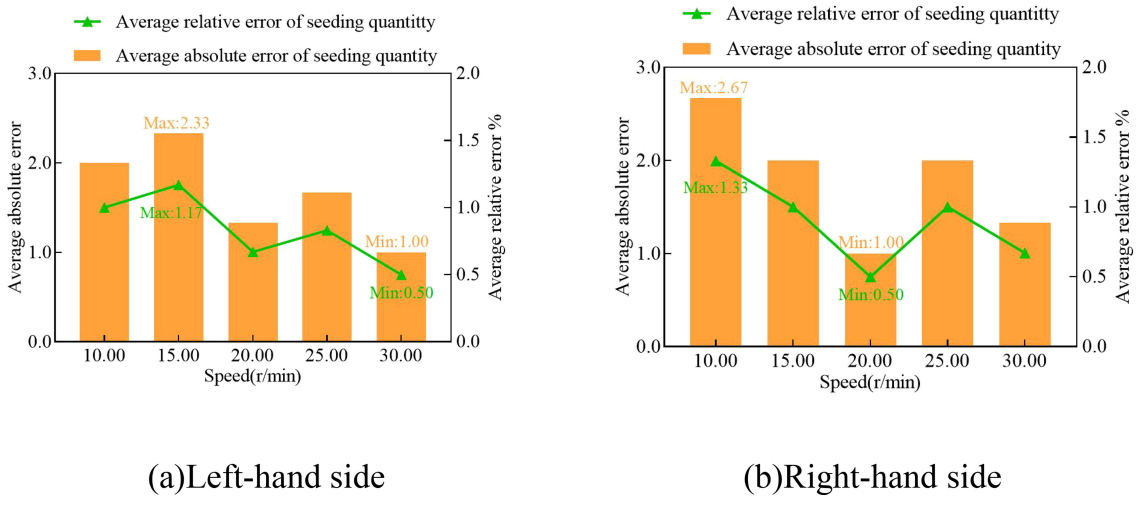


FIGURE 8
Seeding parameter monitoring comparison test results. **(A)** Left-hand side test results. **(B)** Right-hand side test results.

TABLE 2 Analysis of variance of seeding quantity under speed processing.

Monitoring channel	Cognitive	Sum of squares	df	Mean square	F	P
Left-hand side	Between groups	3.333	4	0.833	1.389	0.306
	Within groups	6.000	10	0.600		
	Total	9.333	14			
Right-hand side	Between groups	5.067	4	1.267	0.826	0.538
	Within groups	15.333	10	1.533		
	Total	20.400	14			

value for the right row was 0.538. Both P-values were significantly greater than 0.05, indicating that the differences were not statistically significant. This further indicates that the monitoring accuracy of the system remains relatively stable within the tested range of 10 to 30 r/min for the cutter spindle's rotational speed, and the impact of rotational speed on the monitoring accuracy is minimal, thus meeting the requirements for normal operation.

In summary, the seeding quality monitoring system developed by this institute demonstrates an accuracy comparable to manual statistics, with a minimal relative error. Moreover, the monitoring accuracy remains stable under varying rotational speeds, with minimal influence from the cutter spindle's rotational speed, thus meeting practical requirements.

4.2 Analysis of fault alarm response reliability test results

The fault simulation response test encompassed both missing seed segments and sticky seed segments, with the time interval from the initiation of the fault to the system alarm being recorded. A missing seed segment occurs when the manual seed release is delayed, causing the combine harvester to continue operating without any change in the system level within the monitoring window. Seed segment sticking involves a change in the relative position or sharpness of the seed cutting device's blade, leading to incomplete cutting and resulting in seed segments remaining in the monitoring system longer than the typical drop time. Test results indicate a 100% success rate for fault detection, with no instances of missed fault alarms, demonstrating the reliability of the system's audible and visual alarm functions. The scatter plot of fault alarm response times for the monitoring system is presented in Figure 9. Assuming that the manual timing response is negligible, the figure demonstrates that the response times for both faults are relatively stable, with an average value of less than 0.4 seconds. Therefore, the fault alarm response time of the sowing quality monitoring system is minimal and meets the requirements for practical operation. When a seeding operation fault occurs, the monitoring system provides timely and accurate alarms, alerting the operator to stop and inspect the equipment to prevent further economic losses.

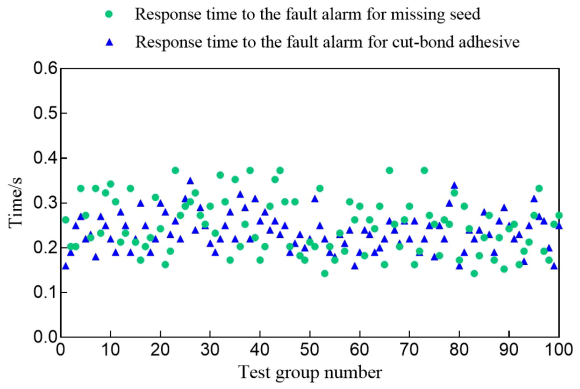


FIGURE 9
Scatter plot of the response time of monitoring system fault alarm.

In summary, the seeding quality monitoring system designed by this institute has a 100% success rate in response to faults when there are faults of missing seed segments and sticky seed segments, and the response time of both faults is relatively stable, with an average value of less than 0.4s, so that timely and accurate alarms can be issued.

4.3 Analysis of experimental results for real-time online monitoring of planter seeding parameters in field operations

The results of real-time monitoring of sowing parameters during the field operation of the cassava combine planter are presented in Table 3. The accuracy and reliability of the monitoring system under actual field conditions were validated through tests involving human-induced missed sowing and adjustments to the relative position and sharpness of the cutter. The table shows that the relative error interval for seeding volume monitoring ranged from 0.48% to 1.46%, with an average relative error of 0.91%, within the operating speed range of 2 to 4 km/h for agricultural implements. The fault alarm response success rate was 100%. The field tests demonstrate that the monitoring system remained operational over extended periods, with measurement results consistently stable. The system effectively monitored agricultural equipment across varying operating speeds and met the practical work requirements.

TABLE 3 Field performance experiment results of monitoring system.

Forward speed (km/h)	Rotational speed of a rotating shaft (r/min)	Number of seeds to be sown	Number of qualifications	Number of monitors	Relative error (%)	Number of faults (missed seeding and cutted section sticking)	Fault alarm rate (%)
2	11.89	215	202	200	0.99	13	100
		218	206	203	1.46	12	100
		220	209	207	0.96	11	100
3	17.84	221	207	205	0.97	14	100
		216	206	205	0.49	10	100
		220	208	207	0.48	12	100
4	23.78	224	214	211	1.42	10	100
		217	205	204	0.49	12	100
		225	213	211	0.94	12	100

5 Conclusions

This paper addresses a seeding quality monitoring system specifically designed for real-time cut-seed planters used for cassava. This research area has predominantly focused on monitoring the sowing of small and medium-sized seeds, and to date, no model or system for real-time cut-seed segment seeding quality monitoring has been identified. The starter cassava combine planter developed by the research team rarely suffers from ground wheel slippage in field operations. Therefore, the potential effects of ground wheel slippage have not been incorporated into the consideration of the system's monitoring performance in the current research design. Accordingly, the system was able to accurately monitor the seeding parameters of the combine planter, including the total amount of seed sown, amount of seed passed, amount of seed missed, and amount of cut-section sticking. In the event of a planter failure, the system promptly activates both sound and light alarms to prevent losses due to missed seeding and cut-section sticking. The fiber optic sensor exhibits strong dust penetration capabilities, making the monitoring system less susceptible to the complex operating environment in the field. However, since the entire system relies on light interactions, it appears infeasible to fully eliminate the impact of dust. In the complex field environment, the sensors are prone to dust accumulation during prolonged monitoring, which may reduce monitoring accuracy. The primary conclusions are as follows:

1. Based on an analysis of the working process of the seed-cutting device in the combine planter, a seeding quality monitoring system for cassava combine planters was designed. This system employs opposed fiber-optic sensors and rotary encoders as the primary monitoring elements, details the structure and working principle of the system, and establishes monitoring algorithms for each seeding state based on three potential operational scenarios in real-time seed-cutting operations.
2. According to the indoor test results, when the simulated rotational speed of the opposing cutter ranges from 10 to 30 rpm, the average absolute error between the monitored sowing volumes of the left and right rows and the actual values is 2.67, with an average relative error of 1.33%. Duncan's multiple comparison analysis indicates that the differences between the system's monitored values and manually measured values are not significant. The system's fault alarm response function is both accurate and timely, achieving a 100% alarm success rate and an average response time of less than 0.4 seconds. The indoor bench tests demonstrate that the monitoring system maintains high and relatively stable accuracy, detects faults promptly and accurately, and exhibits minimal sensitivity to variations in sowing speed.
3. According to the results of the field test, natural light conditions do not affect the performance of the monitoring system, which operates effectively in the complex field environment. At operating speeds of 2 to 4 km/h, the relative error range for seeding amount monitoring is between 0.48% and 1.46%, with an average relative error of 0.91%. The alarm success rate is 100%. The monitoring system's error is within a reasonable range, making it suitable for monitoring actual seeding operations.

Future research will explore the use of dust-proof components or clean fiber optic sensors, such as the installation of dust-proof glass or the integration of a self-detection system to optimize sensor dust levels and mitigate the negative impact of dust on monitoring accuracy. Additionally, automation and multi-source information fusion technologies will be integrated to develop an intelligent seed replenishment system, enabling precise seeding even in cases of missed seeding, thereby better aligning the system with real-world field production conditions.

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 in the article/supplementary material.

Author contributions

BY: Writing – original draft. ZC: Writing – original draft. GD: Writing – review & editing. GL: Writing – review & editing. ShZ: Writing – review & editing. FH: Writing – review & editing. LL: Writing – review & editing. PC: Writing – original draft. XW: Writing – original draft. SiZ: Writing – review & editing. YD: Writing – review & editing. SQ: Writing – review & editing. ZL: Writing – original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Central Public-interest Scientific Institution Basal Research Fund (1630132022002),China Agriculture Research System (CARS-11-GDDGR), Guangxi Science and Technology Major Program (Guike AA23073003).

Acknowledgments

We would like to express our gratitude to our supervisor for their project funding support and guidance on this paper, as well as to our fellow researchers for their help and 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.

References

Abdolahzare, Z., and Abdanan Mehdizadeh, S. (2018). Real time laboratory and field monitoring of the effect of the operational parameters on seed falling speed and trajectory of pneumatic planter. *Comput. Electron. Agric.* 145, 187–198. doi: 10.1016/j.compag.2018.01.001

Albarenque, S., Basso, B., Davidson, O., Maestrini, B., and Melchiori, R. (2023). Plant emergence and maize (zea mays l.) Yield across multiple farmers' fields. *Field Crops Res.* 302, 109090. doi: 10.1016/j.fcr.2023.109090

Bai, J., Hao, F., Cheng, G., and Li, C. (2021). Machine vision-based supplemental seeding device for plug seedling of sweet corn. *Comput. Electron. Agric.* 188, 106345. doi: 10.1016/j.compag.2021.106345

Bai, S., Yuan, Y., Niu, K., Shi, Z., Zhou, L., Zhao, B., et al. (2022). Design and experiment of a sowing quality monitoring system of cotton precision hill-drop planters. *Agriculture* 12, 1117. doi: 10.3390/agriculture12081117

Basso, B., Fiorentino, C., Cammarano, D., and Schulthess, U. (2016). Variable rate nitrogen fertilizer response in wheat using remote sensing. *Precis. Agric.* 17, 168–182. doi: 10.1007/s11119-015-9414-9

Cay, A., Kocabiyik, H., and May, S. (2018). Development of an electro-mechanic control system for seed-metering unit of single seed corn planters part i: design and laboratory simulation. *Comput. Electron. Agric.* 144, 71–79. doi: 10.1016/j.compag.2017.11.035

Chen, J., Li, Y., Qin, C., and Liu, C. (2019). Design and test of capacitive detection system for wheat seeding quantity. *Trans. Chin. Soc. Agric. Eng.* 50, 77–83. doi: 10.11975/j.issn.1002-6819.2018.18.007

Chen, L., Xue, J., Ma, X., Xu, Y., and Mou, X. (2023). Design and test of double row whole rod feeding cassava precision planter. *J. Agric. Mechanization Res.* 45, 75–82. doi: 10.13427/j.cnki.njyi.2023.03.006

Chen, J., Zhang, H., Pan, F., Du, M., and Ji, C. (2022). Control system of a motor-driven precision no-tillage maize planter based on the canopen protocol. *Agriculture* 12, 932. doi: 10.3390/agriculture12070932

Chenna, S., Ivanov, M., Nielsen, T. K., Chalenko, K., Olsen, E., Jorgensen, K., et al. (2024). A data-driven genome annotation approach for cassava. *Plant J.* 119, 1596–1612. doi: 10.1111/tpj.16856

He, C., Zhou, B., Wang, H., Wei, Y., and Huang, J. (2023). A first-year maize/cassava relay intercropping system improves soil nutrients and changes the soil microbial community in the symbiotic period. *Front. Microbiol.* 14. doi: 10.3389/fmicb.2023.1087202

Jiang, M., Liu, C., Du, X., Huang, R., Dai, L., and Yuan, H. (2021). Research on continuous granular material flow detection method and sensor. *Measurement* 182, 109773. doi: 10.1016/j.measurement.2021.109773

Jin, X., Chen, Z., Zhao, B., Liu, M., Li, M., Li, Z., et al. (2024). Design and experiment of high-speed and precise positioning seeding control system for rice seedlings based on dual-position feedback adjustment (dpfa). *Comput. Electron. Agric.* 217, 108548. doi: 10.1016/j.compag.2023.108548

Karimi, H., Navid, H., Besharati, B., and Eskandari, I. (2019). Assessing an infrared-based seed drill monitoring system under field operating conditions. *Comput. Electron. Agric.* 162, 543–551. doi: 10.1016/j.compag.2019.04.045

Ling, L., Wu, G., Wen, C., Xiao, Y., Fu, W., Dong, J., et al. (2024). Influence of speed measurement method on performance of an electric-drive maize precision planter. *Biosyst. Eng.* 238, 175–187. doi: 10.1016/j.biosystemseng.2024.01.014

Lungkapin, J., Salokhe, V. M., Kalsirisilp, R., and Nakashima, H. (2009). Design and development of a cassava planter. *Trans. Asabe* 52, 393–399. doi: 10.13031/2013.26823

Rossi, S., Rubio Scola, I., Bourges, G., Šarauskis, E., and Karayel, D. (2023). Improving the seed detection accuracy of piezoelectric impact sensors for precision seeders. Part i: a comparative study of signal processing algorithms. *Comput. Electron. Agric.* 215, 108449. doi: 10.1016/j.compag.2023.108449

Sheat, S., Mushi, E., Gwandu, F., Sikirou, M., Baleke, P., Kayondo, S. I., et al. (2024). Cut, root, and grow: simplifying cassava propagation to scale. *Plants* 13. doi: 10.3390/plants13040471

Tang, H., Xu, C., Wang, Z., Wang, Q., and Wang, J. (2022). Optimized design, monitoring system development and experiment for a long-belt finger-clip precision corn seed metering device. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.814747

Wang, Z., Pei, J., He, J., Zhang, M., Yang, W., and Luo, X. (2020). Development of the sowing rate monitoring system for precision rice hill-drop drilling machine. *Trans. Chin. Soc. Agric. Eng.* 36, 9–16. doi: 10.11975/j.issn.1002-6819.2020.10.002

Wang, G., Sun, W., Zhang, H., Liu, X., Li, H., Yang, X., et al. (2020). Research on a kind of seeding-monitoring and compensating control system for potato planter without additional seed-metering channel. *Comput. Electron. Agric.* 177, 105681. doi: 10.1016/j.compag.2020.105681

Wang, F., Yang, L., and Wang, H. (2022). Design and test of electric driving pneumatic carrot planter in greenhouse. *Trans. Chin. Soc. Agric. Machinery* 53, 64–73. doi: 10.6041/j.issn.1000-1298.2022.08.007

Wang, J., Zhang, Z., Wang, F., Jiang, Y., and Zhou, W. (2019). Design and experiment of monitoring system for rice hill-direct-seeding based on piezoelectric impact method. *Trans. Chin. Soc. Agric. Machinery* 50, 74–84. doi: 10.6041/j.issn.1000-1298.2019.06.008

Xie, C., Yang, L., Zhang, D., Cui, T., He, X., Du, Z., et al. (2022). A signal output quantity (soq) judgment algorithm for improving seeding quantity accuracy. *Comput. Electron. Agric.* 201, 107321. doi: 10.1016/j.compag.2022.107321

Xie, C., Yang, L., Zhang, D., Cui, T., and Zhang, K. (2021a). Seeding parameter monitoring method based on laser sensors. *Trans. Chin. Soc. Agric. Eng.* 37, 140–146. doi: 10.11975/j.issn.1002-6819.2021.03.017

Xie, C., Zhang, D., Yang, L., Cui, T., He, X., and Du, Z. (2021b). Precision seeding parameter monitoring system based on laser sensor and wireless serial port communication. *Comput. Electron. Agric.* 190, 106429. doi: 10.1016/j.compag.2021.106429

Xie, C., Zhang, D., Yang, L., Cui, T., He, X., Du, Z., et al. (2023). Research and analysis on the influence of different speed measurement methods on the monitoring accuracy of seed spacing. *Agriculture-Basel* 13, 128. doi: 10.3390/agriculture13010128

Xie, C., Zhang, D., Yang, L., Cui, T., Yu, T., Wang, D., et al. (2021c). Experimental analysis on the variation law of sensor monitoring accuracy under different seeding speed and seeding spacing. *Comput. Electron. Agric.* 189, 106369. doi: 10.1016/j.compag.2021.106369

Xing, H., Wang, Z., Luo, X., He, S., and Zang, Y. (2020). Mechanism modeling and experimental analysis of seed throwing with rice pneumatic seed metering device with adjustable seeding rate. *Comput. Electron. Agric.* 178, 105697. doi: 10.1016/j.compag.2020.105697

Xu, L., Hu, B., Li, J., Ren, L., Guo, M., Mao, Z., et al. (2023). An efficient seeding state monitoring system of a pneumatic dibbler based on an interdigital capacitive sensor. *Comput. Electron. Agric.* 209, 107856. doi: 10.1016/j.compag.2023.107856

Zhang, X., Zhang, H., Shi, Z., Jin, W., Chen, Y., and Yu, Y. (2022). Design and experiments of seed pickup status monitoring system for cotton precision dibblers. *Trans. Chin. Soc. Agric. Eng.* 38, 9–19. doi: 10.11975/j.issn.1002-6819.2022.05.002

Zhao, L., Han, Z., Yang, J., and Qi, H. (2018). Single seed precise sowing of maize using computer simulation. *PloS One* 13, e193750. doi: 10.1371/journal.pone.0193750

Zhao, J., Wang, X., Wang, J., and Han, Z. (2023). A high-precision strain seeding spacing monitoring system based on a combined bionic strain sensor and strain peak recognition algorithm. *Comput. Electron. Agric.* 212, 108061. doi: 10.1016/j.compag.2023.108061

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 24 December 2024
DOI 10.3389/fpls.2024.1509746

Check for updates

OPEN ACCESS

EDITED BY
Lei Shu,
Nanjing Agricultural University, China

REVIEWED BY
Wei Lu,
Nanjing Agricultural University, China
Elio Romano,
Centro di ricerca per l'Ingegneria e le Trasformazioni agroalimentari (CREA-IT), Italy

*CORRESPONDENCE
Jianjun Yin
yinjianjun@ujs.edu.cn

RECEIVED 11 October 2024
ACCEPTED 05 December 2024
PUBLISHED 24 December 2024

CITATION
Chen Z, Fang Y, Yin J, Lv S, Sheikh Muhammad F and Liu L (2024) A novel lightweight YOLOv8-PSS model for obstacle detection on the path of unmanned agricultural vehicles.
Front. Plant Sci. 15:1509746.
doi: 10.3389/fpls.2024.1509746

COPYRIGHT
© 2024 Chen, Fang, Yin, Lv, Sheikh Muhammad 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.

A novel lightweight YOLOv8-PSS model for obstacle detection on the path of unmanned agricultural vehicles

Zhijian Chen[1], Yijun Fang[1], Jianjun Yin[1]*, Shiyu Lv[1], Farhan Sheikh Muhammad[1] and Lu Liu[2]

[1]School of Agricultural Engineering, Jiangsu University, Zhenjiang, China, [2]Institute of Technology, Anhui Agricultural University, Hefei, China

Introduction: The rapid urbanization of rural regions, along with an aging population, has resulted in a substantial manpower scarcity for agricultural output, necessitating the urgent development of highly intelligent and accurate agricultural equipment technologies.

Methods: This research introduces YOLOv8-PSS, an enhanced lightweight obstacle detection model, to increase the effectiveness and safety of unmanned agricultural robots in intricate field situations. This YOLOv8-based model incorporates a depth camera to precisely identify and locate impediments in the way of autonomous agricultural equipment. Firstly, this work integrates partial convolution (PConv) into the C2f module of the backbone network to improve inference performance and minimize computing load. PConv significantly reduces processing load during convolution operations, enhancing the model's real-time detection performance. Second, a Slim-neck lightweight neck network is introduced, replacing the original neck network's conventional convolution with GSConv, to further improve detection efficiency and accuracy. This adjustment preserves accuracy while reducing the complexity of the model. After optimization, the bounding box loss function is finally upgraded to Shape-IoU (Shape Intersection over Union), which improves both model accuracy and generalization.

Results: The experimental results demonstrate that the improved YOLOv8_PSS model achieves a precision of 85.3%, a recall of 88.4%, and an average accuracy of 90.6%. Compared to the original base network, it reduces the number of parameters by 55.8%, decreases the model size by 59.5%, and lowers computational cost by 51.2%. When compared with other algorithms, such as Faster RCNN, SSD, YOLOv3-tiny, and YOLOv5, the improved model strikes an optimal balance between parameter count, computational efficiency, detection speed, and accuracy, yielding superior results. In positioning accuracy tests, the, average and maximum errors in the measured distances between the camera and typical obstacles (within a range of 2-15 meters) were 2.73% and 4.44%, respectively.

Discussion: The model performed effectively under real-world conditions, providing robust technical support for future research on autonomous obstacle avoidance in unmanned agricultural machinery.

KEYWORDS

UAV, YOLOv8, depth camera, obstacle detection, PConv, slim-neck, shape-IoU

1 Introduction

The fast rate of urbanization, along with an aging rural population and migration of agricultural workers, has resulted in a considerable labor shortage in rural regions. These issues make it increasingly difficult to sustain high-quality farming techniques while failing to fulfill the expectations of modern agricultural growth. As a result, the automation of agricultural technology has become an absolute requirement for agricultural advancement. At present, such difficulties can be effectively solved through the development of smart agriculture (Garg and Alam, 2023) and unmanned farms are an important way to realize smart agriculture (Ming et al., 2023). Many developed countries in Europe and the United States have streamlined operations using unmanned farm technologies such as automated navigation for smart farm machinery (Yang et al., 2022; Ji et al., 2023). Compared to conventional agricultural production techniques, automation can increase operating hours, reduce operator labor intensity, and enhance overall efficiency (Wang et al., 2022). However, the unstructured features of agricultural landscapes create considerable obstacles. The random and discrete distribution of obstacles, such as pedestrians in the field or other agricultural machinery operating simultaneously, poses serious risks to the autonomous navigation of agricultural equipment. Accurate obstacle detection during operation is essential for effective path planning and obstacle avoidance (Zhao et al., 2023). It is significant to ensure the operational efficiency and driving safety of agricultural machinery.

Obstacle detection can be achieved by employing various obstacle detection sensors, including vision sensors (Fu et al., 2020)、laser radar (Yang et al., 2022)、and ultrasonic sensors (Li et al., 2018). Among these methods, vision-based detection methods have the advantages of low price, rich information, and wide detection range, which are widely used in theoretical research and practical application exploration. Traditional target detection algorithms rely on feature extraction and sliding window techniques to identify specific objects in an image and determine their location (Dong et al., 2023). However, in complex and dynamic environments manually extracted features may not effectively describe and distinguish, objects, affecting the accuracy and robustness of detection. With the rise of deep learning technology, target detection algorithms based on this technology have gradually replaced traditional algorithms, becoming mainstream, emerging and widely used in the agricultural field in recent years (Ren et al., 2016; Ji et al., 2022). These algorithms are mainly categorized into two-stage target detection algorithms represented by Faster R-CNN and one-stage target detection algorithms SSD and YOLO detection algorithms, which are now widely used in target detection (Wang et al., 2021). To address obstacle detection in complex orchard environments, Liu et al (Hui et al., 2019). proposed a real-time pedestrian identification approach based on an enhanced SSD architecture that utilizes MobileNetV2 as the backbone. To anticipate position, the auxiliary network layer uses an inverted residual structure and null convolution. When tested on an open dataset, the upgraded SSD model obtained an average accuracy of 97.46% and a recall of 89.72% in pedestrian recognition, despite only identifying pedestrians as a single obstacle type. Chen et al (Bin et al., 2021). created an enhanced YOLOv3-tiny target detection model to facilitate in obstacle avoidance during autonomous agricultural machinery navigation. This model incorporates shallow characteristics via a splicing layer before the second prediction layer, as well as a residual module, to improve the depth and learning capacity of the YOLOv3-tiny backbone. This advancement makes it possible to use a panoramic camera installed on the agricultural machine to detect impediments in real time with greater accuracy. The results show that the average accuracy and recall rates increased by 5.6% and 5.2%, respectively, satisfying the requirement for real-time obstacle detection while moving. Wei et al (Jiansheng et al., 2021). developed an obstacle sensing system for autonomous tractors based on an upgraded YOLOv3 model for field detection and binocular vision for obstacle localization. The experimental findings demonstrated an average accuracy of 89.54%, a recall of 90.18%, a depth estimate error of 4.66% for dynamic situations, and an average processing time of 0.573 seconds. However, the large number of YOLOv3 model parameters reduces its speed on embedded systems, emphasizing the need for a more lightweight model to increase inference speed. Su et al (Su et al., 2022). enhanced the YOLOv5s model for real-time detection of orchard obstacles, such as people, by using K-means clustering, the Senet module, and pruning. Experimental findings indicated that the revised model decreased size by 13.6 MB while boosting accuracy and average accuracy by 5.60% and 1.30%, respectively. The average detection time reaches 33 ms, which meets the requirements for obstacle detection in orchards.

The above research has promoted the development of obstacle detection technology, and a certain degree of progress has been made in terms of detection accuracy, detection speed, and lightweight. However, the unstructured field environment also brings huge challenges. The obstacles in the field are not single types of objects, usually composed of static and dynamic types of obstacles. This often leads to the model ignoring some objects during detection, which poses certain safety hazards; In addition, most existing obstacle detection models are complex large models that cannot run when deployed on low performance embedded devices on agricultural machinery. Considering the safety and work efficiency of unmanned agricultural machines in the process of travelling, it is particularly important to achieve fast and accurate detection and localization of different types of obstacles. To address the problem that the complexity of the model leads to difficulties in mobile deployment, from the perspective of balancing the detection speed and detection accuracy, this paper proposes an improved network model based on YOLOv8, which achieves obstacle detection and localization after the model is deployed in the edge computer, to better leave a safe distance for the unmanned agricultural vehicle to avoid obstacles. The main contributions of our paper are summarized as follows:

- This article constructs an obstacle dataset suitable for unmanned agricultural vehicle driving scenarios by collecting images of different types of obstacles.
- In order to reduce the model size and maintain high detection speed while ensuring accuracy, this paper proposes a lightweight YOLOv8-PSS model.
- This article deploys the model on agricultural machinery and detects and locates different types of obstacles, verifying that the proposed YOLOv8-PSS model can maintain good detection performance after deployment, laying the foundation of the subsequent unmanned agricultural vehicle intelligentsia.

2 Materials and methods

2.1 Dataset creation

2.1.1 Data source

To increase dataset variety and assure robust detection and generalization during real-world operations, this research combines experimental data from both open-source and self-constructed datasets. The open-source data is sourced from Carnegie Mellon University's National Robotics Engineering Center (https://www.nrec.ri.cmu.edu/solutions/agriculture/human-detection-and-tracking-2/), Flying Plasma AI Studio's public dataset, and images collected from the web. The self-constructed field obstacle dataset was collected from Rungo Farm Co. in Zhenjiang City, Jingkou District, Jiangsu Province, China (Longitude 119°43'55.664″ East, Latitude 32°8'12.998″ North) and Green Spring Leisure Agricultural Demonstration Park in Huimin County, Binzhou City, Shandong Province, China (Longitude 117°32'5.136″ East, Latitude 37° 27'32.342″ North), between June 2023 and early November 2023. Images were captured from 8:00 to 18:00, and the acquisition equipment was an iPhone XS and Canon camcorder with lens resolutions of 3624 × 2448 and 3456 × 2304 pixels, respectively (https://www.kaggle.com/datasets/chenzhijian37/agricultural-obstacles). All data images were collected in the natural environments along farm roads, and unclear images caused by weather or human error were removed. An example of the dataset is shown in Figure 1, which mainly includes common obstacles such as people, farm machinery, trees, vehicles, and utility poles.

2.1.2 Image pre-processing

To further enrich the dataset, this research uses online data enhancement techniques to augment the data volume, specifically including luminance degradation adjustment, random colors, random noise, etc. Considering that the obstacles are normal during the agricultural machine operation, this research does not add other data enhancement methods, such as rotation and flip. LabelImg software is used to label the obstacles in the image with the minimum outer rectangular box, and the format of the LabelImg is a Txt file in YOLO format (Ji et al., 2021; Hu et al., 2023). The dataset was randomly divided into a training set (7344 images), validation set (909 images), and test set (908 images) in a 7:2:1 ratio, as detailed in Table 1.

2.2 YOLOv8

The YOLOv8 network, launched by Ultralytics in January 2023, introduces several improvements over the widely used YOLOv5. It consists of four main components: input, backbone, neck, and head (Yang et al., 2023). YOLOv8 officially provides a series of models in various sizes (Zhang et al., 2024), including YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x, which differ in network width (number of channels) and depth (number of layers), aiming to satisfy the speed and accuracy requirements in different scenarios. In this research, the smallest YOLOv8n is selected as the baseline model for target detection due to its speed, stability, and efficiency. It also offers scalability, making it well-suited for real-time obstacle detection in various target detection scenarios.

2.3 Obstacle recognition based on lightweight YOLOv8-PSS model

The key improvements of the model presented in this paper include: the incorporation of PConv in the backbone network, which reduces redundant computations and memory accesses while efficiently extracting feature information; the introduction of a Slim-neck in the neck network, replacing the original convolution module with GSConv, and designing a cross-level local network, VoVGSCSP, based on it. This reduces computational complexity and network structure while maintaining sufficient accuracy. Additionally, the original YOLOv8's CIoU loss function is replaced with the Shape-IoU loss function, enhancing both detection accuracy and convergence speed. The structure of the improved detection model is shown in Figure 2, with the marked areas indicating the improvements.

FIGURE 1
Sample of obstacles ((**A**) person, (**B**) tree, (**C**) pole, (**D**) agm, (**E**) vehicle).

2.3.1 PConv

The C2F module in YOLOv8 was originally designed to extract effective features. However, in certain cases, feature redundancy may still occur, leading to wasted model parameters and computational resources, ultimately affecting the model's generalization ability. While many lightweight network improvements focus on reducing computation, this often increases the number of memory accesses required during image processing, thereby reducing computational efficiency. This trade-off can not only slow down the model's overall speed but also degrade its accuracy to some extent. Therefore, this research introduces the partial convolution (PConv) proposed in FasterNet, PConv (Chen et al., 2023) compared to conventional convolutional operations. In partial convolution, regular convolution is applied to only a subset of the input feature map's channel, while the remaining channels remain unchanged. This enables PConv to more efficiently select the first or last consecutive channels as representatives of the entire feature map for computation when performing consecutive memory accesses. By doing so, it optimizes the use of the device's computational power, reduces the amount of model computation and memory consumption, avoids performance degradation caused by inter-channel redundancy, and improves the model's generalization ability. During calculations, simplifying the PConv calculation as $h \times w \times k^2 \times c_p^2$, reduces the need for memory accesses with the following expression for memory accesses:

$$h \times w \times k^2 \times 2c_p + k^2 \times c_p^2 \approx h \times w \times 2c_p \quad (1)$$

where *h and w* denote the length and width of the input feature map, c denotes the number of input channels, c_p denotes the channels involved in convolution, and k is the convolution kernel size.

In this paper, PConv is used to streamline and improve the C2f module, maintaining the extended feature extraction capabilities of C2f. A Faster Block is introduced to replace the Bottleneck in C2f, further reducing the computational load during the calculation process. The structure of PConv and the modified module is shown in Figure 3.

TABLE 1 Classification of different obstacle samples.

Type	Train	Validation	Test
Person	1679	190	187
Tree	1154	168	173
Pole	1226	177	175
Agm	1536	186	181
Vehicle	1749	188	192

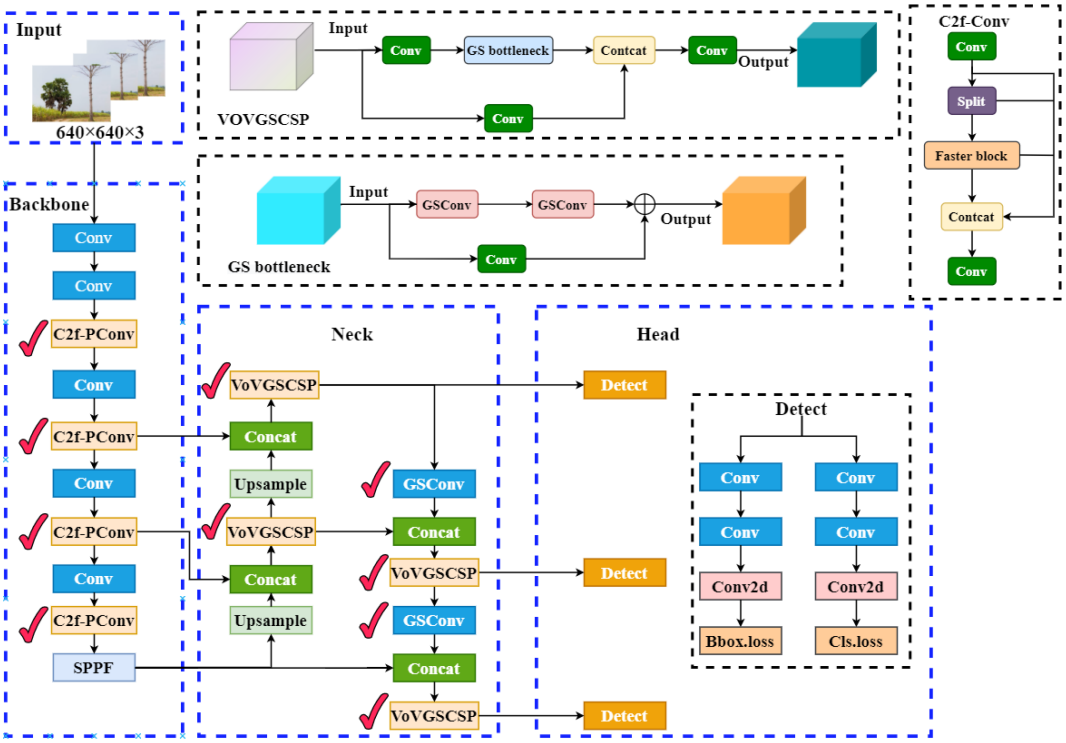


FIGURE 2
YOLOv8-PSS model structure diagram.

2.3.2 Slim neck

In traditional CNNs, spatial information is gradually transformed into channel information, leading to a partial loss of semantic data during spatial compression and channel expansion at each feature map layer. When using standard convolution (SC) for feature extraction, the number of parameters grows with the increasing depth of the network, impacting detection efficiency. Depth-wise separable convolution (DSC) has become popular in many lightweight designs to reduce computation and improve efficiency. DSC minimizes parameters by performing stepwise convolution, which processes channel information separately. However, this approach captures complex spatial information and feature relationships less effectively than SC. Therefore, a balance must be struck between the computational efficiency of DSC and the performance of SC. Li et al (Li et al., 2024). proposed GSConv, a novel approach designed to maintain hidden connections between channels as much as possible while keeping the time complexity low. The core idea of GSConv is to enhance the fusion of feature information between different groups by introducing a shuffle operation that mixes the information generated by the SC operation into each part of the information generated by the DSC. The shuffle operation improves the performance of the model by mixing information from different groups by exchanging local feature information uniformly over different channels. Therefore, in this paper, GSConv is adopted as a lightweight convolution method, replacing both SC and DSC convolutions and integrating the Slim-neck lightweight neck network. This ensures a balance between the detection accuracy and computational efficiency of the network. The structure of GSConv is shown in Figure 4. The Slim-Neck module introduces a GS bottleneck and VoVGSCSP lightweight module based on GSConv, to solve the problem of increased computation caused by many C2F modules extracting features in the YOLOv8 neck network as well as solving the problem of difficulty in real-time detection of the deployed devices at runtime.

To enhance the inference speed of the network model while preserving its detection accuracy, the cross-level partial network module VoV-GSCSP was designed to replace the original C3 module, utilizing the GSConv module. This design cleverly integrates GSConv with VoVNet, reducing the number of parameters and computational complexity. The VoV-GSCSP module ensures high computational efficiency while improving the overall network performance and maintaining sufficient accuracy. Figure 5A illustrates the structure of the GS bottleneck, and Figure 5B shows the structure of the VoV-GSCSP.

2.3.3 Shape-IoU

The loss function serves to quantify the difference or error between the model's predicted labels and the actual labels. By calculating the loss value, it is clear how well the model performs on a given dataset, i.e., the accuracy of the model prediction. The CIoU loss function used in YOLOv8 only considers the overlapping area of the two bounding boxes but does not consider the target's size, shape, and positional feature information. In some cases, it may not accurately reflect the actual situation, leading to a decrease in detection accuracy.

In contrast to CIoU, the Shape-IoU loss function calculates the loss by focusing on the shape and scale of the bounding box, resulting in more precise border regression and enhanced detection accuracy and

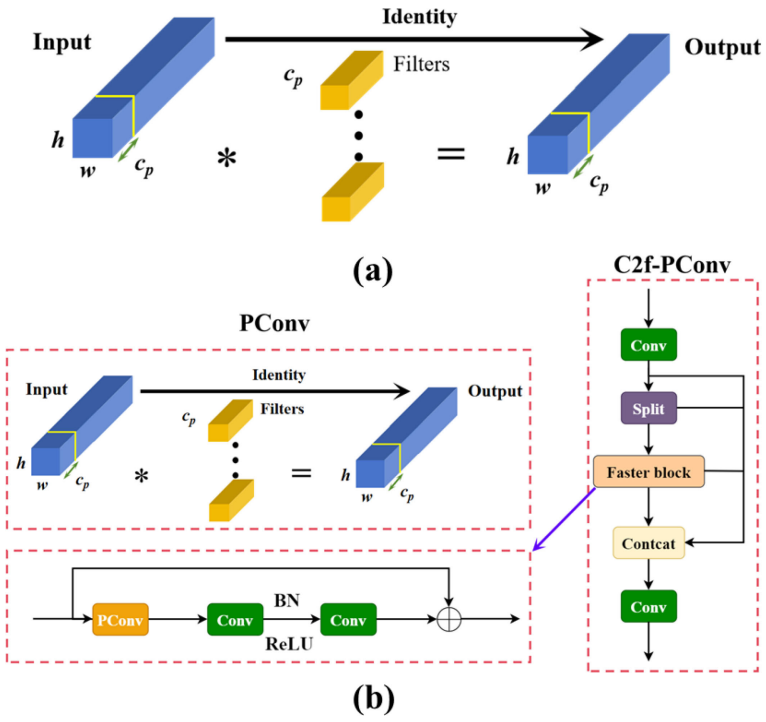


FIGURE 3
(A) Processing of the output feature maps from the regular convolution module; **(B)** Diagram of the structure of the C2f-PConv module; * denotes convolution processing.

robustness (Zhang and Zhang, 2023), as shown in Figure 6. Therefore, in this paper, Shape-IoU is used to replace CIoU in the original model. This modification not only improves the network's convergence performance but also enhances its ability to detect obstacles more accurately. The expression for Shape-IoU is shown below.

$$L_{Shape-IoU} = 1 - IoU + disance^{shape} + 0.5 \times \Omega^{shape} \tag{2}$$

$$distance^{shape} = hh \times \frac{(x_c - x_c^{gt})^2}{c^2} + ww \times \frac{(y_c - y_c^{gt})^2}{c^2} \tag{3}$$

$$\Omega^{shape} = \sum_{t=w,h} (1 - e^{-\omega_t})^{\theta} \tag{4}$$

$$ww = \frac{2 \times (w^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}} \tag{5}$$

$$hh = \frac{2 \times (h^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}} \tag{6}$$

$$\Omega^{shape} = \sum_{t=w,h} (1 - e^{-\omega_t})^{\theta} \tag{7}$$

$$\begin{cases} \omega_w = hh \times \frac{|w - w^{gt}|}{max(w, w^{gt})} \\ \omega_h = ww \times \frac{|h - h^{gt}|}{max(h, h^{gt})} \end{cases} \tag{8}$$

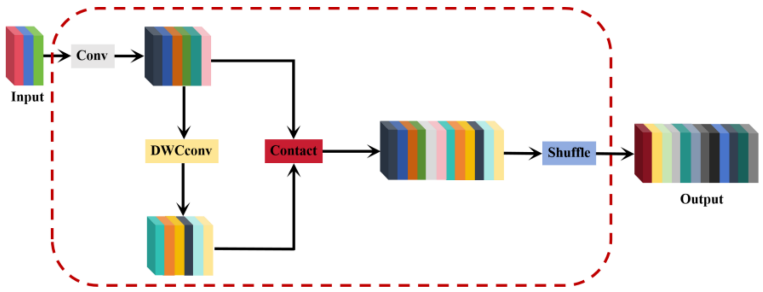


FIGURE 4
GSConv structure diagram.

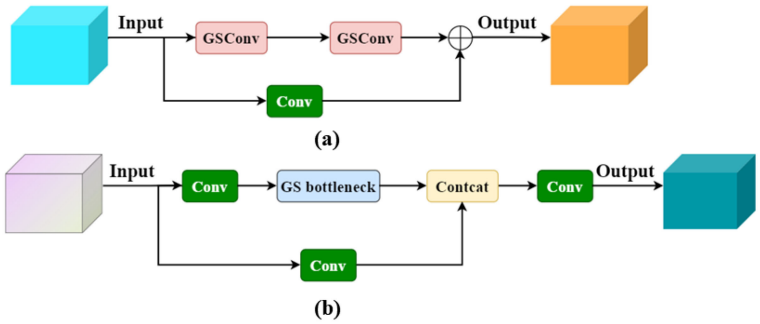


FIGURE 5
(A) Structure of GS bottleneck; **(B)** Structure of VoV-GSCSP.

Where Ω^{shape} is the shape loss; w^{gt}, h^{gt} denotes the width and height of the real bounding box, respectively; θ is the degree of control attention to shape loss and takes the value of 4.

2.4 Visualization of heat maps

Gradient-weighted Class Activation Heatmap (Grad-CAM) can be used to interpret the results predicted by the model by visualizing the heatmap that the model considers most significant. In obstacle detection, Grad-CAM generates heatmaps for the key regions making it evident where the model focuses its attention on the obstacle's critical information. The basic idea of Grad-CAM (Liu et al., 2024) is to use the gradient of the neural network in the output layer for a specific class to determine which feature maps are the most important for predicting a specific class by calculating the gradient weight of the feature map of the last convectional layer. To showcase the improved detection performance of the model, Grad-CAM is applied to visualize the feature extraction at various layers, with red representing higher weights, where the model pays more attention to the region, and blue represents lower weights, which means that the model pays less attention to the region. To observe the heat maps more clearly, the heat maps are superimposed on top of the original maps. Thus, the different features and the areas on which the model focuses can be seen clearly. As shown in Figure 7,

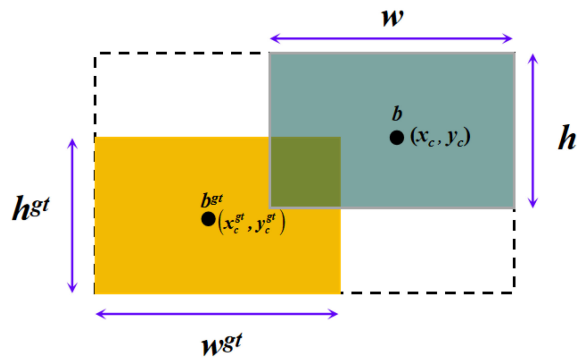


FIGURE 6
Schematic diagram of the intersection of the true and predicted boxes.

the embodiment of the improved model's focus on features at different stages can be seen, and the experimental results show that the algorithm can better solve the obstacle recognition problem in different scenarios.

2.5 Experimental platforms

In this paper, we use our own paired workstation for the training and validation of the deep learning part, the hardware configuration includes: In this study, the training and validation of the deep learning model were conducted on a custom-built workstation. CPU is Intel Core i9 -13900kf, running memory is 32GB, GPU is NVIDIA RTX 4090 graphic card, 1T solid state hard drive; the software is running on the 64-bit operating system of Windows 10 (22H2)All programs are written in python language using the Pytorch1.12 deep learning framework with training acceleration powered by NVIDIA CUDA 11.6 parallel computing driver. Deep learning framework PyTorch 2.1.2, programming platform PyCharm, programming language Python 3.8, CUDA version 11.3, and CUDNN version 7.6 were utilized. These optimizations contribute to improved model accuracy and generalization after applying lightweight techniques.

The training parameters were all adopted from the default hyper-parameters of The input image resolution was set to 640×640 pixels with a batch size of 32, and pre-trained weights were utilized. Stochastic Gradient Descent (SGD) was employed to optimize the network and accelerate the convergence process. The learning rate was set at 0.01, with weight decay coefficients of 0.0005, and a momentum factor of 0.937 A total of 300 rounds of iterations were conducted followed by an optimal results analysis.

2.6 Evaluation metrics

This research utilizes standard performance evaluation metrics commonly used in target detection, including precision (P), recall (R), mean average precision (mAP), the number of parameters, floating point operations (FLOPs), model size, and frames per second (FPS) (Zhang et al., 2023). mAP is particularly important for assessing the overall detection performance of the target detection model; a higher mAP value indicates better obstacle detection capability. To evaluate

FIGURE 7
Visualization results of heat maps under different network modules (**(A)** Original image, **(B)** PConv, **(C)** Slim-neck).

model complexity, three main metrics are considered: the number of parameters, FLOPs, and the model size. P and R are calculated using true positive (TP), false positive (FP), true negative (TN) and false negative (FN). mAP is particularly important for assessing the overall detection performance of the target detection model; a higher mAP value indicates better obstacle detection capability. To evaluate model complexity, three main metrics are considered: the number of parameters, FLOPs, and the model size. FLOPs represent the speed of floating-point operations, which can be used to measure the complexity of a model. Parameters represent the computational memory resources consumed by the model. The formulas are as follows:

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$AP = \int_0^1 P(R)dR \tag{11}$$

$$mAP = \frac{\sum_{n=1}^{N} AP_n}{N} \tag{12}$$

$$Parameters = r^2 \times a \times v + v \tag{13}$$

$$FLOP_S = 2 \times H \times W \times C_{out} \times (C_{in} \times K^2 + 1) \tag{14}$$

where N is the number of categories. where a is the input size, r is the size of the convolution kernel, v is the output size, $H \times W$ is the size of the output feature map, C_{in} is the input channel, K is the kernel size, and C_{out} is the output channel.

2.7 Detection and positioning system design

In this study, the OAK-D-Pro camera from Luxonis is selected for the study, featuring two infrared cameras, one RGB camera, and an infrared laser dot-matrix transmitter. The depth measurement principle mainly relies on the parallax generated by the infrared cameras in the imaging plane to calculate the distance to the object. Firstly, the RGB and depth images of the front view are captured by the two IR cameras and the RGB camera. The improved YOLOv8 model is then applied to detect obstacles within the RGB image. The parameters of the OAK-D-Pro camera are shown in Table 2.

To obtain the precise position point of the obstacle in real space, it is necessary to go through the conversion from the pixel coordinate system, and image coordinate system to the camera coordinate system, and finally transform the 2-D image coordinates into 3-D spatial coordinates. In the obstacle positioning system, the camera is used as the origin of the coordinate system to locate the obstacle, and after identifying the pixel coordinates of the obstacle,

TABLE 2 OAK-D-Pro Parameters.

Parameters of OAK-D-Pro	RGB camera	Infrared camera
Resolution	4032×3040	1280× 800
Baseline length	/	75 mm
Camera focal length	4.81	2.35
FOV	81°D/69°H/55°V	81°D/72°H/49°V
Size	97 × 29.5 × 22.9 mm	
Power	2 ~ 5.5 W	

the 3-D coordinates of the obstacle in the camera coordinate system combined with the depth information are used to compute the absolute coordinates in the world coordinate system after the final coordinate transformation, to realize the detection and positioning.

To ensure the accuracy of camera positioning, OAK-D-Pro needs to be calibrated by displaying a checkerboard calibration image on a 24 inch flat display, capturing a total of 13 multi border positions for calibration. After capturing images of all multi border positions, the calibration image processing step will begin. Upon successful completion, a green background interface will pop up at the end. The checkerboard positions at different calibration angles and the interface after successful calibration are shown in Figure 8.

After calibration is completed, generate the cam_chain.yaml file, which contains the internal and external parameters of the camera after calibration, as shown in Table 3.

To simulate real-world conditions, the OAK-D-Pro camera was mounted on a tracked rice and wheat harvester capturing the video information in the actual field environment. The improved YOLOv8 model was then deployed on the NVIDIA Jetson TX2 to detect and identify the obstacle targets, which still rely on the manually driving operation in the process. The overall framework of the system is shown in Figure 9.

3 Discussion and analysis

3.1 Ablation experiment

The ablation test results, using the same obstacle dataset and training environment, are presented in Table 4. The model incorporating the PConv module is referred to as YOLOv8_P, and the model utilizing the Slim-neck module is designated as YOLOv8_S, with other models named similarly based on their improvements. As observed in Table 4, replacing the C2f module in YOLOv8 with the C2f-PConv in Test 1 resulted in the YOLOv8_P model. While there was a slight decrease in precision, recall, and average precision compared to the original YOLOv8, the model's weight was reduced by 36.2%. Additionally, the number of parameters and computational load were reduced to 67.3% and 79.3%, respectively, of the original model. The main reason for the decrease in the precision rate and other values is that only some channels are involved in the convolution operation of PConv, which has a certain feature extraction effect, but at the same time reduced the amount of computation and size of the model, it also loses some of the feature information in the remaining channels, leading to a reduction in the precision of the training. It is worth noting that the results of experiment 2 show that although there is a slight decrease

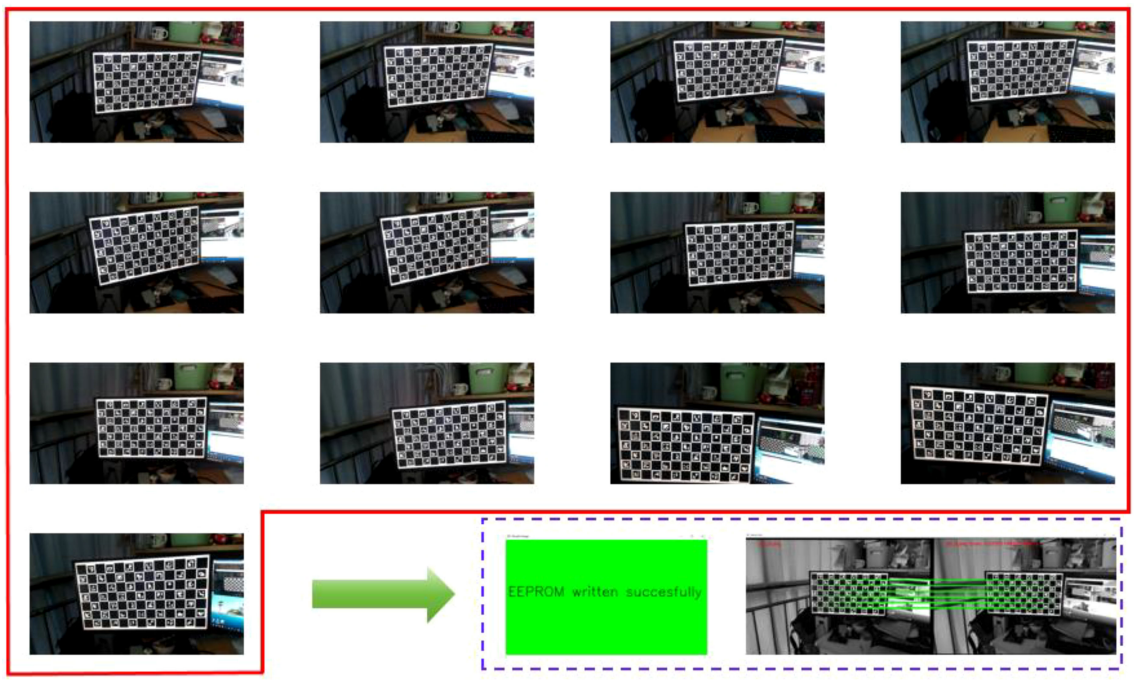


FIGURE 8
Camera calibration diagram.

TABLE 3 Calibration parameters.

Calibration parameters	center camera parameters	Right camera parameters
Radial distortion parameter/k_1	0.0281	0.0129
Radial distortion parameter/k_2	-0.0682	-0.0458
Tangential distortion coefficient/p_1	-0.0013	-0.0014
Tangential distortion coefficient/p_2	0.0024	0.0047
U-axis scale factor/fx	806.2	797.1
V-axis scale factor/fy	665.4	642.6
Rotation matrix/R	$\begin{bmatrix} 0.9999 & 0.0027 & 0.014 \\ -0.0023 & 0.9999 & 0.0031 \\ -0.0056 & -0.0029 & 0.9999 \end{bmatrix}$	
Offset matrix/T	$\begin{bmatrix} 3.7351 \\ -0.1476 \\ 0.0041 \end{bmatrix}$	

in the model detection performance after adding PConv, the magnitude of the decrease is within an acceptable range. By sacrificing a small part of the accuracy, the computational complexity and complexity of the model are greatly reduced, which is more conducive to the deployment of the model on the platform in the future. Experiment 3 based on experiment 1, the neck network was redesigned using the Slim-neck network. At that time, the precision, recall, and average accuracy of the model were 84.3%, 88.1%, and 89.5%, compared with the baseline model, the precision and detection speed of the redesigned model were slightly improved. The number of participants, the amount of computation, and the size of the model were 51.9% of that of the original model, respectively, 55.3%, and 48.3%, indicating that the Slim-neck lightweight network used in this research can meet the demand of real-time detection in terms of model detection accuracy. The main reason for the improved performance is the replacement of traditional convolution and the C3 module in the original model with GSConv and VoV-GSCSP modules. This modification allows the structure of the Slim-neck network to preserve the important channel connections in the neck structure. The GSConv module can effectively process channel feature information to avoid redundant information from being compressed repeatedly. As a result, the model's feature extraction capability and detection speed are enhanced without sacrificing accuracy. Experiment 4 is based on experiment 2, and the original model neck network and backbone network are simultaneously improved by light-weighting, which greatly reduces the complexity and computation amount of the model while maintaining a slight decrease in model accuracy. The size of the improved model is 4.7 MB, the number of parameters is 6.0 G, the number of model parameters is 2.3 × 106 M, the precision, recall, and average accuracy are 83.9%, 85.8% and 89.4%, the mAP decreases by 0.5%, and the detection speed is 113 frames/s. The number of parameters, the amount of computation, and the size of the improved model are about half of the previous ones, and the detection speed is increased by 14%. Experiment 5 based on Experiment 4; the original loss function is further replaced by Shape-IoU. The model has improved precision, recall, and average precision compared with the previous one. The complexity of the model remains unchanged, which indicates that Shape-IoU can improve the fitting effect of the model and further improve the detection accuracy of the model.

The detection performance across different obstacle types in complex scenes is a key factor affecting the accuracy of the detection task. The detection results of the improved model for different obstacle types are shown in Table 5. From Table 5, the model's detection effect for dynamic obstacles is better than that for static ones, the model has the best detection effect for pedestrians, with P, R, and AP of 88.5%, 90.2%, and 92.3%, respectively, and the model has the worst detection effect for utility poles, with P, R and AP of 81.3%, 86.7% and 88.4%, respectively. There are some differences in the model's handling of different obstacles, and the known static obstacles detected during travelling can be labelled in subsequent obstacle avoidance studies to further enhance the safety of travelling.

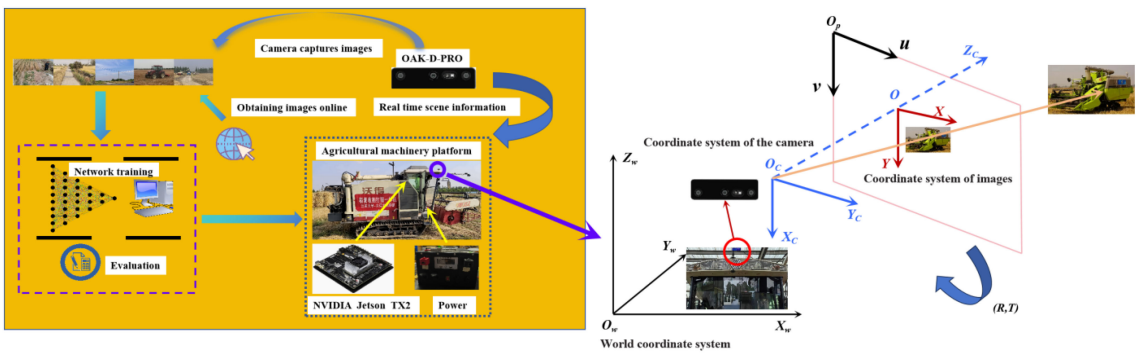


FIGURE 9
Overall framework diagram.

TABLE 4 Improved model ablation test results.

Number	Model	PConv	Slim-neck	Shape-IoU	P %	R/ %	mAP/ %	Size/ MB	Parameters/ $\times10^6$M	Computation/ G	FPS
1	YOLOv8	×	×	×	83.5	86.7	88.9	11.6	5.2	12.3	92
2	YOLOv8_P	√	×	×	83.2	86.4	88.3	7.4	3.5	7.5	107
3	YOLOv8_S	×	√	×	84.3	88.1	89.5	5.6	2.7	6.8	101
4	YOLOv8_PS	√	√	×	83.9	85.8	89.4	4.7	2.3	6.0	113
5	YOLOv8_PSS	√	√	√	85.3	88.4	90.6	4.7	2.3	6.0	111

√ represents the use of the module, and x represents that the module has not been used.

3.2 Comparison of results from different detection models

To further verify the actual performance of the improved model proposed in this paper, this paper compares the improved model with the mainstream two-stage target detection algorithm Faster R-CNN and the one-stage target detection algorithms SSD, YOLOv3-tiny and YOLOv5 models under the same conditions of test parameters and configuration environments, and the comparison results are shown in Table 6.

As demonstrated in Table 6, the enhanced YOLOv8 model outperforms Faster R-CNN, SSD, YOLOv3-tiny, and YOLOv5 by 11%, 0.7%, 2.5%, and 4.6%, respectively. Furthermore, the mean Average Precision (mAP) increases by 46.9%, 10.5%, 11.1%, and 6.7% when compared to these models. The improved YOLOv8 model also exhibits a significant reduction in model size, number of parameters, and computational load. Although the recall rate of Faster R-CNN is higher than that of YOLOv8-PSS during obstacle detection, the other performance metrics of YOLOv8-PSS are superior. Despite being a two-stage target detection model, Faster R-CNN's substantial parameter count renders it unsuitable for lightweight deployment requirements. The SSD algorithm demonstrates higher precision and accuracy compared to Faster R-CNN. Furthermore, YOLOv3-tiny and YOLOv5 have a greater number of parameters and higher floating-point computation requirements than the proposed improved method, and they also exhibit lower average accuracy, which is not conducive to the lightweight design of the network. The model in this research is improved based on YOLOV8, compared with other mainstream models, the improved algorithm proposed in this research achieves the lowest number of parameters and the computational requirements, and shows the highest detection accuracy, which meets the lightweight network at the same time, and meets the deployment requirements of the obstacle detection effect. It is worth noting that the YOLOv8-PSS model proposed in this article has certain limitations in detecting objects with similar shapes, and there is still room for improvement in the detection performance of the model. The main reason may be that many obstacles in the self-built dataset in this article were collected in different scenarios, and in order to run the model on mobile platforms, the focus of this article is more on the lightweight research of the model. Therefore, in the subsequent research process, in order to further improve the detection performance of the model, more datasets will be obtained for training, and attention mechanisms will be added to the model to enhance detection performance. In summary, this article introduces a lightweight YOLOv8-PSS model for obstacle detection in agricultural machinery during operation, which has high efficiency and accuracy. This is of great significance for the research of obstacle avoidance in unmanned agricultural machinery.

To check whether the model can recognize the obstacle information, various scene input models were tested for obstacle detection, and the detection results are shown in Figure 10 From the figure, Faster R-CNN and YOLOV8_PSS performed better for the detection of obstacle information, SSD, YOLOv3-tiny and YOLOv5 have some degree of omission when recognizing different scenes, including people, poles, and trees. Notably, YOLOv5 encountered issues with misidentification, mistakenly recognizing a single farm machine as two separate entities. Although the algorithm proposed in this paper demonstrates strong performance in obstacle detection, it still faces challenges with tree detection. This is mainly because when the tree overlaps or occludes the phenomenon. This overlap complicates the model's ability to extract distinguishing features, ultimately leading to instances of missed detection.

TABLE 5 Information on different obstacle detection accuracy.

Type	P(%)	R(%)	mAP(%)
Person	88.5	90.2	92.3
Tree	82.1	86.2	89.6
Poles	81.3	86.7	88.4
Vehicle	86.7	88.6	90.9
Agm	87.9	90.3	91.8
Average value	85.3	88.4	90.6

3.3 Comparison of different modules

To verify the effectiveness of the PConv and Slim neck modules used, this paper uses an improved YOLOv8 object detection network as the base network. By replacing the mainstream lightweight backbone network and neck network, while maintaining the same parameters except for the backbone or neck network, the effects of different modules on target training are compared. The specific results are shown in Tables 7 and 8. According to Table 1, PConv has better advantages in mAP compared to ShuffleNetV2, MobileNetV2, and GhostNet backbone networks, with improvements of 9.2%, 6.4%, and 2.7%, respectively. Although MobileNetV2 has lower computational

TABLE 6 Comparison of detection results of different models.

References	Model	P/%	R%	mAP/%	Size/MB	Parameters/×10^{6}M	Computation/G	FPS
(Ren et al., 2016)	Faster RCNN	74.3	91.9	43.7	108	86.9	129.8	41
(Hui et al., 2019)	SSD	84.6	80.3	80.1	93.1	31.4	58.7	65
(Bin et al., 2021)	YOLOv3-tiny	82.8	81.2	79.5	24.5	20.7	46.3	77
(Su et al., 2022)	YOLOv5	80.7	84.6	83.9	13.2	9.0	15.2	86
	YOLOv8_PSS	85.3	88.4	90.6	4.7	2.3	6.0	111

complexity and model size than PConv, its obstacle calibration performance is poor and it is not suitable for practical deployment situations. According to Table 3, the Slim neck module has a certain improvement effect compared to other neck network improvement methods, with mAP increased by 10.3%, 10.9%, and 4.1%, respectively. Therefore, the use of Slim neck for lightweight improvement of the neck network has significantly improved the overall detection performance of the model compared to the original network, which can meet real-time detection requirements and is more suitable for mobile deployment of later models.

To further assess the effectiveness of the Shape-IoU bounding box loss function, we compared its convergence with that of the SIoU, CIoU, EIoU, and WIoU loss functions in enhancing the network's performance. The improved model comprises 2.3×10^6 parameters, with a computational load of 6.0 G and a model size of 4.7 MB when different bounding box loss functions are employed. The results indicate that substituting these loss functions does not impact the model's complexity. From Table 9, changing the bounding box loss function affects the precision, accuracy, and recall of the model to a certain extent. The recall of the improved model using Shape-IoU is

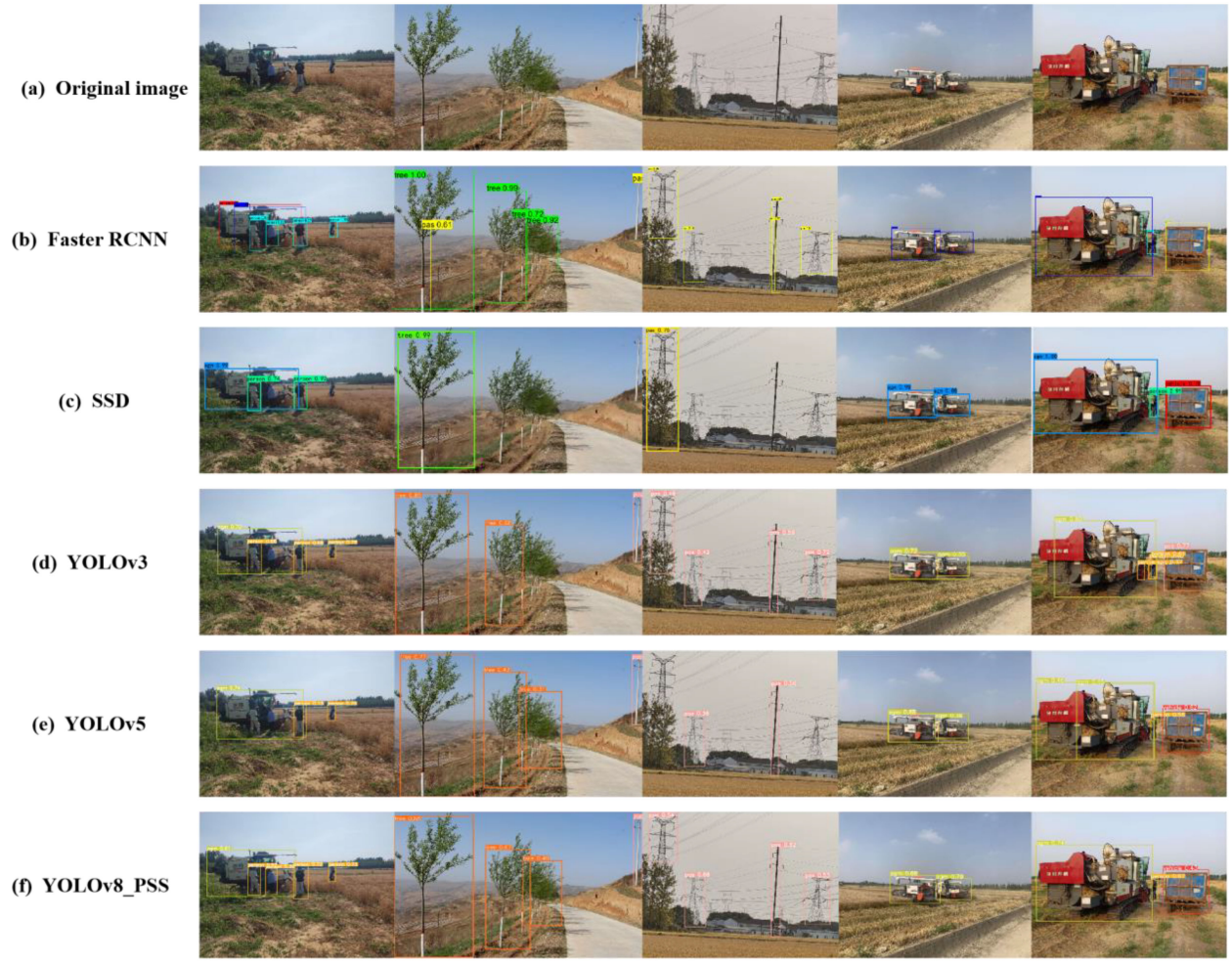


FIGURE 10
Comparison of the effectiveness of different detection models (**(A)** original image, **(B)** Faster RCNN, **(C)** SSD, **(D)** YOLOv3, **(D)** YOLOv5, **(F)** YOLOv8_PSS).

10.3389/fpls.2024.1509746

TABLE 7 Comparison of different lightweight backbone networks.

Backbone	P/ %	R/ %	mAP/ %	Size	Computation/ G
MobileNetV2	83.4	73.9	81.4	4.5	5.8
ShuffleNetV2	82.1	76.6	84.2	6.9	8.1
GhostNet	82.3	83.6	87.9	6.7	8.4
PConv	85.3	88.4	90.6	4.7	6.0

88.4%, which is 2.8%, 2.7%, 3.3%, and 4.6% higher than that of CIoU, SIoU, EIoU, and WIoU, respectively. Additionally, the mAP for the shape-IoU is 90.6%, which is 1.2%, 2.1%, 1.9%, and 0.7% higher than the corresponding values of other models. Although Shape-IoU precision is 0.9, 0.1, and 0.2% lower than SIoU, EIoU and WIoU, respectively, the optimal equilibrium between recall and mean average precision is achieved.

As shown in Figure 11, it is the convergence of YOLOv8 lightweight improved model respectively using different bounding box loss functions during training. As depicted, the loss values for the various bounding box loss functions begin to converge after 50 iterations on the validation set. Notably, the regression optimization of the bounding box using the EIoU loss function proves inefficient for datasets with complex scenarios, resulting in the highest bounding box loss value and the slowest convergence speed. In contrast, the SIoU, CIoU, WIoU, and Shape-IoU loss functions converge at a similar rate, achieving significantly lower convergence values compared to EIoU. Among these, the Shape-IoU loss function exhibits a smaller loss value upon convergence, indicating reduced susceptibility to overfitting and stronger generalization ability. Therefore, the Shape-IoU loss function demonstrates the best overall performance in this study.

3.4 Model positioning accuracy experiments

To further verify the positioning accuracy of the model in real scenarios, the YOLOv8_PSS model training file was converted into a blob file for compilation operation. Testing was conducted on June 12, 2024, at Rungo Farm Co. in Jingkou District, Zhenjiang City, Jiangsu Province, China. The obstacle detection and positioning distances ranged from 2 to 15 meters. Depth distance measurements were utilized for evaluation and analysis, with the DM120 laser rangefinder produced by DELIXI employed to detect

TABLE 8 Comparison of improvements in different neck networks.

Backbone	P/ %	R/ %	mAP/ %	Size	Computation/ G
Bifpn	77.7	82.1	80.3	6.3	8.2
RepGFPN	79.4	83.5	79.7	5.7	7.6
GD-YOLO	80.2	81.9	86.5	5.8	6.3
Slim neck	85.3	88.4	90.6	4.7	6.0

TABLE 9 Performance comparison of different bounding box loss functions.

Loss function	P/%	R/%	mAP/%
CIoU	83.9	85.6	89.4
SIoU	86.2	85.7	88.5
EIoU	85.4	85.1	88.7
WIoU	85.5	83.8	89.9
Shape-IoU	85.3	88.4	90.6

obstacles (measurement accuracy of ± 2 mm and a measurement range of 120 m). The measured results were subsequently analyzed, and the experimental findings are presented in Table 10, where Z represents the measurements from the camera and Zd corresponds to the measurements from the laser rangefinder.

The test results show that within a distance range of 2 to 15 m, the maximum average relative error of the measured distance is 2.73% while the maximum relative error of the measured distance is 4.44%, respectively. Meanwhile, from Table 10, the measurement accuracy for obstacles at middle and near distances is relatively high. However, with the increase of the distance between the obstacles and the camera, the measurement error also increases gradually. This is mainly due to the low maximum resolution of the camera itself during the acquisition, which makes the feature information collected in the collection of information on the objects at a distance is not rich enough. Additionally, outdoor lighting conditions contribute to background noise in the captured images, further affecting positioning accuracy. Positioning accuracy, as can be seen from the table, the overall accuracy of positioning during the test meets the needs of real-time detection. The results of on-site detection are shown in Figure Before testing, the depth camera is mounted on top of the agricultural machine. During operation, the depth camera identifies and detects objects in its field of view, providing their corresponding category, confidence level, and 3D spatial coordinates. As shown in Figure 12A, the detected object is identified as a person, with a confidence level of 0.86 and spatial 3D

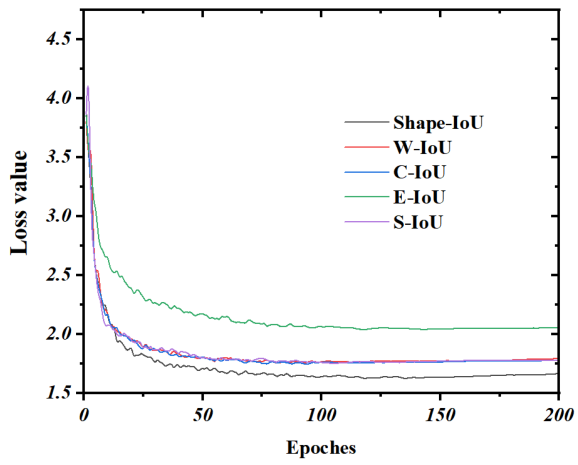


FIGURE 11
Iterative variation curves of the bounding box loss function.

TABLE 10 Obstacle positioning accuracy test results.

Number	Tree			Person			Pole			Vehicle			Agm		
	Z/m	Z_d/m	E/%	Z/m	Z_d/m	E/%	Z/m	Z_d/m	E/%	Z/m	Z_d/m	E/%	Z/m	Z_d/m	E/%
1	2.42	2.43	0.41	2.66	2.64	0.75	2.47	2.48	0.40	2.48	2.49	0.40	2.86	2.88	0.42
2	3.25	3.28	0.91	3.48	3.52	1.13	3.87	3.83	1.04	3.69	3.72	0.81	3.88	3.92	1.02
3	4.37	4.31	1.39	4.23	4.29	0.93	4.59	4.65	1.29	4.63	4.55	1.76	4.54	4.59	1.09
4	5.41	5.52	1.99	5.48	5.57	1.62	5.09	5.21	2.30	5.81	5.72	1.57	5.34	5.23	2.10
5	6.59	6.73	2.08	6.47	6.65	2.71	6.24	6.03	3.48	6.87	7.03	2.26	6.15	6.34	2.99
6	7.39	7.56	2.25	7.09	7.26	2.34	7.21	7.37	2.17	7.35	7.21	1.94	7.12	7.27	2.06
7	8.23	8.48	2.95	8.87	9.11	2.63	8.56	8.80	2.72	8.52	8.76	2.74	8.96	9.23	2.93
8	9.51	9.78	2.76	9.85	10.15	2.95	9.38	9.66	2.89	9.60	9.91	3.13	9.76	10.12	3.56
9	10.63	10.98	3.19	10.01	10.36	3.39	10.74	11.09	3.15	10.61	10.92	2.84	10.56	10.84	2.58
10	11.02	11.42	3.50	11.25	11.66	3.52	11.71	12.13	3.46	11.03	11.46	3.75	11.82	12.21	3.19
11	12.10	12.55	3.59	12.29	12.74	3.53	12.82	13.28	3.46	12.37	12.81	3.43	12.39	12.84	3.50
12	13.97	14.53	3.85	13.60	14.11	3.61	13.96	14.45	3.39	13.33	13.83	3.62	13.14	13.65	3.73
13	14.77	15.39	4.03	14.39	15.04	4.32	14.94	15.62	4.35	14.86	15.54	4.38	14.68	15.31	4.11
14	15.05	15.75	4.44	15.42	16.12	4.34	15.10	15.77	4.25	15.92	16.63	4.27	15.33	15.99	4.13
Average			2.67			2.69			2.73			2.64			2.67
Maximum			4.44			4.34			4.35			4.27			4.13

coordinates of (-0.04, -0.28, 4.24) (Figures 12A–E). During testing, the detection targets including people, farm machinery, vehicles, utility poles, and trees were 44, 33, 36, 37, and 42, respectively. The number of successfully detected targets was 44, 31, 35 and 38, with the success rates of 100%, 93.9%, 97.2%, 94.5%, and 90.4%, respectively, and the overall success rate of detection was 95.2%.

3.5 Research contribution to phytoprotection

Plants are often affected by pests, diseases and factors such as drought and salinity during their complex and varied growth process. To ensure their survival and fulfill environmental protection requirements, it is essential to investigate and develop effective pest and weed control strategies, as well as plant protection technologies, to achieve optimal pest management (ZFrische et al., 2018). With the increasing demand for mechanization in agricultural production, the role of agricultural machinery has gradually shifted from manual to autonomous operation. The use of unmanned agricultural machinery for plant protection has become increasingly widespread. However, due to the complexity and diversity of the unstructured farmland environment, the autonomous operation task of plant protection by unmanned agricultural vehicles require additional attention on obstacle avoidance, in addition to the challenge of achieving full-coverage

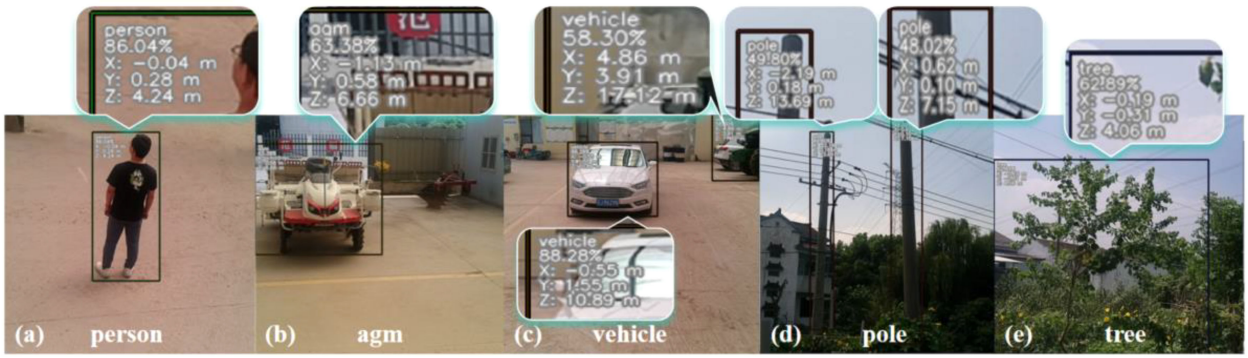


FIGURE 12
Visualization of the results of field detection and positioning (**(A)** person, **(B)** agm, **(C)** vehicle, **(D)** pole, **(E)** tree).

path planning. Farmland obstacles are not only diverse but also widely distributed, posing a significant threat to the traveling safety of unmanned agricultural vehicles. After detecting obstacles using the YOLOv8-PSS model proposed in this paper, the unmanned farm machine can adjust its travelling path based on the detection results to avoid damaging crops or equipment. For example, an unmanned sprayer used for plant protection can adjust its route according to the location of the obstacle, thus ensuring precise operation and preventing conflicts with crops or other agricultural machinery. The YOLOv8-PSS model is ideal for unmanned agricultural platforms with high efficiency, real-time processing and improved precision in farmland obstacle identification. By accurately identifying obstacles the model enables agricultural equipment to perform precise obstacle avoidance, optimize pesticide application, and enhance crop protection. This improves the intelligence of plant protection machinery and addresses challenges such as high operational precision requirements, complex operating environments, and safety risks associated with unmanned plant protection systems.

4 Conclusions

In this research work, to enable rapid obstacle detection during the autonomous navigation of unmanned agricultural machines in complex environments, an obstacle recognition method based on an improved YOLOv8 model is proposed. This approach offers a novel solution for obstacle detection and positioning in agricultural machinery and holds significant theoretical and practical implications for advancing intelligent unmanned agricultural systems. The key conclusions of this study are as follows:

1. Through the construction of farmland obstacle datasets, the YOLOv8_PSS obstacle detection model is proposed. The C2f-PConv module is added to the backbone to reduce network size, while the original neck is replaced by the lightweight Slim-Neck with GSConv and the VoVGSCSP module, cutting model parameters, computation, and size to 55.8%, 51.2%, and 59.5% of the original. Shape-IoU loss is also introduced to stabilize training, reduce regression and shape deviation, and improve convergence speed. The final model achieves 85.3% accuracy, 88.4% recall, and 90.6% mAP, balancing speed and precision better than Faster R-CNN, SSD, YOLOv3-tiny, and YOLOv5, making it ideal for mobile deployment

2. In the obstacle positioning accuracy test, the maximum average error and maximum error for the distance between five types of obstacles and the camera, measured within a range of 2 to 15 meters, were 2.73% and 4.44%, respectively. In the detection test, the model achieved a combined detection rate of 95.2% for obstacles. These results demonstrate the model's excellent performance in both positioning accuracy and obstacle detection, fulfilling the requirements for precise detection and localization of common obstacles in practical applications. Furthermore, this study sets a benchmark for obstacle detection and localization in unmanned agricultural vehicle environments, offering valuable insights for technological innovation and advancements in related fields. In practical applications, the improved model experiences some degree of obstacle detection failure, particularly in environments with strong or backlighting, where obstacles may blend with the background due to insufficient contrast. Additionally, the model demonstrates lower success rates in recognizing static obstacles compared to dynamic ones. To address these issues, the dataset used in this study primarily consists of images captured during moderate daylight conditions. Future research will aim to enhance the model by expanding the dataset with more diverse samples and further optimizing the model's architecture.

3. In this study, obstacle detection relied solely on a single stereo camera, which has a limited field of view. The detection system has certain limitations and can only obtain obstacle information on the front side of the agricultural unmanned operation platform. However, in actual road environments, many dynamic obstacles may move towards agricultural machinery from the side. In subsequent research, it is necessary to combine other sensors to detect obstacles, in order to improve the robustness and comprehensiveness of the perception system as much as possible. Different speed ranges and behavioral orientations should be matched for different obstacles, and obstacle avoidance strategies corresponding to different obstacles should be studied to increase the universality of the obstacle avoidance system. These improvements will strengthen obstacle detection in real-world scenarios, especially for dynamic obstacles. This will enhance safety during obstacle avoidance maneuvers, ensuring safer navigation.

Data availability statement

The datasets presented in this article are not readily available because the data is confidential. Any enquires can be directed to the corresponding author.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

ZC: Writing – original draft, Writing – review & editing. YF: Conceptualization, Writing – review & editing. JY: Conceptualization, Data curation, Formal analysis, Supervision, Writing – review & editing. SL: Data curation, Formal analysis, Investigation, Writing – review & editing. FM: Funding acquisition, Writing – review & editing. LL: Data curation, Formal analysis, 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 research was funded by the National Natural Science Foundation of China under Grant No. 52375248 and 52350410469.

Conflict of interest

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

Generative AI statement

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

Publisher's note

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

References

Bin, C., Man, Z., Hongzhen, X., Han, L., and Yanxin, Y. (2021). Farmland Obstacle Detection in Panoramic Image Based on Improved YOLO v3 tiny. *Trans. Chin. Soc. Agric. Machinery (Transactions CSAM)* 52, 58–65. doi: 10.6041/j.issn.1000-1298.2021.S0.008

Chen, J., S-h, K., He, H., Zhuo, W., Wen, S., Lee, C.-H., et al. (2023). Run, don't walk: chasing higher FLOPS for faster neural networks. *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.* 2023, 12021–12031. doi: 10.48550/arXiv.2303.03667

Dong, H., Zhang, Y., Gu, H., Konz, N., Zhang, Y., and Mazurowski, M. A. (2023). SWSSL: Sliding window-based self-supervised learning for anomaly detection in high-resolution images. *IEEE Trans. Med. Imaging.* 42 (12), 3860–3870. doi: 10.1109/TMI.2023.3314318

Fu, Q., Yang, Y., Chen, X., and Shang, Y. (2020). Vision-based obstacle avoidance for flapping-wing aerial vehicles. *Sci. China Inf. Sci.* 63, 1–3. doi: 10.1007/s11432-019-2750-y

Garg, D., and Alam, M. (2023). Smart agriculture: A literature review. *J. Manage. Analytics* 10, 359–415. doi: 10.1080/23270012.2023.2207184

Hu, T., Wang, W., Gu, J., Xia, Z., Zhang, J., and Wang, B. (2023). Research on apple object detection and localization method based on improved yolox and rgb-d images. *Agronomy* 13, 1816. doi: 10.3390/agronomy13071816

Hui, L., Lishuai, Z., Yue, S., Jian, Z., and Bian, W. (2019). Real-time pedestrian detection in orchard based on improved SSD. *Trans. Chin. Soc. Agric. Machinery (Transactions CSAM)* 50, 29–35. doi: 10.6041/j.issn.1000-1298.2019.04.003

Ji, W., Gao, X., Xu, B., Pan, Y., Zhang, Z., and Zhao, D. (2021). Apple target recognition method in complex environment based on improved YOLOv4. *J. Food Process Eng.* 44, e13866. doi: 10.1111/jfpe.13866

Ji, W., Pan, Y., Xu, B., and Wang, J. (2022). A real-time apple targets detection method for picking robot based on ShufflenetV2-YOLOX. *Agriculture* 12, 856. doi: 10.3390/agriculture12060856

Ji, W., Wang, J., Xu, B., and Zhang, T. (2023). Apple grading based on multi-dimensional view processing and deep learning. *Foods* 12, 2117. doi: 10.3390/foods12112117

Jiansheng, W., Shuguo, P., Guangzhao, T., Wang, G., and Yingchun, S. (2021). Design and experiments of the binocular visual obstacle perception system for agricultural vehicles. *Trans. Chin. Soc. Agric. Eng. (Transactions CSAE)* 37, 55–63. doi: 10.11975/j.issn.1002-6819.2021.09.007

Li, H., Li, J., Wei, H., Zheng, L., Zhan, Z., and Ren, Q. (2024). Slim-neck by GSConv: a lightweight-design for real-time detector architectures. *J. Real-Time Image Process.* 21, 62. doi: 10.48550/arXiv.2206.02424

Li, S., Feng, C., Liang, X., Qin, H., Li, H., and Shi, L. (2018). A guided vehicle under fire conditions based on a modified ultrasonic obstacle avoidance technology. *Sensors* 18, 4366. doi: 10.3390/s18124366

Liu, Q., Lv, J., and Zhang, C. (2024). MAE-YOLOv8-based small object detection of green crisp plum in real complex orchard environments. *Comput. Electron. Agric.* 226, 109458. doi: 10.1016/j.compag.2024.109458

Ming, R., Jiang, R., Luo, H., Lai, T., Guo, E., and Zhou, Z. (2023). Comparative analysis of different uav swarm control methods on unmanned farms. *Agronomy* 13, 2499. doi: 10.3390/agronomy13102499

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

Su, F., Zhao, Y., Shi, Y., Zhao, D., Wang, G., Yan, Y., et al. (2022). Tree trunk and obstacle detection in apple orchard based on improved YOLOv5s model. *Agronomy* 12, 2427. doi: 10.3390/agronomy12102427

Wang, P., Gao, M., Sun, Y., Zhang, H., Liao, Y., and Xie, S. (2022). A vibration-powered self-contained node by profiling mechanism and its application in cleaner agricultural production. *J. Cleaner Production* 366, 132897. doi: 10.1016/j.jclepro.2022.132897

Wang, J., Gao, Z., Zhang, Y., Zhou, J., Wu, J., and Li, P. (2021). Real-time detection and location of potted flowers based on a ZED camera and a YOLO V4-tiny deep learning algorithm. *Horticulturae* 8, 21. doi: 10.3390/horticulturae8010021

Yang, Y., Li, Y., Wen, X., Zhang, G., Ma, Q., Cheng, S., et al. (2022). An optimal goal point determination algorithm for automatic navigation of agricultural machinery: Improving the tracking accuracy of the Pure Pursuit algorithm. *Comput. Electron. Agric.* 194, 106760. doi: 10.1016/j.compag.2022.106760

Yang, S., Wang, W., Gao, S., and Deng, Z. (2023). Strawberry ripeness detection based on YOLOv8 algorithm fused with LW-Swin Transformer. *Comput. Electron. Agric.* 215, 108360. doi: 10.1016/j.compag.2023.108360

Yang, Y., Xing, W., Qianglong, M., et al. (2022). Real time planning of the obstacle avoidance path of agricultural machinery in dynamic recognition areas based on Bezier curve. *Trans. Chin. Soc. Agric. Eng. (Transactions CSAE)* 38, 34–43. doi: 10.11975/j.issn.1002-6819.2022.06.004

ZFrische, T., Egerer, S., Matezki, S., and Wogram, J. (2018). 5-Point programme for sustainable plant protection. *Environ. Sci. Europe* 30, 1–17. doi: 10.1186/s12302-018-0136-2

Zhang, Z., Li, J., Su, C., Wang, Z., Li, Y., Li, D., et al. (2024). A method for counting fish based on improved YOLOv8. *Aquacultural Eng.* 107, 102450. doi: 10.1016/j.aquaeng.2024.102450

Zhang, Z., Lu, Y., Zhao, Y., Pan, Q., Jin, K., Xu, G., et al. (2023). Ts-yolo: an all-day and lightweight tea canopy shoots detection model. *Agronomy* 13, 1411. doi: 10.3390/agronomy13051411

Zhang, H., and Zhang, S. (2023). Shape-iou: More accurate metric considering bounding box shape and scale. *arXiv preprint arXiv*, 231217663. doi: 10.48550/arXiv.2312.17663

Zhao, X., Wang, K., Wu, S., Wen, L., Chen, Z., Dong, L., et al. (2023). An obstacle avoidance path planner for an autonomous tractor using the minimum snap algorithm. *Comput. Electron. Agric.* 207, 107738. doi: 10.1016/j.compag.2023.107738

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 January 2025
DOI 10.3389/fpls.2024.1491397

Check for updates

OPEN ACCESS

EDITED BY
Jun Ni,
Nanjing Agricultural University, China

REVIEWED BY
Renata Retkute,
University of Cambridge, United Kingdom
Nedeljko Latinovic,
University of Montenegro, Montenegro
Jie Guo,
Zhejiang University, China

*CORRESPONDENCE
Zhou Yang
Yangzhou@scau.edu.cn
Xing Xu
xuxing@scau.edu.cn

RECEIVED 04 September 2024
ACCEPTED 16 December 2024
PUBLISHED 09 January 2025

CITATION
Yu J, Xu X, Duan J, Jiang Y, Yuan H, Liang H, Jing S and Yang Z (2025) Effect of operational parameters on droplet deposition characteristics using an unmanned aerial vehicle for banana canopy. *Front. Plant Sci.* 15:1491397. doi: 10.3389/fpls.2024.1491397

COPYRIGHT
© 2025 Yu, Xu, Duan, Jiang, Yuan, Liang, Jing and Yang. 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.

Effect of operational parameters on droplet deposition characteristics using an unmanned aerial vehicle for banana canopy

Jiaxiang Yu[1,2], Xing Xu[1]*, Jieli Duan[1,2], Yinlong Jiang[1,2], Haotian Yuan[1,2], Huazimo Liang[1], Shuaijie Jing[1,2] and Zhou Yang[1,2,3]*

[1]College of Engineering, South China Agricultural University, Guangzhou, China, [2]Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, China, [3]School of Mechanical Engineering, Guangdong Ocean University, Zhanjiang, China

In recent years, as an important part of precision agricultural aviation, the plant protection unmanned aerial vehicle (UAV) has been widely studied and applied worldwide, especially in East Asia. Banana, as a typical large broad-leaved crop, has high requirements for pests and diseases control. The mechanization degree of plant protection management in banana orchard is low. Therefore, our study focuses on the effects of different flight heights (3-5 m) and droplet sizes (50-150 μm) of plant protection UAV on the droplet deposition distribution characteristics of banana canopy. And the droplet deposition distribution in banana canopy and spraying drift of plant protection UAV and ground air-assisted sprayer were compared. The results showed that droplet size was the main factor affecting droplet deposition density, coverage, uniformity and penetration on both sides of banana canopy leaves. The droplet deposition density and coverage on the adaxial side of leaves were mostly significantly larger than that on the abaxial side. The flight height of 4 m and the droplet size of 100 μm could make the adaxial side of banana canopy leaves have higher droplet deposition density (63.77 droplets per square cm) and coverage (12.75%), and can make the droplets effectively deposit on the abaxial side of banana canopy leaves, with droplet deposition density of 17.46 droplets per square cm and coverage of 1.24%. Choosing an appropriate flight height and a droplet size could improve the droplet deposition uniformity on both sides of banana canopy leaves, but the improvement was not significant. Moreover, at a same operational parameter combination, it was difficult to achieve the best droplet deposition density, coverage, uniformity and penetration at the same time. In addition, appropriately increasing the flight height and droplet size could help to improve the droplet deposition penetration on the adaxial side of banana canopy leaves, but there were few significant improvements. Compared with the plant protection UAV, the ground air-assisted sprayer had higher droplet deposition density and coverage on the abaxial side of banana canopy leaves, but had smaller droplet deposition coverage on the adaxial side. The droplet deposition density and coverage on the abaxial side of banana canopy leaves were obviously larger than the adaxial side during the spraying of ground air-assisted sprayer. The droplet drift distance of the ground air-assisted sprayer was

farther than the plant protection UAV. The test results of this study can provide practical and data support for the UAV aerial application in banana orchard, and provide a valuable reference for the implementation of air-ground cooperation spraying strategy in banana orchard, which is of great significance to promote sustainable and intelligent phytoprotection of banana orchard.

KEYWORDS

aerial application, unmanned aerial vehicle, banana, droplet deposition, operational parameters

1 Introduction

The banana (Musa spp.), is the most traded and consumed fruit in the world, mainly planted in tropical and subtropical area, and it is an important economic crop and food crop (Guo et al., 2020; Xie, 2019). The banana industry, as the industry with the largest output value of single crop in the tropical region China, is very important for increasing farmers' income, agricultural efficiency and promoting rural development and revitalization (Jiang T. et al., 2023). As a typical large broad-leaved crop, banana has the characteristics of closed canopy and interlaced branches and leaves, and has high requirements for pests and diseases control (Jiang et al., 2022; Jiang Y. et al., 2023; Zhang et al., 2021). Spraying chemical pesticides to control pests and diseases is one of the important links in the production and management of the banana orchard which directly affects the yield and fruit quality of bananas, and is crucial to the economic benefits of the entire banana industry (Jiang T. et al., 2023; Robson et al., 2007). Therefore, it is of great significance to enhance the effective chemical control of banana pests and diseases for the development of banana industry.

Banana orchards in China are mainly distributed in hilly and mountainous areas such as Guangdong, Guangxi and Fujian (Xu and Liu, 2021). Due to the lack of special pesticide spraying equipment for banana orchard, traditional ground spraying equipment such as self-propelled sprayers or knapsack sprayers are usually used to improve pesticide deposition by over-spraying method, resulting in serious pesticide residues and ecological environment pollution (Barraza et al., 2020; Sinha, 2020). The use of self-propelled sprayer or artificial knapsack sprayer in banana orchard requires the construction of roads where the land is sufficiently flat for the sprayer and the operator to pass through, which will greatly increase the cost and reduce the banana planting area (Matthews, 2009). In addition, where the self-propelled sprayer is used to spray bananas the nozzles need to be elevated above the banana canopy in order for droplets to be dispersed horizontally across the canopy to reach the young leaves. When using the knapsack sprayer, the lower banana leaves will prevent the droplets from reaching the leaves at the top of the canopy. At present, these problems have seriously affected the economic benefits of banana industry in China, and it is necessary to apply more efficient and safe pesticide spraying technology and equipment to banana orchard management.

In recent years, plant protection unmanned aerial vehicle (UAV) has gradually become the preferred plant protection equipment in China due to its special advantages over traditional pesticide spraying equipment (Lan and Chen, 2018). The rotor downwash airflow of the plant protection UAV can disturb the crop leaves, so that droplets can be attached on both sides of the leaves. In addition, the operation of UAV is not limited by terrain and crop growth, and does not destroy the soil structure, while the operator has no risk of poisoning (Xue et al., 2008; Zhou et al., 2014). Therefore, using the plant protection UAV to spray pesticides in banana orchards is an important method to improve the utilization rate of pesticides and achieve high deposition of pesticides on banana leaves. It is also an important way to solve the bottleneck of pesticide application in the banana industry.

With the rapid development of China's agricultural aviation, the spraying technology of plant protection UAV has become a research hotspot (Chen et al., 2020). Due to the application of plant protection UAV in orchards is still in the initial stage, there are still a series of serious problems in UAV spraying technology, such as unclear optimal operational parameters, poor droplet penetration, low droplet coverage, and uneven droplet distribution (Qin et al., 2016). Aiming at these problems, some researchers have conducted a series of field spray researches on the effects of plant protection UAV operational parameters on the droplet deposition distribution characteristics of fruit crop, such as areca (Wang et al., 2019b; Wang J. et al., 2020; Wang et al., 2023), peach (Meng et al., 2020), citrus (Chen et al., 2017b; Nascimento and Vitória, 2022; Tang et al., 2018), apple (Li et al., 2018; Liu et al., 2020), pineapple (Wang et al., 2019a), grape (Biglia et al., 2022), guava (Verma et al., 2022), coconut (Pandiselvam et al., 2024). These researches mainly focus on forest fruit crops with spindle-shaped canopy, with a few studies on broad-leaved fruit crops, only large broad-leaved woody crops such as coconut and areca, but the research on large broad-leaved herb crops such as banana is almost zero. The shielding effect between leaves of the forest crop is small, and the droplets wrapped by airflow are easy to enter the canopy through the gap between leaves (Guo X. et al., 2021). However, the canopy of large broad-leaved crops such as banana is inverted

triangle shaped, with leaves are similar to cattail fan, resulting in a great shielding effect between the leaves. Different from banana leaves, the leaves of coconut and areca have feathery compound leaf structure, which can reduce the shielding effect between leaves. The difference in crop shape and structure makes it difficult to apply the existing research results to banana pest and disease control. In addition, the existing research on broad-leaved fruit crops does not consider the droplet deposition effect on the abaxial side of crop canopy leaves (Wang et al., 2019b; Wang J. et al., 2020; Wang et al., 2023; Pandiselvam et al., 2024). A large number of research results show that the abaxial side of banana leaves is the main occurrence area of pests and diseases because of its suitable temperature, not being directly washed by rain, and not being directly exposed to sunlight (Huang and Wei, 2017). Therefore, the effective droplet deposition on the abaxial side of banana leaves is of positive significance for crop protection by chemical pesticide spraying. Flight height and droplet size are the main factors affecting the application effect of plant protection UAV (OECD, 2021). Too high flight height may cause a significant decrease in droplet deposition on crop canopy leaves; too low flight height may aggravate the occlusion between the crop canopy leaves, thus influencing the droplet deposition effect on the lower canopy leaves (Chen et al., 2022). The droplet size is not only a key index to characterize the pesticide properties, but also an important factor affecting the droplet deposition distribution in crop canopy (Wang et al., 2015; Yao et al., 2017). The droplets of a small size can easily penetrate the crop canopy to deposit with a high deposition, while the droplets of a large size tend to accumulate, slide or bounce, thus affecting the droplet deposition effect (Ding et al., 2022; Qin et al., 2016).

In summary, the study of droplet deposition distribution of banana canopy by plant protection UAV is crucial for the pests and diseases control of banana orchard. This study aimed to evaluate the effect of droplet deposition on both sides of banana canopy leaves at different flight heights and droplet size parameters. The plant protection UAV operational parameters suitable for banana orchard were selected from the perspective of improving droplet deposition, droplet deposition uniformity and penetration. In addition, the droplet deposition distribution in banana canopy and spraying drift of plant protection UAV and ground air-assisted sprayer were compared. These provide data and practical basis for UAV aerial application in banana orchard, and guide the decision-making of the operation model of mechanized pesticide spraying in banana orchards.

2 Materials and methods

2.1 Plant protection UAV and its spraying system

As shown in Figure 1, the model of UAV used in this spraying test was eight-rotor electric UAV (MG-1P, Shenzhen DJI Technology Co., Ltd., Shenzhen, China). The size of the UAV was 1460 mm × 1460 mm × 616 mm, the rotor diameter was 533 mm, the maximum pesticide tank load was 10 L, the maximum flight speed was 7 m/s, the maximum flight height was 30 m, and the flying time was about 10 min with a full tank. The spraying system included a diaphragm pump and two centrifugal atomizers developed by our team (Yang et al., 2021). The two atomizers were installed under the two main propeller shafts on the rear side of the UAV. The effective spray width was 4-6 m. The atomizers and water pump were both powered by the UAV battery, and the parameters were manually set by the controller before each flight.

FIGURE 1
UAV and spraying system used in the test.

2.2 Experimental design

2.2.1 Experimental area and crop characteristics

The tests were conducted on February 23, 2023 at the banana experimental base of the Banana and Vegetable Institute of Dongguan Agricultural Science Research Center located at Dongguan, Guangdong Province, China (latitude 113.698862, longitude 23.006825) (Figure 2). The tested crop was Brazil banana in the budding stage. The average height was 3.5 m, and the row spacing was 2.0 m × 2.5 m. Budding stage is an important period in the whole growth stage of banana, which is not only the key to high yield, but also related to the quality of banana (Jing et al.,

FIGURE 2
Experimental area.

2022). This stage is also a period of high incidence of pests and diseases, such as leaf spot disease, yellow leaf disease, nematodes, flower thrips, weevil (Constantinides and Jr. McHugh, 2003). Therefore, the budding stage is the key stage of pest and disease management in the whole growth stage of banana.

2.2.2 Sampling scheme

In order to reduce the experimental error, the UAV flied along the crop line (northwest - southeast) during spraying test and three healthy banana plants with similar growth condition were selected on the flight path for sampling in each test treatment. The site layout and flight path are shown in Figure 3.

The water sensitive paper (40 mm × 30 mm, Chongqing Liuliushanxia Co., Ltd., Chongqing, China) was used to evaluate droplet deposition effect at the sampling location. As shown in Figure 4A, there were three sampling layers on the canopy in each sample banana plant, which were upper, middle and lower layers, respectively. Three sampling layers were 3.0 m, 2.1 m and 1.3 m above the ground, respectively. The angles between the upper, middle and lower layer leaves and pseudo-stem were 30°, 60° and 120°, respectively. Among them, the transverse area of the upper layer is small, and the angle between the leaves and the pseudo-stem is small, which is the leaf center part that is prone to pests and diseases. The middle and lower layers have larger transverse area, dense leaves and increased angle with pseudo-stem, which are important parts of banana canopy prone to pests and diseases. In each layer, four sampling points were arranged in a circular and equal array around the pseudo-stem and the diameters of the circular array of the upper, middle and lower layers were 1.5 m, 1.6 m and 1.4 m, respectively (Figure 4B). Each sampling point was marked with two numbered white filter paper fixed on both sides of the leaf in advance, so as to accurately place and collect the water sensitive paper on both sides of the leaf at each sampling point, thus avoiding variability. A total of 72 water sensitive paper was collected from three banana plants in each test treatment. This configuration of water sensitive paper and filter paper is widely used to evaluate the actual droplet deposition in crop canopy during spraying tests (Braekman et al., 2009; Miranda-Fuentes et al., 2016; Nuyttens et al., 2004; Zhou and He, 2016).

2.2.3 Operational parameters

Combined with the previous spraying experience of the eight-rotor plant protection UAV operator and banana farmers, it was found that the flight height had a great influence on the spraying effect. Furthermore, in our previous pre-test, it was found that when the flight height was 3-5 m from the top of the banana canopy and the droplet size was 50-150 μm, the droplets could be well deposited in the banana canopy, especially on the abaxial side of banana leaves. Therefore, this experiment mainly aimed at changing the flight height and droplet size of the UAV under the same area spraying amount condition. The flight speed was set to 1.5 m/s and the spraying amount was 75 L/ha. The effects of plant protection UAV flight height (3-5 m) and droplet size (50-150 μm) on droplet deposition distribution were studied by a two-factor and three-level full-factor test. The specific factors and parameters are shown in Table 1. The droplet volume median diameter (VMD) (Czaczyk, 2012) was selected as a parameter to characterize the droplet size. The value was controlled by the rotation speed of the centrifugal atomizer and measured by the laser particle-size analyzer in the wind-tunnel lab of the National Center for International Collaboration Research on Precision Agricultural Aviation Pesticide Spraying Technology (NPAAC) in South China Agricultural University based on the ASABE S572.3 standard (ASAE, 2020) according to the method described in our previous study (Yang et al., 2023). The measurement result of VMD (Dv50), Dv10 and Dv90 were shown in Table 2, where Dvn represents the sum of the volume of the given size and below, accounting for n% of the total volume of the sample. In addition, the 1‰ OP-10 (DOWSIL, Miland, USA) aqueous

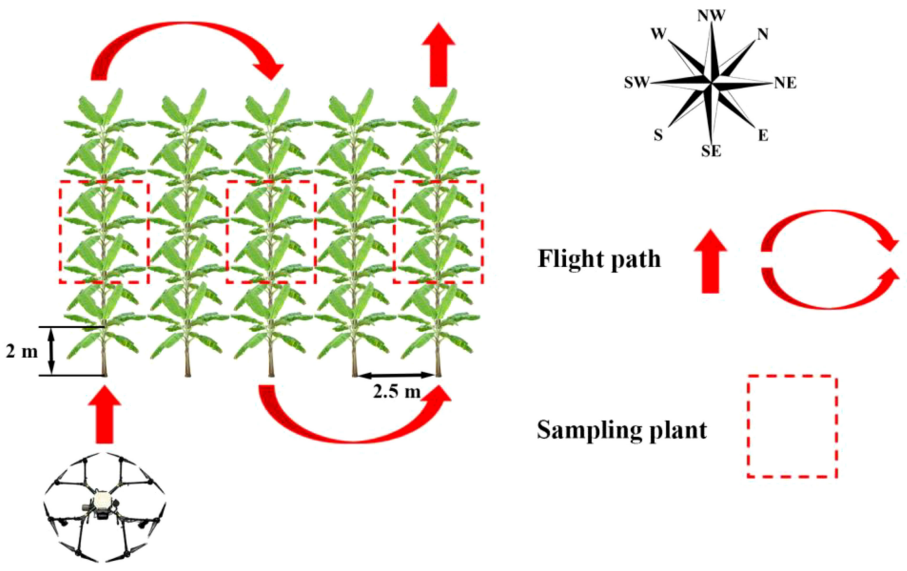


FIGURE 3
Schematic diagram of site layout and flight path in the test.

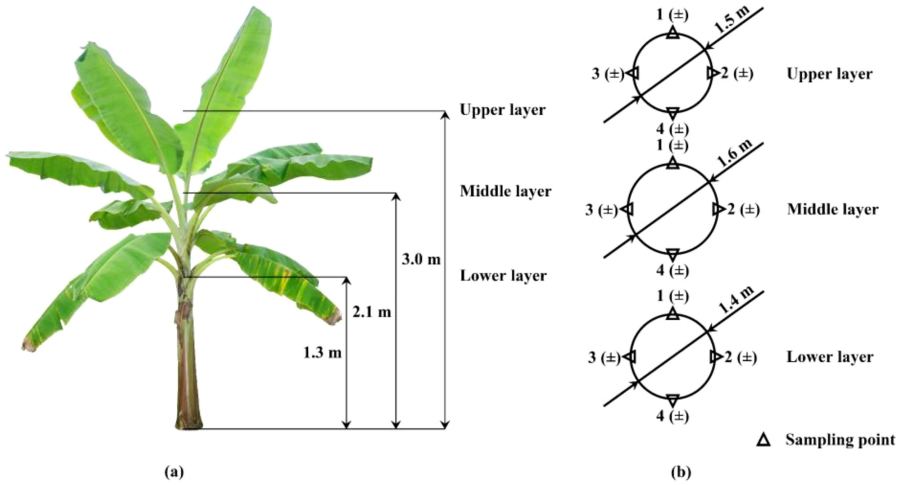


FIGURE 4
Sampling point layout schematic diagram: **(A)** Canopy sampling layer division; **(B)** Each layer sampling point layout.

surfactant solution was used in the experiment, which has similar physical properties to pesticides (Wang G. et al., 2020). Solvents were prepared before each application and fully mixed.

2.2.4 Weather conditions parameters

Weather conditions parameters were monitored and recorded by portable weather station (Kestrel 5500, Nielsen-Kellerma, Minneapolis, USA) over the full duration of the trials, including wind speed, temperature and humidity. The measurement accuracy of the wind speed, temperature and humidity are ±3%, 1°C and 3%, respectively. The weather station was placed 15 m from the experimental area with a recording frequency of 1 Hz.

The temperature and humidity ranged between 17.1°C and 23.7°C, and 39.6% and 69.1%, respectively, while the wind speed recorded in each test treatment was always below 1.5 m/s. The weather was in the optimal conditions for spraying applications according to the definition of the Best Management Practices (TOPPS-Prowadis Project, 2014). In addition, the angle between the wind direction and the flight direction was less than 30°, which met the requirement of ASAE Standards (ASAE, 2018). In all cases, the experimental trials were conducted in "light air" conditions, making the data obtained from the different test treatments broadly comparable (Barua, 2005). The weather data corresponding to each test treatment are shown in Table 3.

2.2.5 Comparison of droplet deposition and drift with ground air-assisted sprayer

The tests were conducted on November 7, 2024 at the banana orchard of South China Agricultural University. The tested banana plant was Brazil banana in the budding stage, and the growth was similar to that of the banana plant tested at the Banana and Vegetable Institute of Dongguan Agricultural Science Research Center in February 23, 2023.

Two treatments were set up, which were spraying by plant protection UAV and spraying by ground air-assisted sprayer, respectively. In the treatment of spraying by plant protection UAV, the UAV flied along crop line from north to south and the optimized operational parameter of flight height and droplet size were set. The ground air-assisted sprayer developed by our team was selected as the comparison machine type of the treatment of spraying by ground air-assisted sprayer (Xie, 2024). According to the optimal operational parameters of ground air-assisted sprayer for banana leaves obtained in our previous study, the nozzle type of hollow cone nozzle, the spray pressure of 0.5 Mpa and the wind speed of 4 m/s were set, and the spraying mode was inter-row spraying mode (Jiang Y. et al., 2023). Weather parameters of the two treatments are shown in Table 4.

The site layout and spraying path of plant protection UAV and ground air-assisted sprayer are shown in Figure 5. The drift sampling zone was a bare soil ground without obstacles and vegetation downwind the edge of the sprayed area. Spray drift samples were collected from ground sediment spraying drift. The route of the ground air-assisted sprayer to end the operation was set as the starting edge of the spray drift zone, and a total of 9 drift sampling points were set at 1, 2, 3, 4, 5, 7.5, 10, 15, and 20m away from it. A total of 3 drifting bands were used as 3 replicates, with a total of 27 sampling points per treatment.

TABLE 1 Factors and parameters of experiment.

Factors	Parameters
Flight height (m)	3, 4, 5
Droplet size (μm)	50, 100, 150

TABLE 2 Results of droplet size measurement.

Droplet size (μm)	Rotation Speed (r/min)	VMD (μm)	Dv0.1 (μm)	Dv0.9 (μm)
50	9800	50.30 ± 0.95	16.30 ± 0.13	100.51 ± 1.71
100	1650	100.08 ± 0.40	36.61 ± 1.71	333.54 ± 2.36
150	600	150.48 ± 0.70	72.96 ± 1.97	353.32 ± 1.31

The measurement result of droplet size in this table represent mean ± standard deviation.

TABLE 3 Weather conditions parameters during the test.

Test Treatment	Flight height (m)	Droplet size (μm)	Mean Temperature (°C)	Mean Humidity (%)	Mean Wind Speed (m/s)	Mean Wind Direction
1	3	50	23.0	39.6	0.9	NW
2	3	100	18.8	64.3	0.9	ESE
3	3	150	17.1	68.8	0.8	WNW
4	4	50	23.0	40.7	0.4	WNW
5	4	100	19.7	56.5	0.9	NNW
6	4	150	17.4	69.1	0.5	ESE
7	5	50	23.7	39.5	0.7	NW
8	5	100	23.4	41.0	0.9	WNW
9	5	150	17.9	67.3	0.6	ESE

2.3 Characterization of the deposition effect and statistical analysis

After each test treatment, the water sensitive paper was immediately collected and marked by operators wearing rubber gloves after the droplets had dried, then placed into sealed bags and stored in a cool and dry environment. After all tests, all collected water sensitive paper were taken back to the laboratory and scanned one by one at high resolution (600 dpi) using a printer with an integrated scanner (Hewlett-Packard, Palo Alto, USA). The scanned water sensitive paper was then analyzed by using an open source image processing software ImagePy (Zhang et al., 2024), which can obtain data such as droplet coverage density (the number of water sensitive paper droplets per unit area), droplet deposition coverage (the ratio of the target surface area covered by droplets to the total target surface area) and droplet deposition amount (the deposition amount of droplet per unit area).

According to the existing UAV application standards, the droplet deposition density (15 droplets per square cm) is used to characterize the droplet deposition effect at the sampling position and no indicator for the droplet coverage (Cerruto et al., 2019; Ministry of Agriculture of the PRC, 2018; Fox et al., 2003). At present, there is no evaluation index of canopy effective droplet deposition suitable for large broad-leaved crops. Although droplet deposition density was used as the most important evaluation index when measuring the spray quality of UAVs, it has been proved that percentage area seems to be the most reliable (Salyani and Fox, 1999). In orchard studies, some researchers used droplet deposition coverage of 1% as an evaluation indicator for effective droplet deposition (Wang et al., 2022). Therefore, combined with the previous analysis, this study used droplet deposition density of 15 droplets per square cm and droplet deposition coverage of 1% as the evaluation index for effective droplet deposition. And the droplet deposition amount (μl/cm^2) was used to characterize the deposition of droplet drift sampling points.

The droplet deposition uniformity was characterized by calculating the mean value of the coefficient of variation (CV) of the droplet deposition density and coverage at each layer's sampling points. The droplet deposition penetration was characterized by calculating the CV of the mean droplet deposition density and coverage of each sampling layer. Usually, the CV is used as a measure of the uniformity of a set of sampling data by calculating the extent of variation of this set of data. The smaller the CV is, the smaller the change range of the representative data is, which means that the more uniform the droplet deposition, the better the penetration (Liao et al., 2015). The CV calculation equation is as follows:

$$CV = \frac{S}{\overline{X}} \times 100\,\% \tag{1}$$

$$S = \sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)} \tag{2}$$

Where X_i is the droplet deposition density (droplets per square cm) or coverage (%) on the i-th water sensitive paper, $\overline{X}$ is the mean value of X_i, n is the total number of samples from the water sensitive paper, and S is the standard deviation of X_i.

In this study, all statistical analyses were performed using IBM SPSS Statistics 27 software. Multivariate analysis of variance at the significance level of 0.05 was used to obtain the effects of flight height and droplet size on droplet deposition density, coverage, uniformity and penetration on both sides of banana canopy leaves. The differences of different levels of each factor were analyzed and compared using Duncan's *post hoc* test at a significance level of 0.05.

TABLE 4 Weather conditions parameters during the treatment of spraying by plant protection UAV and spraying by ground air-assisted sprayer.

Treatment	Mean Temperature (°C)	Mean Humidity (%)	Mean Wind Speed (m/s)	Mean Wind Direction
Plant protection UAV	22.3	56.8	1.1	ENE
Air-assisted sprayer	21.3	60.2	0.9	ESE

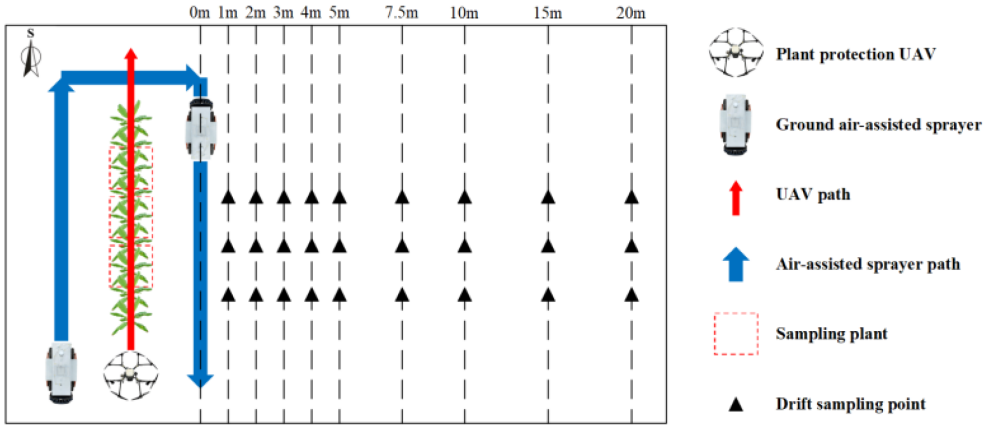


FIGURE 5
Schematic diagram of site layout and spraying path of plant protection UAV and ground air-assisted sprayer.

3 Results

3.1 Droplet deposition

The variance analysis results of the effects of flight height, droplet size and their interaction on droplet deposition density and coverage on both sides of banana canopy leaves were carried out, as shown in Table 5. The results showed that the droplet size had a significant effect on droplet deposition density on both sides of banana canopy leaves ($P < 0.05$), and the flight height and their interaction had no significant effect on droplet deposition density ($P > 0.05$). The droplet size had a significant effect on droplet deposition coverage on the adaxial side of banana canopy leaves ($P < 0.05$) and had no significant effect on the abaxial side ($P > 0.05$), and the flight height and their interaction had no significant effect on droplet deposition coverage ($P > 0.05$). The strongest factor influencing the droplet deposition density and coverage on both sides of the banana canopy leaves was the droplet size, with F values of 8.988, 3.623 and 37.464, 1.344, respectively. The second was the flight height, with F values of 0.998, 2.256 and 1.467, 0.800, respectively. In order to further study the effects of flight height and droplet size on the droplet deposition density and coverage of banana canopy leaves, we conducted a statistical analysis of the test results, as shown in Figure 6.

It can be observed from Figure 6 that when the flight height was 4 m and the droplet size was 100 μm, the droplet deposition density on the adaxial side of the banana canopy leaves was the largest (63.77 droplets per square cm). However, the droplet deposition coverage on the adaxial side of 4 m flight height and 150 μm droplet size was the largest (14.33%). When the flight height was 4 m and the droplet size was 100 μm, both the droplet deposition density and coverage on the abaxial side of the banana canopy leaves was the largest, were 17.46 droplets per square cm and 1.24%, respectively.

When the flight height was constant, the droplet deposition density on the adaxial side of the banana canopy leaves increased first and then decreased with increasing droplet size, but the droplet

TABLE 5 Variance analysis of droplet deposition density and coverage of banana canopy leaves.

Source	Degree of Freedom	Adaxial Side				Abaxial Side			
		Droplet Deposition Density		Droplet Deposition Coverage		Droplet Deposition Density		Droplet Deposition Coverage	
		F	P	F	P	F	P	F	P
Flight height	2	0.998	0.388	1.467	0.257	2.256	0.134	0.800	0.465
Droplet size	2	8.988	0.002	37.464	< 0.01	3.623	0.048	1.344	0.286
Flight height *Droplet size	4	0.145	0.963	0.202	0.934	0.669	0.622	0.306	0.870
Error	18								
Total	26								

The P and F value indicate the significance level, and $P < 0.05$ represents factors that have a significant effect on the test result.

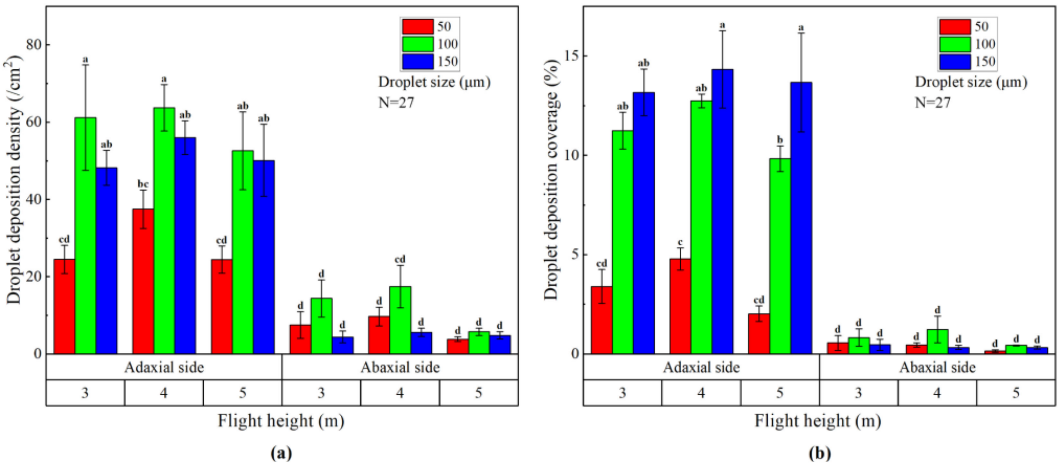


FIGURE 6
The histogram of droplet deposition density **(A)** and coverage **(B)** on banana canopy leaves. The data in this figure represent mean ± standard deviation. Columns with different lowercase letters are significantly different at the P <0.05 level based on Duncan's new multiple range test.

deposition coverage gradually increased with increasing droplet size. However, both the droplet deposition density and coverage on the abaxial side increased first and then decreased with increasing droplet size. But there was no significant difference in droplet deposition density and coverage on the abaxial side between different droplet sizes. Among them, the droplet deposition density and coverage on the adaxial side of 100 μm and 150 μm droplet size was mostly significantly larger than that of 50 μm. When the droplet size was constant, the droplet deposition density and coverage on the adaxial side of the banana canopy leaves increased first and then decreased with increasing flight height, but there was no significant difference in the change. The droplet deposition density on the abaxial side of the banana canopy leaves increased first and then decreased with increasing flight height, however, the droplet deposition coverage on the abaxial side of 50 μm and 150 μm droplet size decreased with increasing flight height, and the droplet deposition coverage 100 μm droplet size increased first and then decreased with increasing flight height. But there was no significant difference in droplet deposition density and coverage on the abaxial side of banana canopy leaves with the change of flight height. In addition, the droplet deposition density and coverage on the adaxial side of banana canopy leaves were mostly significantly larger than that on the abaxial side.

3.2 Droplet deposition uniformity

We calculated the droplet deposition density and coverage uniformity of different operational parameter combinations based on the collected droplet deposition data. Table 6 shows the variance analysis results of the effects of flight height, droplet size and their interaction on the droplet deposition uniformity on both sides of banana canopy leaves. The results showed that the flight height, droplet size and their interaction had no significant effect on the

TABLE 6 Variance analysis of droplet deposition uniformity of banana canopy leaves.

Source	Degree of Freedom	Adaxial Side				Abaxial Side			
		Droplet Deposition Density Uniformity		Droplet Deposition Coverage Uniformity		Droplet Deposition Density Uniformity		Droplet Deposition Coverage Uniformity	
		F	P	F	P	F	P	F	P
Flight height	2	0.172	0.843	0.470	0.633	0.082	0.922	0.020	0.981
Droplet size	2	2.819	0.086	1.238	0.313	0.314	0.734	0.461	0.638
Flight height *Droplet size	4	1.081	0.395	0.746	0.574	1.070	0.400	1.202	0.344
Error	18								
Total	26								

The P and F value indicate the significance level, and P < 0.05 represents factors that have a significant effect on the test result.

droplet deposition density and coverage uniformity on both sides of banana canopy leaves ($P > 0.05$). However, the most important factor affecting the droplet deposition density and coverage uniformity on both sides of banana canopy leaves was droplet size, with F values of 2.819, 0.314 and 1.238, 0.461, respectively. The second was the flight height, with F values of 0.172, 0.082 and 0.470, 0.020, respectively.

Figure 7 shows the droplet deposition density and coverage uniformity on both sides of banana canopy leaves with different operational parameter combinations of flight height and droplet size. It can be seen that when the flight height was 4 m and the droplet size was 150 μm, the droplet deposition density and coverage uniformity on the adaxial side was the best, were 37.58% and 53.79% respectively. When the flight height was 5 m and the droplet size was 50 μm, the droplet deposition density and coverage uniformity on the abaxial side was the best, were 29.06% and 38.83%, respectively. It can be seen that the operational parameter combination that achieved the best droplet deposition uniformity might not be able to achieve the largest droplet deposition density or coverage, and the operational parameter combination that achieved the largest droplet deposition density and coverage mostly had a worse droplet deposition uniformity. We believe that this may be due to the difference in statistical calculation methods, which makes it difficult to balance the optimal droplet deposition density, coverage and uniformity. In addition, when the flight height was constant, there was mostly no significant difference in droplet deposition density and coverage uniformity on both sides of banana canopy leaves with the change of droplet size. When the droplet size was constant, there was also no significant difference in droplet deposition density and coverage uniformity on both sides of banana canopy leaves with the change of flight height. The droplet deposition density and coverage uniformity on the adaxial side of banana canopy leaves was not significantly different from that on the abaxial side.

3.3 Droplet deposition penetration

Table 7 shows the variance analysis results of the effects of flight height, droplet size and their interaction on droplet deposition density and coverage penetration on both sides of banana canopy leaves. The results showed that the flight height, droplet size and their interaction had no significant effect on the droplet deposition density and coverage penetration on both sides of banana canopy leaves ($P > 0.05$). The most important factor affecting the droplet deposition density and coverage penetration on the adaxial side of banana canopy leaves was droplet size, with F values of 3.142 and 1.504, followed by flight height with F values of 1.025 and 0.450. The most important factor affecting the droplet deposition density and coverage penetration on the abaxial side was the flight height with F values of 2.763 and 0.909, followed by the droplet size with F values of 1.527 and 0.226.

Figure 8 shows the droplet deposition density and coverage penetration on both sides of banana canopy leaves for different flight heights and droplet sizes. It can be seen that when the flight height was 4 m and the droplet size was 100 μm, the droplet deposition density and coverage penetration on the adaxial side of the banana canopy leaves was the best, which were 29.74% and 35.09%, respectively. The droplet deposition density penetration on the abaxial side of 4 m flight height and 100 μm droplet size was the best (23.51%), and the droplet deposition coverage penetration on the abaxial side of 5 m flight height and 50 μm droplet size was the best (26.98%). When the flight height was constant, the difference of droplet deposition density and coverage penetration on both sides of banana canopy leaves was mostly not significant with the change of droplet size. When the droplet size was constant, the difference of droplet deposition density and coverage penetration on both sides of banana canopy leaves was mostly not significant with the change of flight height. In addition, the droplet deposition density and coverage penetration on the adaxial side of banana canopy leaves was mostly not significantly different from that on the abaxial side.

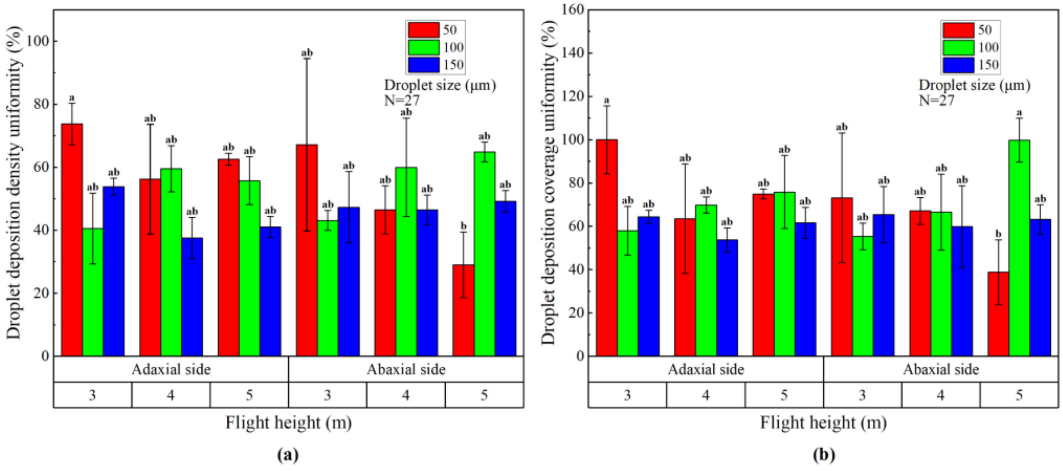


FIGURE 7
The histogram of droplet deposition density uniformity **(A)** and coverage uniformity **(B)** on banana canopy leaves. The data in this figure represent mean ± standard deviation. Columns with different lowercase letters are significantly different at the $P < 0.05$ level based on Duncan's new multiple range test.

TABLE 7 Variance analysis of droplet deposition penetration of banana canopy leaves.

Source	Degree of Freedom	Adaxial Side				Abaxial Side			
		Droplet Deposition Density Penetration		Droplet Deposition Coverage Penetration		Droplet Deposition Density Penetration		Droplet Deposition Coverage Penetration	
		F	P	F	P	F	P	F	P
Flight height	2	1.025	0.379	0.450	0.645	2.763	0.090	0.909	0.421
Droplet size	2	3.142	0.068	1.504	0.249	1.527	0.244	0.226	0.800
Flight height *Droplet size	4	1.128	0.375	1.585	0.221	1.560	0.228	1.654	0.204
Error	18								
Total	26								

The P and F value indicate the significance level, and $P < 0.05$ represents factors that have a significant effect on the test result.

3.4 Comparison of plant protection UAV and ground air-assisted sprayer

3.4.1 Droplet deposition

Table 8 shows the droplet deposition density and droplet deposition coverage on both sides of banana canopy leaves by two kinds of spraying machines, plant protection UAV and ground air-assisted sprayer. It can be seen that the droplet deposition density of the ground air-assisted sprayer on both sides of the banana canopy leaves was 1.6 times and 7.4 times that of the plant protection UAV, respectively. However, the droplet coverage of the ground air-assisted sprayer on the adaxial side of the banana canopy leaf was only 74.6% of that of the plant protection UAV. Different from the plant protection UAV, the droplet deposition density and coverage of the ground air-assisted sprayer on the abaxial side of the banana canopy leaves were obviously larger than the adaxial side.

3.4.2 Droplet drift

The ground droplet drift deposition of plant protection UAV and ground air-assisted sprayer within 20 m of the drift zone is shown in Figure 9. It can be seen that the droplet drift deposition amount of the plant protection UAV was only larger than that of the ground air-assisted sprayer before 2 m. The droplet drift deposition amount of the plant protection UAV and the ground air-assisted sprayer approaches 0 at 7.5 m and 15 m, respectively.

4 Discussion

The results of droplet deposition density showed that neither too large or too small droplet size showed a good droplet deposition density on both sides of banana canopy leaves. Smaller droplets are more likely to drift affected by environmental wind, while larger droplets are less affected by environmental wind and can reach the

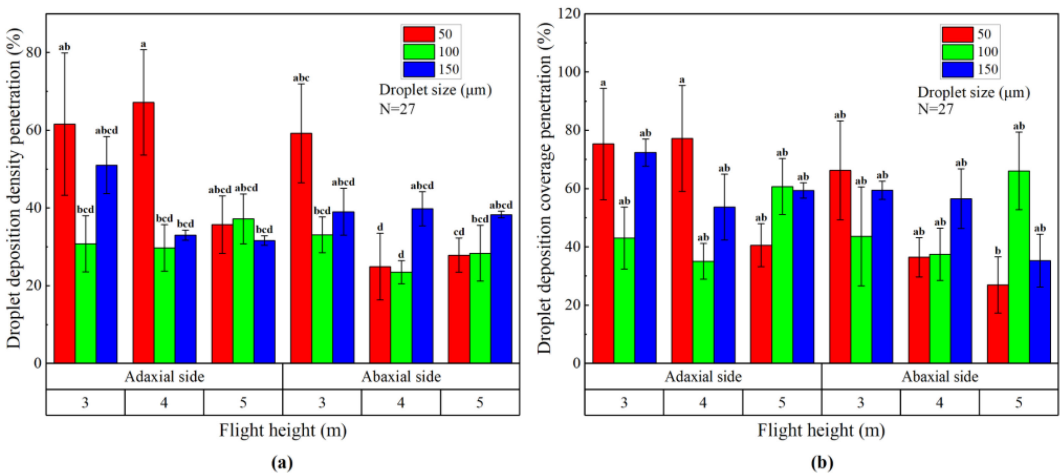


FIGURE 8
The histogram of droplet deposition density penetration **(A)** and coverage penetration **(B)** on banana canopy leaves. The data in this figure represent mean ± standard deviation. Columns with different lowercase letters are significantly different at the $P < 0.05$ level based on Duncan's new multiple range test.

TABLE 8 The droplet deposition on both sides of banana canopy leaves of plant protection UAV and ground air-assisted sprayer.

Machine Type	Adaxial Side		Abaxial Side	
	Droplet Deposition Density (/cm^2)	Droplet Deposition Coverage (%)	Droplet Deposition Density (/cm^2)	Droplet Deposition Coverage (%)
Plant protection UAV	68.21 ± 6.19	17.41 ± 1.97	17.66 ± 2.35	1.31 ± 0.35
Ground air-assisted sprayer	107.23 ± 8.42	12.99 ± 0.81	130.94 ± 3.29	20.69 ± 1.04

The data in this table represent mean ± standard deviation.

adaxial side of crop leaves under the action of self-weight, but difficult to be transported to the abaxial side by UAV downwash airflow (Chen et al., 2020). At an appropriate flight height, the appropriate downwash airflow generated by the UAV rotor can force the droplets to be well deposited (Tang et al., 2018). When the flight height was 4 m and the droplet size was 100 μm, the droplet deposition density on both sides of the banana canopy leaves was the largest, and only this operational parameter combination could meet the effective droplet deposition density index of 15 droplets per square cm on the abaxial side.

Different from the results of droplet deposition density, when the flight height was constant, the droplet deposition coverage on the adaxial side of banana canopy leaves increased with droplet size increasing from 100 μm to 150 μm. The reason is that when the application amount is constant, as the droplet size decreases, the number of droplets can increase at a geometric speed, and the probability of droplets hitting the surface of the target crop increases, further increasing the deposition density of droplets on the surface of the target crop. However, the possibility of drifting affected by airflow becomes larger, and the decrease in droplet volume, resulting in a decrease in deposition coverage (Lan et al., 2016). In addition, the droplet deposition coverage on the adaxial side of 4 m flight height and 150 μm droplet size was the largest, but its droplet deposition coverage on the abaxial side did not reach the effective droplet deposition coverage index of 1%. When the flight height was 4 m and the droplet size was 100 μm, the droplet deposition coverage on abaxial side of the banana canopy leaves was the largest and its droplet deposition coverage on the adaxial side was second only to 4 m flight height and 150 μm droplet size. Moreover, only this operational parameter combination could meet the effective droplet deposition coverage index on the abaxial side. Therefore, combined with the results of droplet deposition density we believe that the flight height of 4 m and the droplet size of 100 μm are the best operational parameters for UAV spraying in banana orchard.

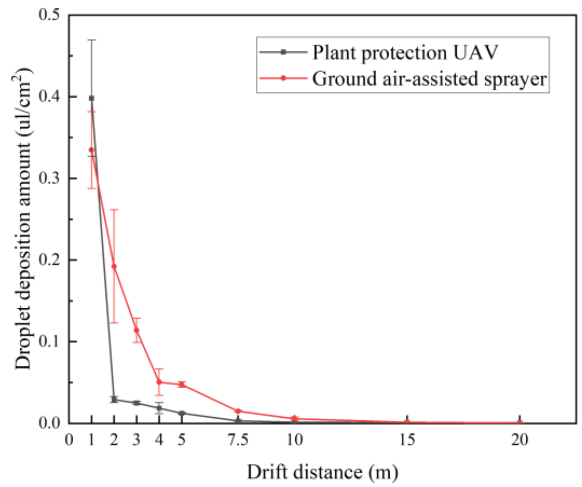


FIGURE 9
Droplet drift deposition of plant protection UAV and ground air-assisted sprayer.

The droplet deposition density and coverage on the adaxial side of banana canopy leaves was mostly significantly larger than that on the abaxial side. We believe that this is due to the special spraying method of aerial spraying makes the adaxial side of banana canopy leaves can directly collide with the spray droplets. However, the droplet deposition on the abaxial side always depends on the forced disturbance of the UAV downwash airflow to the leaves, so as to optimize the droplet transport channel and enhance the droplet's orbital movement ability to indirectly realize the interaction between the droplets and the leaves. This is also the biggest difference between UAV spray and traditional ground spraying equipment (Jiang Y. et al., 2023). In addition, the large banana leaves will be elastically deformed after being subjected to the strong wind load of the UAV downwash airflow, and the leaves on both sides of the veins tend to be parallel to the flow direction of the droplets, resulting in a decrease in the contact area. This is a typical feature of large broad-leaved herbaceous crops such as banana, and it is also the biggest difference from woody plants such as citrus and peach.

Combined with the results of droplet deposition uniformity and penetration, it can be seen that under the same operational parameters, it is difficult to achieve the best droplet deposition, droplet deposition uniformity and penetration at the same time. The droplet deposition reflects the amount of droplets received by the target crop, and the high droplet deposition indicates that the target crop accepts a large number of droplets. The droplet deposition uniformity and penetration reflect the difference of droplet distribution in the horizontal and vertical spraying area, and a large droplet deposition does not mean a good droplet deposition uniformity or penetration. In addition, during the

UAV spraying, the deformation degree and inclination degree of leaves at different sampling points was different under the influence of UAV downwash airflow. We believe that this is also a factor that leads to the inconsistency of the optimal operational parameter for droplet deposition, droplet deposition uniformity and penetration. Similar results were found in other herbaceous crops, such as rice (Guo S. et al., 2021).

The results of droplet deposition penetration showed that the droplet deposition density and coverage penetration on both sides of banana canopy leaves at flight height of 3 m was mostly less than that of 4 m and 5 m. When the flight height is low, the wind field under UAV rotor is too strong, causing the upper canopy leaves to swing and tilt around with great deformation. This not only affects the transport of droplets through the upper canopy to the adaxial side of the middle and lower canopy leaves, but also blocks the transport channel of droplets to the abaxial side of the leaves. When the droplet size was 50 μm, the droplet deposition density and coverage penetration on the adaxial side of banana canopy leaves is mostly less than 100 μm and 150 μm. It shows that the penetration ability of larger droplets may be better than that of smaller droplets within a certain range of droplet size for the plant protection UAV. This is because the downwash rotor wind field of plant protection UAV can accelerate the deposition speed of droplets, and blow the upper leaves of crop canopy at the same time, so that a large number of large size droplets with a faster falling speed can reach the lower canopy of the crop within the action time of the rotor wind field (Chen et al., 2017a; Ferguson et al., 2016; Fritz et al., 2009). There is a similar result of the research carried out by some researchers on droplet deposition of different droplet sizes in rice using an UAV (Chen et al., 2020).

Due to the different spraying methods, the ground air-assisted sprayer could obtain larger droplet deposition density and coverage on the abaxial side of banana canopy leaves than the plant protection UAV, but the plant protection UAV had larger droplet deposition coverage on the adaxial side of banana canopy leaves. This also caused the drift droplets of the plant protection UAV to deposit in a large amount in the drift zone near the banana plant, while the droplet drift distance of the ground air-assisted sprayer was farther. Therefore, the application of ground air-assisted sprayers in banana orchards requires a larger drift isolation belt to avoid the harm of pesticide droplet drift to non-target organisms. In addition, the droplet deposition density and coverage on both sides of the banana canopy leaves by the ground air-assisted sprayer were obviously smaller than those of the banana leaves with the petiole fixed on the horizontal beam in our previous study (Jiang Y. et al., 2023). We believe that this is due to the airflow generated by the sprayer fan aggravates the mutual occlusion between the canopy leaves.

Although the results of this study were of great significance for the use of the plant protection UAV in the droplet deposition distribution characteristics on banana plants, the banana morphological structure varies with growth stages and varieties. However, the crop canopy structure directly affects the setting of operational parameters of plant protection UAV (Leandro Da Vitória et al., 2022; Liu et al., 2022; Mahmud et al., 2023; Sun et al., 2022). Further studies are needed to clarify new operational parameters, such as new spraying amounts, flight models, and flight speeds, in order to increase the UAV aerial application efficiency and effectiveness in pests and diseases control in banana orchard. Meanwhile, the results of this study were also of great significance to the air-ground cooperation spraying of plant protection UAV and ground air-assisted sprayer in banana orchards. The two machines work together successively, and the spraying process does not interfere with each other, giving full play to the excellent deposition performance of the ground air-assisted sprayer on the abaxial side of the banana canopy leaves, and then the plant protection UAV supplements the disadvantage of insufficient deposition on the adaxial side of the ground air-assisted sprayer to achieve uniform coverage of banana canopy plant protection operations. This is also our future research focus.

5 Conclusions

In this work, aerial spraying tests at different flight heights and droplet sizes were carried out in the banana orchard using an UAV, and the droplet deposition distribution on both sides of banana canopy leaves was studied and analyzed. The main conclusions are shown as follows.

1. The droplet deposition density, coverage, uniformity and penetration of banana canopy leaves were affected by the flight height and droplet size, and the droplet size was the main influencing factor. In addition, different from the ground spraying equipment, the droplet deposition density and coverage on the adaxial side of banana canopy leaves was generally significantly larger than that on the abaxial side during the UAV spraying.
2. The operational parameter combination to achieve the best droplet deposition density and penetration on both sides of banana canopy leaves was the flight height of 4 m, the droplet size of 100 μm, the droplet deposition density and penetration were 63.77 droplets per square cm, 17.46 droplets per square cm and 29.74%, 23.51%, respectively. However, the droplet deposition density uniformity on both sides was worse, which were 59.56% and 60.01% respectively. This operational parameter combination could meet the effective droplet deposition density index of 15 droplets per square cm on the abaxial side of banana canopy leaves. The operational parameter combinations achieved the best droplet deposition density uniformity on both sides of the banana canopy were the flight height of 4 m, the droplet size of 150 μm and the flight height of 5 m, the droplet size of 50 μm, respectively. The best droplet deposition uniformity were 37.58% and 29.06%, respectively. However, the droplet deposition density and penetration on both sides were worse, which were 56.04 droplets per square cm, 24.43 droplets per square cm and 33.02%, 27.91% on the adaxial side, and 5.64 droplets per square cm, 3.86 droplets per square cm and 35.76%, 39.86%

on the abaxial side, respectively. Using these two optimal parameter combinations, the droplets could not be effectively deposited on the abaxial side of the banana canopy leaves.

3. Using droplet deposition coverage as an evaluation indicator. The operational parameter combinations achieved the best droplet deposition coverage, uniformity and penetration on the adaxial side of the banana canopy were 4 m flight height and 150 μm droplet size, 4 m flight height and 150 μm droplet size, and 4 m flight height and 100 μm droplet size, were 14.33%, 53.79% and 35.09%, respectively. However, these operational parameter combinations could not reach the best result on the abaxial side, were 0.33%, 59.93%, 37.49%. On the abaxial side, the operational parameter combinations achieved the best droplet deposition coverage, uniformity and penetration were 4 m flight height and 100 μm droplet size, 5 m flight height and 50 μm droplet size, 5 m flight height and 50 μm droplet size, were 1.24%, 38.83%, 26.98%. However, these operational parameter combinations could not reach the best result on the adaxial side, were 12.75%, 75.00%, 40.58%. Only the operational parameter combination of 4 m flight height and 100 μm droplet size could reach the effective droplet deposition coverage index of 1% on the abaxial side.
4. Compared with the plant protection UAV, the ground air-assisted sprayer had higher droplet deposition density and coverage on the abaxial side of banana canopy leaves. However, the droplet deposition coverage on the adaxial side was smaller. The droplet deposition density and coverage on the abaxial side of banana canopy leaves were obviously greater than the adaxial side during the spraying of ground air-assisted sprayer. In addition, the droplet drift distance of the ground air-assisted sprayer was farther than the plant protection UAV.

According to the effective deposition evaluation index of 15 droplets per square cm droplet deposition density and 1% droplet deposition coverage, we believe that the operational parameter combination of flight height of 4 m and droplet size of 100 μm can be used to obtain a better droplet deposition distribution effect on banana canopy leaves without considering the drift risk. Based on the results of this work, our study is still ongoing and further field trials are planned to determine the optimal operational parameters for banana plants at different growth stages. In addition, there were repeated spraying tests in this study, solvent was used instead of pesticide to avoid phytotoxicity. In order to further verify the reliability of plant protection UAV application technology in the banana orchard, it is also necessary to conduct field trials to study the control effect of diseases and pests. And the cooperation spraying strategy of plant protection UAV and ground air-assisted sprayer will be further applied to banana orchards to achieve full-scale droplet coverage of banana canopy. The results of this study provide a valuable reference for the adjustment of the operational parameters of plant protection UAV in the banana orchard and the implementation of air-ground cooperation spraying strategy in banana orchard.

Data availability statement

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

Author contributions

JY: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. XX: Funding acquisition, Investigation, Validation, Writing – review & editing. JD: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. YJ: Funding acquisition, Investigation, Writing – review & editing. HY: Investigation, Writing – review & editing. HL: Writing – review & editing, Investigation. SJ: Writing – review & editing, Investigation. ZY: Writing – review & editing, Funding acquisition, Project administration, Resources, Supervision.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Laboratory of Lingnan Modern Agriculture Project (Grant No. NT2021009), China Agriculture Research System of MOF and MARA (Grant No. CARS-31-11), the Guangdong Basic and Applied Basic Research Foundation (2023A1515110797), the Key-Area Research and Development Program of Guangzhou (Grant No. 2024B03J1355).

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

ASAE (2018). *Procedure for Measuring Drift Deposits from Ground, Orchard, and Aerial Sprayers, ANSI/ASAE S561.1 APR2004 (R2018ED)* (St. Joseph, Michigan: ASABE). Available at: https://elibrary.asabe.org/abstract.asp?aid=44207 (Accessed March 30, 2023).

ASAE (2020). *Spray nozzle classification by droplet spectra, ANSI/ASAE S572.3 (FEB2020)* (St. Joseph, Michigan: ASABE). Available at: https://elibrary.asabe.org/abstract.asp?aid=51101 (Accessed March 30, 2023).

Barraza, D., Jansen, K., Wesseling, C., and van Wendel De Joode, B. (2020). Pesticide risk perceptions among bystanders of aerial spraying on bananas in Costa Rica. *Environ. Res.* 189, 109877. doi: 10.1016/j.envres.2020.109877

Barua, D. K. (2005). "Beaufort wind scale," in *Encyclopedia of Coastal Science*. Ed. M. L. Schwartz (Springer, Dordrecht), 117–215. Available at: https://link.springer.com/referenceworkentry/10.1007/1-4020-3880-1_45 (Accessed November 6, 2023).

Biglia, A., Grella, M., Bloise, N., Comba, L., Mozzanini, E., Sopegno, A., et al. (2022). Uav-spray application in vine-yards: flight modes and spray system adjustment effects on canopy deposit, coverage, and off-target losses. *Sci. Total Environ.* 845, 157292. doi: 10.1016/j.scitotenv.2022.157292

Braekman, P., Foque, D., Messens, W., Van Labeke, M., Pieters, J. G., and Nuyttens, D. (2009). Effect of spray application technique on spray deposition in greenhouse strawberries and tomatoes. *Pest Manage. Sci.* 66, 203–212. doi: 10.1002/ps.1858

Cerruto, E., Manetto, G., Longo, D., Failla, S., and Papa, R. (2019). A model to estimate the spray deposit by simulated water sensitive papers. *Crop Prot.* 124, 104861. doi: 10.1016/j.cropro.2019.104861

Chen, P., Douzals, J. P., Lan, Y., Cotteux, E., Delpuech, X., Pouxviel, G., et al. (2022). Characteristics of unmanned aerial spraying systems and related spray drift: a review. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.870956

Chen, S., Lan, Y., Bradley K, F., Li, J., Liu, A., and Mao, Y. (2017a). Effect of wind field below rotor on distribution of aerial spraying droplet deposition by using multi-rotor UAV. *Trans. Chin. Soc. Agric. Machinery* 48, 105–113. doi: 10.6041/j.issn.1000-1298.2017.08.011

Chen, S., Lan, Y., Zhou, Z., Liao, J., and Zhu, Q. (2017b). Effects of spraying parameters of small plant protection UAV on droplets deposition distribution in citrus canopy. *J. South China Agric. Univ.* 38, 97–102. doi: 10.7671/j.issn.1001-411X.2017.05.017

Chen, S., Lan, Y., Zhou, Z., Ouyang, F., Wang, G., Huang, X., et al. (2020). Effect of droplet size parameters on droplet deposition and drift of aerial spraying by using plant protection uav. *Agronomy* 10, 195. doi: 10.3390/agronomy10020195

Constantinides, L. N., and Jr. Mchugh, J. J. (2003). *Pest management strategic plan for banana production in hawaii* (Hono-lulu, Hawaii). Available at: https://ipmdata.ipmcenters.org/documents/pmsps/HIBananaPMSP.pdf (Accessed October 3, 2023).

Czaczyk, Z. (2012). Spray classification for selected flat fan nozzles. *J. Plant Prot. Res.* 52, 180–183. doi: 10.2478/v10045-012-0027-2

Ding, W., Zhu, F., Jin, M., Xu, L., and Zhang, Y. (2022). Simulation and visualization of spraying droplets behavior and deposition within virtual rice canopy. *Int. J. Agric. Biol. Eng.* 15, 19–27. doi: 10.25165/j.ijabe.20221505.7099

Ferguson, J. C., Chechetto, R. G., Hewitt, A. J., Chauhan, B. S., Adkins, S. W., Kruger, G. R., et al. (2016). Assessing the deposition and canopy penetration of nozzles with different spray qualities in an oat (Avena sativa L.) canopy. *Crop Prot.* 81, 14–19. doi: 10.1016/j.cropro.2015.11.013

Fox, R. D., Derksen, R. C., Cooper, J. A., Krause, C. R., and Ozkan, H. E. (2003). Visual and image system measurement of spray deposits using water–sensitive paper. *Appl. Eng. Agric.* 19, 549–552. doi: 10.13031/2013.15315

Fritz, B. K., Hoffmann, W. C., and Lan, Y. (2009). Evaluation of the EPA drift reduction technology (DRT) low-speed wind tunnel protocol. *J. ASTM Int.* 6, 102129. doi: 10.1520/JAI102129

Guo, J., Duan, J., Li, J., and Yang, Z. (2020). Mechanized technology research and equipment application of banana post-harvesting: a review. *Agronomy* 10, 374. doi: 10.3390/agronomy10030374

Guo, S., Li, J., Yao, W., Hu, X., Wei, X., Long, B., et al. (2021). Optimization of the factors affecting droplet deposition in rice fields by rotary unmanned aerial vehicles (UAVs). *Precis. Agric.* 22, 1918–1935. doi: 10.1007/s11119-021-09818-7

Guo, X., Xue, X., Qin, W., and Liu, X. (2021). Effect of operation parameters of plant protection UAV on droplet deposition of palm tree. *J. Chin. Agric. Mechanization* 42, 35–40. doi: 10.13733/j.jcam.issn.2095-5553.2021.06.06

Huang, L., and Wei, S. (2017). *High-quality production technology of baodao banana. 1st ed.* (Beijing: China Agricultural Science and Technology Press). Available at: http://find.nlc.cn/search/showDocDetails?docId=-2532135815191829362&dataSource=ucs01 (Accessed April 24, 2024).

Jiang, Y., Duan, J., Xu, X., Ding, Y., Li, Y., and Yang, Z. (2022). Measurement of the banana pseudo-stem phenotypic parameters based on ellipse model. *Int. J. Agric. Biol. Eng.* 15, 195–202. doi: 10.25165/j.ijabe.20221503.6614

Jiang, T., Duan, J., Zhang, Z., Xie, B., and Yang, Z. (2023). Performance matching of common pesticides in banana plantations on the surface of banana leaves at different growth stages. *Pest Manage. Sci.* 79, 5116–5129. doi: 10.1002/ps.7713

Jiang, Y., Yang, Z., Xu, X., Shen, D., Jiang, T., Xie, B., et al. (2023). Wetting and deposition characteristics of air-assisted spray droplet on large broad-leaved crop canopy. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1079703

Jing, T., Xie, J., and Zhou, D. (2022). *Color Image for banana cultivation and disease-insect pest control. 1st ed.* (Beijing: China Agriculture Press). Available at: https://find.nlc.cn/search/showDocDetails?docId=8945314841193110 7&dataSource=ucs01 (Accessed April 24, 2024).

Lan, Y., and Chen, S. (2018). Current status and trends of plant protection UAV and its spraying technology in China. *Int. J. Precis. Agric. Aviation.* 1, 1–9. doi: 10.33440/j.ijpaa.20180101.0002

Lan, Y., Peng, J., and Jin, J. (2016). Research status and development of pesticide spraying droplet size. *J. South China Agric. University.* 37, 1–9. doi: 10.7671/i.issn.1001-411X.2016.06.001

Leandro Da Vitória, E., Furtado De Oliveira, R., Hoffmam Crause, D., and Oliveira Ribeiro, L. F. (2022). Effect of flight operative height and genotypes on conilon coffee spraying using an unmanned aerial vehicle. *Coffee Sci.* 17, 1–9. doi: 10.25186/.v17i.2003

Li, L., Liu, Y., He, X., Song, J., Zeng, A., Wang, Z., et al. (2018). *Assessment of spray deposition and losses in the ap-ple orchard from agricultural unmanned aerial vehicle in China 2018 ASABE Annual International Meeting* (Detroit, Michigan: ASABE), 1800504. doi: 10.13031/aim.201800504

Liao, J., Zang, Y., Zhou, Z., and Luo, X. (2015). Quality evaluation method and optimization of operating parameters in crop aerial spraying technology. *Trans. Chin. Soc. Agric. Eng.* 31, 38–46. doi: 10.11975/j.issn.1002-6819.2015.z2.006

Liu, Y., Li, L., Liu, Y., He, X., Song, J., Zeng, A., et al. (2020). Assessment of spray deposition and losses in an apple orchard with an unmanned agricultural aircraft system in China. *Trans. Asabe* 63, 619–627. doi: 10.13031/trans.13233

Liu, Y., Xiao, Q., Han, X., Zeeshan, M., Fang, Z., and Dou, Z. (2022). Effect of aerial application of adjuvants on pepper defoliant droplet deposition and efficacy of defoliation sprayed by unmanned aerial vehicles. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.917462

Mahmud, M. S., He, L., Heinemann, P., Choi, D., and Zhu, H. (2023). Unmanned aerial vehicle based tree canopy characteristics measurement for precision spray applications. *Smart Agric. Technol.* 4, 100153. doi: 10.1016/j.atech.2022.100153

Matthews, G. (2009). Aerial spraying of bananas in dispute. *Outlooks On Pest Manage.* 20, 209–211. doi: 10.1564/20oct05

Meng, Y., Su, J., Song, J., Chen, W., and Lan, Y. (2020). Experimental evaluation of uav spraying for peach trees of different shapes: effects of operational parameters on droplet distribution. *Comput. Electron. Agric.* 170, 105282. doi: 10.1016/j.compag.2020.105282

Ministry of Agriculture of the PRC (2018). *Technical specification of quality evaluation for crop protection uas, NYT3213* (Beijing, China: China Agriculture Press). Available at: http://www.aqsc.org/zlbz/bzgg/202312/P020240219401819200966.pdf (Accessed April 11, 2024).

Miranda-Fuentes, A., Llorens, J., Rodríguez-Lizana, A., Cuenca, A., Gil, E., Blanco-Roldán, G. L., et al. (2016). Assessing the optimal liquid volume to be sprayed on isolated olive trees according to their canopy volumes. *Sci. Total Environ.* 568, 296–305. doi: 10.1016/j.scitotenv.2016.06.013

Nascimento, V. P. D., and Vitória, E. L. D. (2022). Spraying quality using unmanned aerial vehicle in citrus. *Rev. Engenharia na Agricultura - REVENG* 30, 214–221. doi: 10.13083/reveng.v30i1.13700

Nuyttens, D., Windey, S., and Sonck, B. (2004). Optimisation of a vertical spray boom for greenhouse spray applications. *Biosyst. Eng.* 89, 417–423. doi: 10.1016/j.biosystemseng.2004.08.016

OECD (2021). *Report on the state of the knowledge - literature review on unmanned aerial spray systems in agriculture* (Paris: OECD Publishing). Available at: https://www.oecd.org/chemicalsafety/pesticides-biocides/literature-review-on-unmanned-aerial-spray-systems-in-agriculture.pdf (Accessed April 11, 2024).

Pandiselvam, R., Daliyamol,, Imran, S. S., Hegde, V., Sujithra, M., Prathibha, P. S., et al. (2024). Evaluation of unmanned aerial vehicle for effective spraying application in coconut plantations. *Heliyon.* 10, e38569. doi: 10.1016/j.heliyon.2024.e38569

Qin, W., Qiu, B., Xue, X., Chen, C., Xu, Z., and Zhou, Q. (2016). Droplet deposition and control effect of insecticides sprayed with an unmanned aerial vehicle against plant hoppers. *Crop Prot.* 85, 79–88. doi: 10.1016/j.cropro.2016.03.018

Robson, J. D., Wright, M. G., and Almeida, R. P. P. (2007). Effect of imidacloprid foliar treatment and banana leaf age on Pentalonia nigronervosa (Hemiptera, Aphididae) survival. *N. Z. J. Crop Hortic. Sci.* 35, 415–422. doi: 10.1080/01140670709510209

Salyani, M., and Fox, R. D. (1999). Evaluation of spray quality by oiland water-sensitive papers. *Trans. ASAE.* 42, 37–43. doi: 10.13031/2013.13206

Sinha, J. P. (2020). Aerial robot for smart farming and enhancing farmers' net benefit. *Indian J. Agric. Sci.* 90, 258–267. doi: 10.56093/ijas.v90i2.98997

Sun, T., Zhang, S., Xue, X., and Jiao, Y. (2022). Comparison of droplet distribution and control effect of wheat aphids un-der different operation parameters of the crop protection UAV in the wheat flowering stage. *Agronomy* 12, 3175. doi: 10.3390/agronomy12123175

Tang, Y., Hou, C. J., Luo, S. M., Lin, J. T., Yang, Z., and Huang, W. F. (2018). Effects of operation height and tree shape on droplet deposition in citrus trees using an unmanned aerial vehicle. *Comput. Electron. Agric.* 148, 1–7. doi: 10.1016/j.compag.2018.02.026

TOPPS-Prowadis Project (2014). Best management practices to reduce spray drift. Available online at: http://www.topps-life.org/ (Accessed April 25, 2024).

Verma, A., Singh, M., Parmar, R. P., and Bhullar, K. S. (2022). Feasibility study on hexacopter UAV based sprayer for application of environment-friendly biopesticide in guava orchard. *J. Environ. Biol.* 43, 97–104. doi: 10.22438/jeb/43/1/MRN-1912

Wang, G., Han, Y., Li, X., Andaloro, J., Chen, P., Hoffmann, W. C., et al. (2020). Field evaluation of spray drift and environmental impact using an agricultural unmanned aerial vehicle (uav) sprayer. *Sci. Total Environ.* 737, 139793. doi: 10.1016/j.scitotenv.2020.139793

Wang, X., He, X., Song, J., and Herbst, A. (2015). Effect of adjuvant types and concentration on spray drift potential of different nozzles. *Trans. Chin. Soc. Agric. Eng.* 31, 49–55. doi: 10.11975/j.issn.1002-6819.2015.22.007

Wang, J., Lan, Y., Wen, S., Hewitt, A. J., Yao, W., and Chen, P. (2019a). Meteorological and flight altitude effects on deposition, penetration, and drift in pineapple aerial spraying. *Asia-Pac. J. Chem. Eng.* 15, e2382. doi: 10.1002/apj.2382

Wang, J., Lan, Y., Yao, W., Chen, P., Lin, J., and Yan, Y. (2019b). Effects of working height of single-rotor unmanned aerial vehicle on drift and droplets deposition distribution of areca tree. *Trans. Chin. Soc. Agric. Machinery* 50, 109–119. doi: 10.6041/j.issn.1000-1298.2019.07.011

Wang, J., Lan, Y., Yao, W., Chen, P., Wang, G., and Chen, S. (2020). Aerial spraying application of multi-rotor unmanned aerial vehicle on areca trees. *Int. J. Precis. Agric. Aviation* 3, 51–64. doi: 10.33440/j.ijpaa.20200304.134

Wang, C., Liu, Y., Zhang, Z., Han, L., Li, Y., Zhang, H., et al. (2022). Spray performance evaluation of a six-rotor unmanned aerial vehicle sprayer for pesticide application using an orchard operation mode in apple orchards. *Pest Manage. science.* 78, 2449–2466. doi: 10.1002/ps.6875

Wang, J., Ma, C., Chen, P., Yao, W., Yan, Y., Zeng, T., et al. (2023). Evaluation of aerial spraying application of multi-rotor unmanned aerial vehicle for areca catechu protection. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1093912

Xie, J. (2019). Fruit scientific research in New China in the past 70 years: Banana. *J. Fruit Sci.* 36, 1429–1440. doi: 10.13925/j.cnki.gsxb.Z17

Xie, M. (2024). Analysis of Fruit Tree Canopy with Millimeter-Wave Rader for Targeted Pesticide Application. South China Agricultural University, Guangzhou.

Xu, X., and Liu, E. (2021). China 's banana market and industry survey analysis report. *Agric. Products Market* 12, 44–46. Available at: https://kns.cnki.net/kcms/detail/detail.aspx?FileName=NCPS202112020&DbName=CJFN2021 (Accessed April 25, 2024).

Xue, X., Fu, X., and Gu, W. (2008). New development of plant protection machinery technology. *Jiangsu Agric. Mechanization* 06, 19–21. doi: 10.16271/j.cnki.jsnjh.2008.06.007

Yang, Z., Yu, J., Duan, J., Xu, X., and Huang, G. (2023). Optimization-design and atomization-performance study of aerial dual-atomization centrifugal atomizer. *Agriculture* 13, 430. doi: 10.3390/agriculture13020430

Yang, Z., Yu, J., Xu, X., Duan, J., Ou, Y., Shen, D., et al. (2021). A centrifugal atomization nozzle. Available online at: https://kns.cnki.net/kcms/detail/detail.aspx?FileName=CN113385316A&DbName=SCPD2021 (Accessed February 20, 2023).

Yao, W., Lan, Y., Wang, J., Wen, S., Wang, G., and Chen, S. (2017). Droplet drift characteristics of aerial spraying of AS350B3e helicopter. *Trans. Chin. Soc. Agric. Eng.* 33, 75–83. doi: 10.11975/j.issn.1002-6819.2017.22.010

Zhang, R., Liu, Y., Zhu, D., Wu, P., Xia, W., Zheng, C., et al. (2024). Comparative analysis of micro-physical characteristics of sprayed droplets using various measurement technologies. *Irrigation science.* 42, 701–719. doi: 10.1007/s00271-024-00920-y

Zhang, W., Zhu, X., Chen, C., Zeng, H., Jiang, X., Wu, J., et al. (2021). Large broad-leaved canopy of banana (Musa nana Lour.) Induces dramatically high spatial-temporal variability of throughfall. *Hydrol. Res.* 52, 1223–1238. doi: 10.2166/nh.2021.023

Zhou, L., and He, Y. (2016). *Simulation and optimization of multi spray factors in UAV 2016 ASABE Annual International Meeting* (Orlando, Florida: ASABE), 162460907. doi: 10.13031/aim.20162460907

Zhou, Z., Yuan, W., and Chen, S. (2014). Current status and future directions of rice plant protection machinery in China. *Guangdong Agric. Sci.* 41, 178–183. doi: 10.16768/j.issn.1004-874x.2014.15.021

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 27 February 2025
DOI 10.3389/fpls.2025.1510663

Check for updates

OPEN ACCESS

EDITED BY
Jingcheng Zhang,
Hangzhou Dianzi University, China

REVIEWED BY
Elio Romano,
Centro di ricerca per l'Ingegneria e le Trasformazioni agroalimentari (CREA-IT), Italy
Hongbo Qiao,
Henan Agricultural University, China

*CORRESPONDENCE
Haiguang Wang
wanghaiguang@cau.edu.cn

RECEIVED 13 October 2024
ACCEPTED 03 February 2025
PUBLISHED 27 February 2025

CITATION
Qin F, Wang H, Jiang Q and Wang H (2025) A novel method based on lesion expansion to assess plant disease severity.
Front. Plant Sci. 16:1510663.
doi: 10.3389/fpls.2025.1510663

COPYRIGHT
© 2025 Qin, Wang, Jiang and Wang. 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 novel method based on lesion expansion to assess plant disease severity

Feng Qin, Haiguang Wang*, Qian Jiang and Hongli Wang

College of Plant Protection, China Agricultural University, Beijing, China

Introduction: Severity is a key indicator utilized in plant disease monitoring and pathogen-plant interaction phenotyping.

Methods: A plant disease severity assessment method based on lesion expansion was proposed in this study to more accurately and quickly assess the severity of plant diseases for which the lesion area ratio of an investigated plant unit at each severity class in the corresponding severity grading standard is not the actual ratio of the lesion area to the area of the whole investigated plant unit. By taking wheat stripe rust caused by *Puccinia striiformis* f. sp. *tritici* as an example, after image segmentation operations of single diseased wheat leaves with wheat stripe rust, lesion expansion processing was carried out using nine method combinations of three proposed lesion expansion methods and three proposed lesion expansion coefficient determination methods, and then the severity assessments of single diseased wheat leaves were conducted.

Results: The results showed that the accuracy of severity assessments of single diseased wheat leaves in each severity class was in the range of 78.00% to 100.00%. No matter which method was used to determine the lesion expansion coefficient/coefficients, the performance of the severity assessments of the single diseased leaves achieved after lesion expansion using lesion expansion method 3 (the lesion expansion method based on an image scaling algorithm) outperformed that achieved after lesion expansion using the other two lesion expansion methods. The performance of the method combination of lesion expansion method 3 and lesion expansion coefficient determination method 1 with a lesion expansion coefficient of 2.74, achieving an accuracy of 96.16% for severity assessments of all the single diseased wheat leaves, was the optimal method among the nine method combinations.

Discussion: The results demonstrated that satisfactory severity assessment results could be achieved using the proposed method based on lesion expansion. The results indicated that the lesion-expansion-based plant disease severity assessment method is feasible, and can be used to solve the severity assessment problem described above. This study provided a new idea and method for accurate severity assessment of plant diseases and provided support for the automatic and intelligent assessment of plant disease severity.

KEYWORDS

plant disease, severity, disease assessment, image processing, lesion expansion, expansion coefficient, wheat stripe rust

1 Introduction

Plant diseases can reduce the yield and quality of plant products and affect food security worldwide. Some pathogens can produce various toxins when they infect host plants and cause diseases, resulting in serious harm to humans and livestock. Furthermore, some plant diseases can have serious impacts on ecological environments. Therefore, it is of great significance to strengthen the scientific management of plant diseases. Carrying out disease investigations and obtaining information on the occurrence of plant diseases is the basis and premise for the scientific and effective management of diseases. Severity is a key indicator used in plant disease investigations, plant disease monitoring, and diseased plant phenotyping, and the severity assessment is critical to understand the disease intensity, explore the pathogen-plant interaction, and implement disease management (Bock et al., 2022a; Jiang et al., 2022, 2023). The severity assessments of plant diseases mainly rely on the naked-eye observations (visual observations) of raters or assessors, but under laboratory conditions, the graph paper method and paper-weighing method can sometimes be used for disease severity assessments (Li et al., 2011). With the development of science and technology, the methods of plant disease investigation and monitoring supported by information technology have been developing rapidly (Wang, 2022, 2023). There are more and more applications of information technology in plant disease severity assessments, mainly including image processing technology (Li et al., 2011, 2013a; Barbedo, 2014; Pethybridge and Nelson, 2015; Shrivastava et al., 2015; Jiang et al., 2021; Olivoto et al., 2022; Yao et al., 2024), remote sensing technology (Huang et al., 2004; Wang et al., 2007, 2012, 2016; Zhao et al., 2014), and near-infrared spectroscopy technology (Li et al., 2015). In particular, image processing technology is more and more widely used in plant disease severity assessments (Li et al., 2011, 2013a; Barbedo, 2014; Pethybridge and Nelson, 2015; Shrivastava et al., 2015; Jiang et al., 2021; Olivoto et al., 2022; Yao et al., 2024). The applications of image processing technology, remote sensing technology, near-infrared spectroscopy technology, and other information technologies, provide strong support for more convenient and accurate severity assessments of plant diseases.

The severity assessments of plant diseases should be carried out in strict accordance with the severity grading standards of corresponding diseases. Severity is used to describe the disease intensity or infection degree on a plant unit (such as a leaf, fruit, stem, or plant), and in many plant disease severity grading standards, the severity classes are divided based on the ratios of the lesion areas to the areas of the investigated plant units (Nutter et al., 1991; Bock et al., 2022b). However, for some diseases, such as wheat stripe (yellow) rust caused by *Puccinia striiformis* f. sp. *tritici* (*Pst*) and wheat leaf rust caused by *P. triticina* (*Pt*), in their severity grading standards, the ratio of the lesion area to the area of an investigated plant unit corresponding to each severity class is not the actual ratio of the lesion area to the area of the investigated plant unit, which brings difficulties to the accurate severity assessments of these diseases. During the severity assessments of these diseases, it is easy to make errors in the assessment of the severity classes. For instance, according to the National Standard of the People's Republic of China's (GB/T 15795–2011) "Rules for Monitoring and Forecast of the Wheat Stripe Rust (*Puccinia striiformis* West.)", based on the percentages of the lesion areas in the whole areas of the corresponding single diseased wheat leaves, eight severity classes including 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100% are classified for wheat stripe rust; if the percentage of the lesion area in the whole area of a single diseased wheat leaf infected with *Pst* is between two adjacent severity classes, the severity of this diseased wheat leaf is regarded as the nearest severity class; and if the percentage of the lesion area in the whole area of a single diseased wheat leaf infected with *Pst* is less than 1%, the severity of this diseased wheat leaf is recorded as the severity class of 1%. However, for each severity class of wheat stripe rust, the actual percentages of the lesion areas in the areas of the whole diseased wheat leaves are much lower than the lesion area percentage corresponding to the severity class in the severity grading standard of the disease, and even when the severity class of a single diseased wheat leaf infected with *Pst* is 100%, the diseased leaf is still not completely covered with the uredinia produced by *Pst*, and the actual percentage of the lesion area to the whole area of the single diseased leaf is much lower than 100% (Shang et al., 1990; Jiang et al., 2022).

The naked-eye observation method, namely the visual observation method, which is widely used in the severity assessments of plant diseases in practice, is time-consuming and laborious and is susceptible to the subjective experience of raters or assessors, and prone to resulting in assessment errors. To ensure the reliability of the assessment results, raters or assessors should be trained and familiar with the disease severity grading method before they can carry out the severity assessments using the naked-eye observation method. Both the graph paper and the paper-weighing methods are time-consuming, labor-intensive, complicated to carry out, and rarely used in production practice. The plant disease severity assessment methods based on remote sensing technology or near-infrared spectroscopy have a strong dependence on instruments with relatively high prices, thus, the methods are mainly used in research and still have a long way to go before being widely used in production practice. At present, when using image processing technology to assess the severity of a plant disease, a main research idea is to compare the actual ratio of the lesion area to the area of an investigated plant unit with the ratios of lesion areas to areas of the plant units corresponding to the severity classes in the severity grading standard of the disease and then to determine the severity class of the investigated plant unit (Li et al., 2011, 2013a; Pethybridge and Nelson, 2015; Shrivastava et al., 2015; Jiang et al., 2021; Olivoto et al., 2022; Yao et al., 2024). Thus, severity assessment errors can be induced for plant diseases with the severity grading standards in which the ratio of the lesion area to the area of an investigated plant unit corresponding to each severity class is not the actual ratio of the lesion area to the area of the investigated plant unit, exerting great effects on the application of the plant disease severity assessment method based on image processing technology. Therefore, it is of great significance to solve the problem caused by the inconsistency between the actual ratio of the lesion area to the

area of an investigated plant unit and the lesion area ratio corresponding to the severity class in the plant disease severity grading standard, which is crucial to the accurate assessment of plant disease severity and is very important for the realization of automatic assessment of plant disease severity. Jiang et al. (2022) conducted image segmentation of wheat stripe rust and obtained the actual percentages of the lesion areas in the areas of the corresponding whole single diseased wheat leaves using image processing technology, and then proposed two severity assessment methods for wheat stripe rust based on the actual percentages of lesion areas, i.e., the severity assessment methods based on the midpoint-of-two-adjacent-means-based actual percentage reference range and the 90%, 95%, and 99% reference ranges of the actual percentages of lesion areas for each severity class, respectively. To realize the automatic severity assessments of wheat stripe rust based on the actual percentage of lesion areas, Jiang et al. (2024) built an automatic grading system of wheat stripe rust severity using an automatic image segmentation method and a severity assessment method based on the 99% reference range of the actual percentage of lesion areas for each severity class of wheat stripe rust. Furthermore, Jiang et al. (2023) proposed two severity assessment methods for wheat stripe rust based on the actual percentage of lesion areas using supervised learning and unsupervised learning, respectively. These methods proposed by Jiang et al. (2022, 2023, 2024) have provided good solutions to solve the problem described above.

In essence, the issue with the fact that the ratio of the lesion area to the area of an investigated plant unit corresponding to each severity class in the plant disease severity grading standard is not the actual ratio of the lesion area to the area of the investigated plant unit is the fact that the latter is much lower than the former. To better solve the problem regarding the inconsistency between the actual ratio of the lesion area to the area of an investigated plant unit and the lesion area ratio corresponding to the severity class in the plant disease severity grading standard, which can induce severity assessment errors, a severity assessment method based on lesion expansion for plant diseases was proposed in this study. In this study, wheat stripe rust was taken as the research object. In the images of each single diseased wheat leaf, using image processing technology, the segmentation of lesion/lesions was performed, the expansion of the segmented lesion/lesions was carried out to obtain new lesion/lesions, and then the area/pixel number of new lesion/lesions and the area/pixel number of the single diseased wheat leaf were calculated. Subsequently, the actual percentage of the lesion area in the area of the whole diseased wheat leaf was calculated. By comparing the actual percentage of the lesion area in the area of the whole diseased wheat leaf with the percentages in the eight severity classes in the disease severity grading standard, the severity of the single diseased wheat leaf was determined, finally. Using the proposed method, a severity assessment of single diseased wheat leaves with wheat stripe rust can be implemented. This study provides a novel method for the severity assessment of wheat stripe rust and can also provide a methodology and ideas for severity assessments of other similar plant diseases. Furthermore, a basis will be provided for the accurate and automatic severity assessment of plant diseases.

2 Materials and methods

2.1 Workflow for assessing plant disease severity based on lesion expansion

In this study, a severity assessment method for plant diseases based on lesion expansion using image processing technology was proposed, and the workflow to implement the method is shown in Figure 1, mainly including the steps as follows.

1) Image acquisition of individual diseased plant units of each severity class

Various image acquisition devices such as digital cameras, mobile phones, or other devices with camera functions can be used to acquire images of individual diseased plant units of each severity class of plant disease.

2) Disease image segmentation

Image processing software is used to manually or automatically segment the diseased plant units from the backgrounds to obtain the plant unit region images, and then the segmented lesion images are obtained by segmenting the lesion region/lesion regions from the plant unit region images.

3) Calculation of the whole plant unit area and lesion area of each individual diseased plant unit

The area features of the plant unit region images and the corresponding segmented lesion images are extracted using image processing software, respectively, and then the whole plant unit area and lesion area of each individual diseased plant unit are obtained, or, to represent the areas by the pixel numbers, the number of the whole plant unit pixels and the number of the lesion pixels of each individual diseased plant unit are obtained by statistical calculation operations in image processing software.

4) Calculation of the actual ratio of the lesion area to the area of each whole diseased plant unit

Based on the obtained whole plant unit area and lesion area of each individual diseased plant unit, the actual ratio of the lesion area to the area of the corresponding whole diseased plant unit is calculated according to Formula (1), or, based on the obtained number of the whole plant unit pixels and number of the lesion pixels of each individual diseased plant unit, the actual ratio of the lesion area to the area of the corresponding whole diseased plant unit is calculated according to Formula (2).

$$R_d = \frac{A_d}{A_u} \times 100\,\% \tag{1}$$

where R_d is the actual ratio of the lesion area to the area of the whole individual diseased plant unit, A_d is the lesion area in the whole individual diseased plant unit, and A_u is the area of the whole individual diseased plant unit.

$$R_d = \frac{P_d}{P_u} \times 100\,\% \tag{2}$$

where R_d is the actual ratio of the lesion area to the area of the whole individual diseased plant unit, P_d is the pixel number of the lesion region/regions in the whole individual diseased plant unit, and P_u is the pixel number of the whole individual diseased plant unit.

10.3389/fpls.2025.1510663

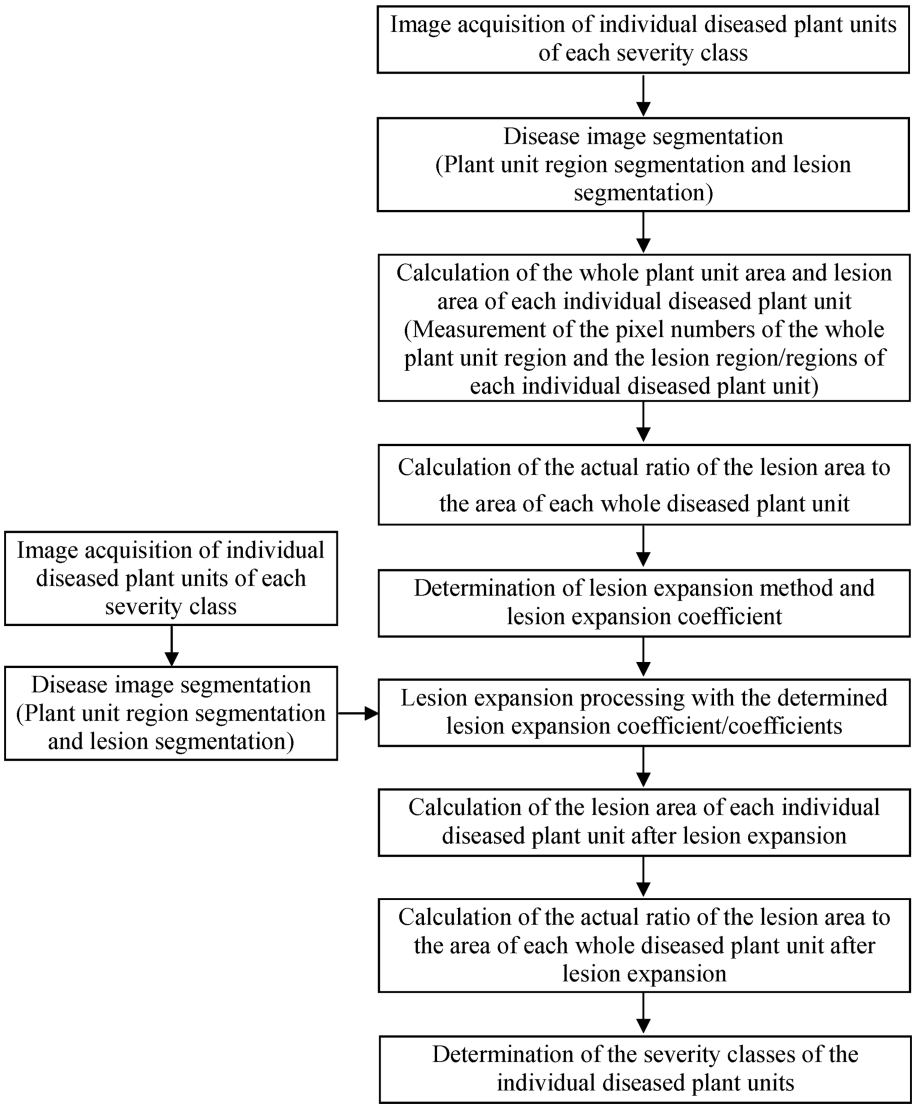


FIGURE 1
Workflow diagram for assessing plant disease severity based on lesion expansion.

5) Determination of lesion expansion method and lesion expansion coefficient

Based on the actual ratios of the lesion areas to the areas of the corresponding whole diseased plant units for all the severity classes of a disease, a variety of methods can be tried to determine the lesion expansion coefficients. For instance, the lesion area ratio in the severity grading standard corresponding to the severity class of a diseased plant unit is divided by the actual ratio of the lesion area to the area of the whole diseased plant unit, the calculation result is recorded as r, and then, the value of r corresponding to the diseased plant unit with the highest actual lesion area ratio among the diseased plant units with the highest severity class can be treated as the lesion expansion coefficient of the diseased plant units in all the severity classes. The value of r corresponding to the diseased plant unit with the highest actual lesion area ratio among the diseased plant units at a severity class can be treated as the lesion expansion coefficient of the diseased plant units in the corresponding severity class. The mean of the r values corresponding to the diseased plant units at a severity class can be treated as the lesion expansion coefficient of the diseased plant units in the corresponding severity class. The mean of the r values corresponding to the diseased plant units in all the severity classes can be treated as the lesion expansion coefficient of the diseased plant units at all the severity classes. Moreover, within a certain range, the enumeration method can be used to determine the lesion expansion coefficient of the diseased plant units in each severity class or the lesion expansion coefficient of the diseased plant units in all the severity classes.

6) Lesion expansion processing with the determined lesion expansion coefficient/coefficients

Using the lesion expansion coefficient/coefficients determined in the above step (Step 5), lesion expansion processing of each lesion image can be carried out based on the centroid of each lesion in image processing software. If the expansion range of a lesion

exceeds the boundary of the diseased plant unit, the exceeded part should be removed to obtain the new expanded lesion.

7) Calculation of the lesion area of each individual diseased plant unit after lesion expansion

Using image processing software, the area features of the lesion images after lesion expansion are extracted, and then the lesion area of each individual diseased plant unit is obtained, or to represent the areas by the pixel numbers, the number of the lesion pixels of each individual diseased plant unit is obtained by statistical calculation operations in image processing software.

8) Calculation of the actual ratio of the lesion area to the area of each whole diseased plant unit after lesion expansion

Based on the obtained whole plant unit area and lesion area or the number of whole plant unit pixels and number of lesion pixels of each individual diseased plant unit after lesion expansion, the actual ratio of the lesion area to the area of the corresponding whole diseased plant unit after lesion expansion is calculated according to Formula (1) or (2) presented above.

9) Determination of the severity classes of the individual diseased plant units

By comparing the actual ratio of the lesion area to the area of each whole diseased plant unit after lesion expansion with the lesion area ratio for each severity class in the disease severity grading standard, the severity of the diseased plant unit is determined, or, efforts can be made to acquire disease image of an individual diseased plant unit, perform segmentation operations to obtain the plant unit region image and segmented lesion image, carry out lesion expansion based on the determined lesion expansion coefficient/coefficients, calculate the lesion area and the whole plant unit area of the individual diseased plant unit after lesion expansion, and calculate the actual ratio of the lesion area to the area of the whole diseased plant unit after lesion expansion, and, finally, the severity of the diseased plant unit can be determined by comparing the actual ratio of the lesion area to the area of the whole diseased plant unit after lesion expansion with the lesion area ratio for each severity class in the disease severity grading standard.

Hereinafter, wheat stripe rust will be taken as an example to illustrate how to implement the severity assessment of wheat stripe rust based on lesion expansion.

2.2 Disease image acquisition, disease image segmentation, and calculation of the actual percentage of the lesion area in the area of each whole single diseased wheat leaf

In the previous study conducted by Jiang et al. (2022), after artificial spray inoculation of *Pst* on the wheat seedlings in the field and indoor environments, single diseased wheat leaves with typical symptoms at the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100% were collected according to the severity grading standard of wheat stripe rust (the National Standard of the People's Republic of China (GB/T 15795–2011) as described above, and then an image of each single diseased wheat leaf was taken with a sheet of A4 white paper as background using a Nikon D700 digital camera (Nikon Corp., Tokyo, Japan), a HUAWEI P30 smartphone, or an iPhone 6S smartphone. For each severity class of wheat stripe rust, 50 images in the JPEG format (one image per single diseased wheat leaf) were acquired. A total of 400 images of the single diseased wheat leaves were acquired for the eight severity classes (i.e., 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%). The resolutions of the acquired disease images were 4256×2832, 3648×2736, and 4032×3024 pixels, respectively, depending on the different devices used for disease image acquisition. Using the Adobe Photoshop 2022 software (Adobe Systems Incorporated, San Jose, CA, USA), the leaf region image and the lesion image, as shown in Figure 2, were manually segmented from each single diseased wheat leaf image, and then pixel statistics of the corresponding leaf region and lesion region/regions were implemented. Subsequently, the actual percentage of the lesion area in the area of each whole single diseased wheat leaf was achieved according to Formula (2) as described above.

2.3 Lesion expansion methods

A novel method based on lesion expansion for severity assessments of wheat stripe rust was proposed in this study because the actual percentage of the lesion area in the area of a whole single diseased wheat leaf infected with *Pst* is much lower than the lesion area percentage of the corresponding severity class in the severity grading standard of wheat stripe rust, as described above. The overall idea of this method was to conduct the expansion of the lesion/lesions in the image of a single diseased wheat leaf, obtain the area/pixel number of the new obtained lesion/lesions and the area/pixel number of the whole single diseased wheat leaf, calculate the actual percentage of the lesion area in the area of the whole single diseased wheat leaf according to Formula (1) or (2) as described above, and finally, compare the actual percentage of the lesion area in the area of the whole single diseased wheat leaf with the percentages for all the severity classes in the disease severity grading standard and achieve the severity class of the single diseased wheat leaf. According to the overall idea, in MATLAB 2019b software (MathWorks, Natick, MA, USA), the segmented lesion image of a whole single diseased wheat leaf infected with *Pst* was converted from the RGB image format to the grayscale image format using the rgb2gray function, and then the grayscale image was binarized using the Otsu method, with the lesion pixels as the foreground (the value of each lesion pixel was regarded as 1) and the remaining pixels in the single diseased wheat leaf image as the background (the value of each pixel in the non-lesion region of the single diseased wheat leaf image was regarded as 0). The connected component extraction algorithm was utilized to extract the connected component/components corresponding to the lesion/lesions from the binary image, and then the expansion of each lesion was carried out based on the corresponding centroid. The expanded lesion image was pasted back into the original single diseased wheat leaf image using centroid alignment, and if there were lesion pixels beyond the leaf region, the values of the corresponding pixels were reset to 0. Subsequently, the area/pixel number of the new obtained lesion/lesions after lesion expansion and the actual percentage of the corresponding lesion area in the

10.3389/fpls.2025.1510663

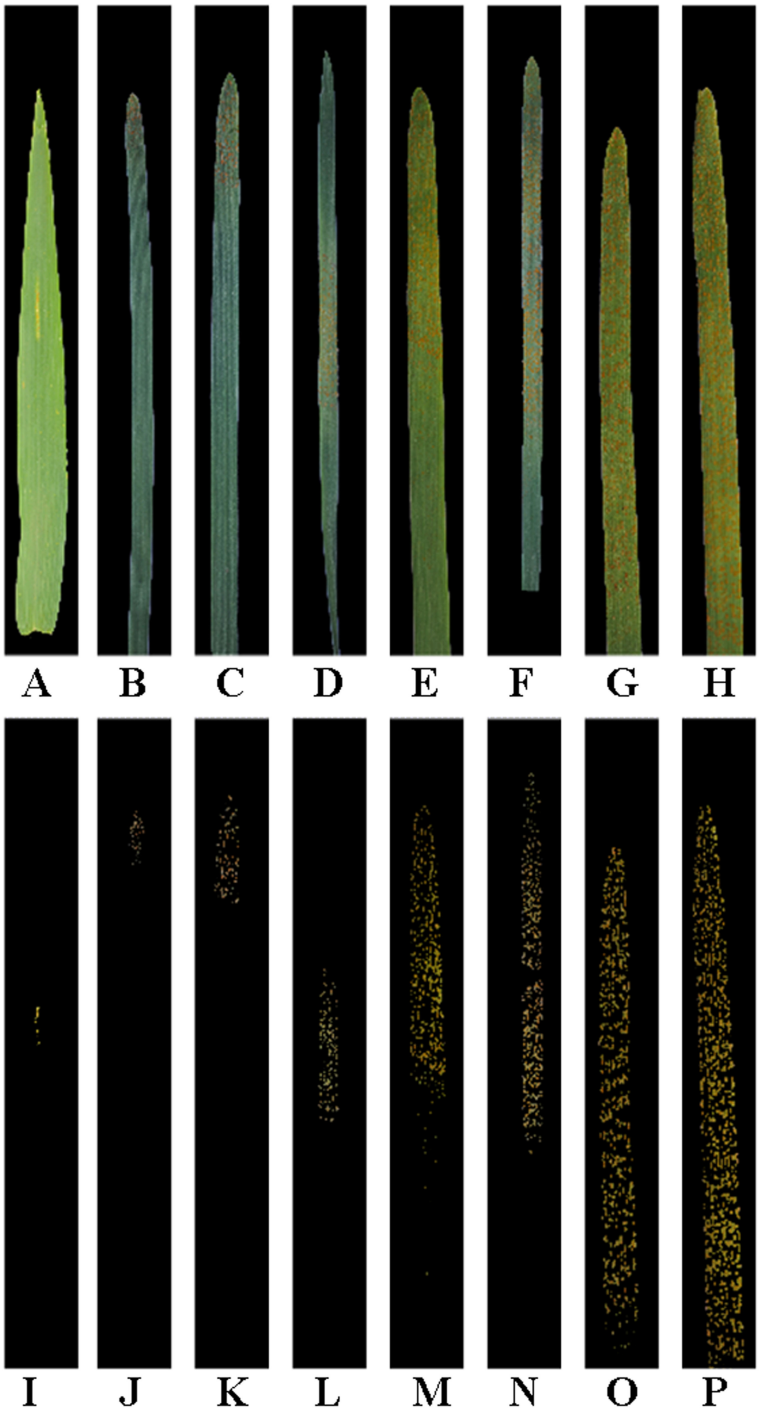


FIGURE 2
Leaf region images and lesion images after manual segmentation of the single wheat leaf images of each severity class of wheat stripe rust. All the images are shown after being cropped uniformly. **(A–H)** Original leaf region images after manual segmentation of the single diseased wheat leaf images in the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, respectively; **(I–P)** Original lesion images after manual segmentation of the single diseased wheat leaf images in the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, respectively.

area of the whole single diseased wheat leaf were calculated. Finally, the severity of the single diseased wheat leaf was assessed by comparing the actual percentage of the lesion area in the area of the whole single diseased wheat leaf with the percentages corresponding to all the eight severity classes in the disease severity grading standard of wheat stripe rust.

From the perspective of the specific implementation of lesion expansion, in this study, three methods to implement lesion expansion were proposed based on two ideas. The following are the two ideas in detail.

Idea 1. For each single diseased leaf image, based on the disease region/regions (namely, lesion/lesions; for wheat stripe rust,

generally, a lesion region is actually the individual region covered with urediospores) in the segmented lesion image, the centroid of each disease region is located. Because the shapes of the lesions of many plant diseases are irregular, it is best to perform image processing based on the centroid of each lesion. Then, based on the distance from the centroid to each point on the boundary of the lesion region, the point extends outward (the line connecting the centroid to each point on the boundary of the lesion region extends outward in the direction from the centroid to the corresponding point on the lesion boundary), according to a determined lesion expansion coefficient. Subsequently, the extended ends are connected together to form a closed region (i.e., a new lesion). Under this idea, two methods for lesion boundary selection were proposed based on the external edge and the internal edge of the lesion, respectively, and thus two specific implementation methods of lesion expansion were developed, which were named the lesion expansion method based on the external edge of the lesion (i.e., lesion expansion method 1—the external edge-based lesion expansion method) and the lesion expansion method based on the internal edge of the lesion (i.e., lesion expansion method 2—the internal edge-based lesion expansion method), respectively.

Idea 2. For each single diseased leaf image, based on the disease region/regions (lesion/lesions) in the segmented lesion image, the centroid of each disease region is located. Because the shapes of the lesions of many plant diseases are irregular, it is best to perform image processing based on the centroid of each lesion. Then, based on the centroid, the lesion expansion is conducted based on a determined lesion expansion coefficient to form a new lesion (the whole lesion region expands outward from the centroid based on the determined lesion expansion coefficient, and thus a larger closed region is formed, which is a new lesion). Under this idea, the developed method for the specific implementation of lesion expansion was named the lesion expansion method based on the image scaling algorithm (i.e., lesion expansion method 3).

The three lesion expansion methods were specifically implemented as follows.

2.3.1 Lesion expansion method 1—the external edge-based lesion expansion method

Based on the connected component/components corresponding to the lesion/lesions extracted from the binary image of the segmented lesion image of a whole single diseased leaf using the connected component extraction algorithm as described above, a morphological hole-filling operation was performed on the extracted connected component/components to remove the internal holes. The location of the centroid of each lesion was calculated. A morphological dilation operation was performed on the connected component/components with a structure element with the size of 3×3, and then the original binary lesion image was subtracted by the obtained image to achieve the edge of each lesion (i.e., the boundary of each lesion region). The external edge of each lesion was found in this way and ensured that the original lesion could be calculated when it was only a line or a point. Then, by taking the centroid of a lesion region as the center point, each edge point (i.e., each point on the boundary of the lesion region) was extended in the opposite direction of its connection with the centroid (that is, the direction away from the centroid). Assuming that the distance between an edge point and the centroid was d and the lesion expansion coefficient was set to k, the distance between the edge point after extension and the centroid was $d \times k$. Subsequently, the ends of the edge points after extension were connected together to form a new expanded lesion, and the morphological hole-filling operation was performed on the new expanded lesion. The implementation of lesion expansion method 1 (the external edge-based lesion expansion method) is illustrated in Figure 3. Finally, the expanded lesion image was pasted back into the original single diseased leaf image using centroid alignment, and if there were lesion pixels beyond the leaf region, the values of the corresponding pixels were reset to 0. The color of the new obtained lesion after the expansion was filled with the average value of the R, G, and B components of the corresponding original lesion, and thus a single leaf image containing the new obtained lesion/lesions after the expansion was achieved.

2.3.2 Lesion expansion method 2—the internal edge-based lesion expansion method

As in lesion expansion method 1 described above, based on the connected component/components corresponding to the lesion/lesions extracted from the binary image of the segmented lesion image of a whole single diseased leaf using the connected component extraction algorithm, first, a morphological hole-filling operation was performed on the extracted connected component/components to remove the internal holes. The centroid of each lesion was located and the corresponding location information was achieved. A morphological erosion operation was performed on the connected component/components and then the obtained image was subtracted by the original binary lesion image to achieve the edge of each lesion (i.e., the boundary of each lesion region). Thus, the internal edge of each lesion was found. If there was only one pixel in a lesion, then the pixel itself was the edge of the lesion, and in the case of a lesion with a single pixel and the case where the length or width of a lesion was one pixel, no expansion processing was performed. By taking the centroid of a lesion region as the center point, each edge point (i.e., each point on the boundary of the lesion region) was then extended in the opposite direction of its connection with the centroid (that is, the direction away from the centroid). Assuming that the distance between an edge point and the centroid was d and the lesion expansion coefficient was k, the distance between the edge point after extension and the centroid was $d \times k$. Subsequently, the ends of the edge points after extension were connected together, and the morphological hole-filling operation was performed on the connected region to form a new expanded lesion. The implementation of lesion expansion method 2 (the internal edge-based lesion expansion method) is illustrated in Figure 4. The expanded lesion image was pasted back into the original single diseased leaf image using centroid alignment, and if there were lesion pixels beyond the leaf region, the values of the corresponding pixels were reset to 0. Finally, the color of the new obtained lesion after the expansion was filled with the average value of the R, G, and B components of the corresponding original lesion, and thus the

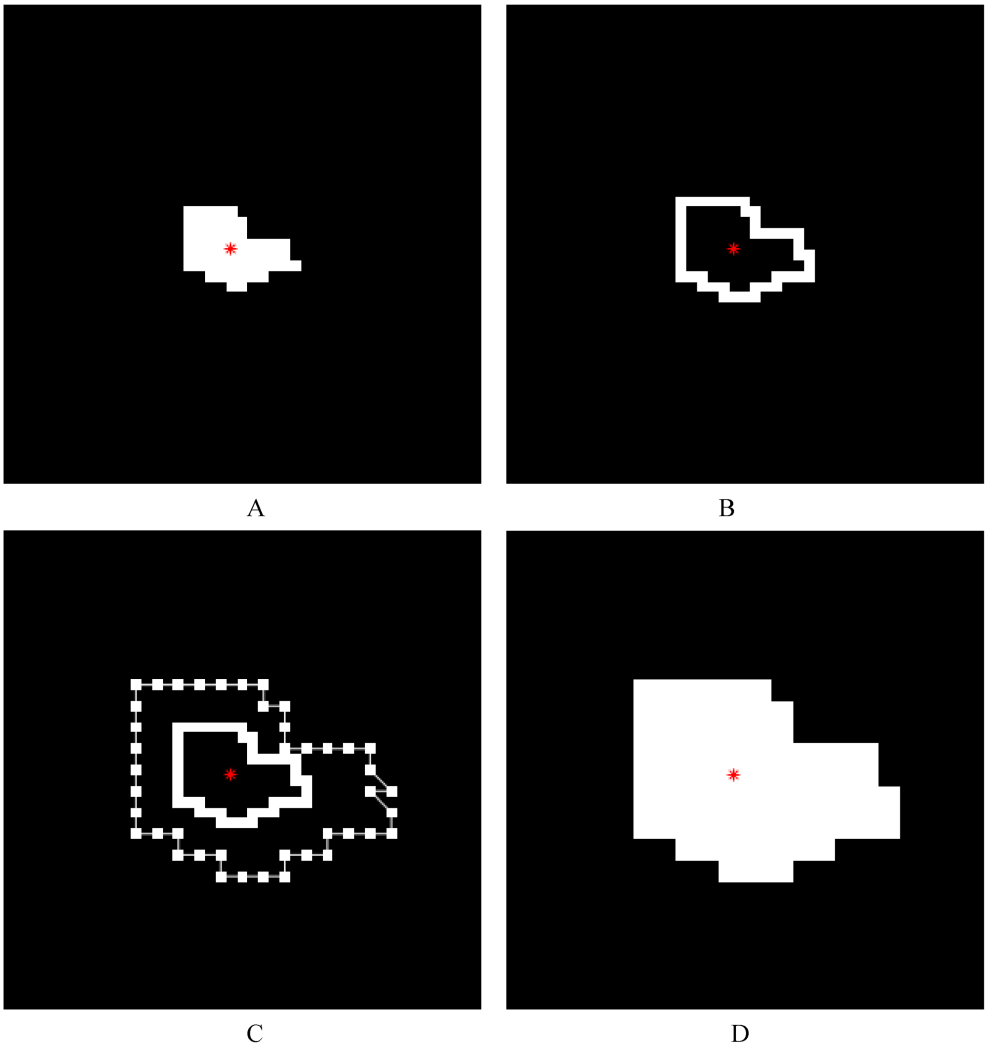


FIGURE 3
Illustrations of lesion image processing using lesion expansion method 1 (the external edge-based lesion expansion method). The red asterisk represents the centroid of a lesion. **(A)** The white region represents a lesion and the black region represents the background. **(B)** The external edge of each lesion (i.e., the boundary of each lesion region) was achieved after the original binary lesion image in **A** was subtracted from the image obtained by performing a morphological dilation operation on the lesion in **A** with a 3×3 structure element. **(C)** Based on the lesion expansion coefficient of 2 (k=2), the edge of the lesion was expanded outward, and then the expanded edge points were connected with straight lines in order. **(D)** The result (i.e., the new lesion obtained after lesion expansion) was achieved after the morphological hole-filling operation was performed on the connected region obtained by connecting the expanded edge points.

single leaf image containing the new obtained lesion/lesions after the expansion was achieved.

2.3.3 Lesion expansion method 3—the lesion expansion method based on an image scaling algorithm

As in lesion expansion method 1 and lesion expansion method 2 described above, based on the connected component/components corresponding to the lesion/lesions extracted from the binary image of the segmented lesion image of a whole single diseased leaf using the connected component extraction algorithm, first, a morphological hole-filling operation was performed on the extracted connected component/components to remove the internal holes. The location of the centroid of each lesion was calculated and the corresponding location information was achieved. Given that the lesion expansion coefficient was set to k, the bicubic interpolation algorithm was used to enlarge each original binary lesion to k times based on the corresponding centroid. The enlarged image was then rebinarized, that is, each non-zero pixel was considered as the lesion foreground and the corresponding value was set to 1, and each pixel with the value of 0 remained unchanged. The implementation of lesion expansion method 3 (the lesion expansion method based on an image scaling algorithm) is illustrated in Figure 5. As described above, the expanded lesion image was pasted back into the original single diseased leaf image using centroid alignment, and if there were any lesion pixels beyond the leaf region, the values of the corresponding pixels were reset to 0. The color of the new obtained lesion after the expansion was filled with the average value of the R, G, and B components of the corresponding original lesion. Thus, a single leaf

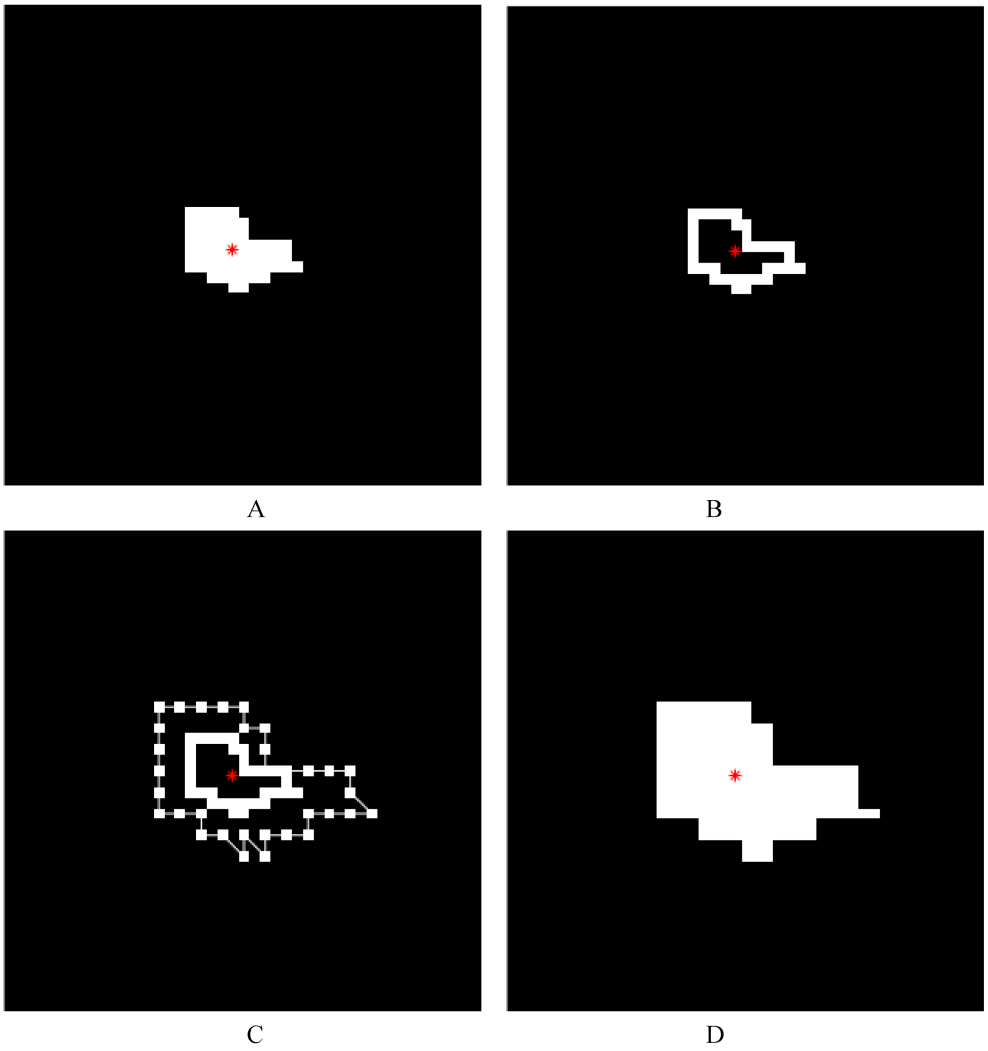


FIGURE 4
Illustrations of lesion image processing using lesion expansion method 2 (the internal edge-based lesion expansion method). The red asterisk represents the centroid of a lesion. **(A)** The white region represents a lesion and the black region represents the background. **(B)** The internal edge of each lesion (i.e., the boundary of each lesion region) was achieved after the image obtained by performing a morphological erosion operation on the lesion in **A** was subtracted from the original binary lesion image. **(C)** Based on the lesion expansion coefficient of 2 (k=2), the edge of the lesion was expanded outward, and then the expanded edge points were connected with straight lines in order. **(D)** The result (i.e., the new lesion obtained after lesion expansion) was achieved after the morphological hole-filling operation was performed on the connected region obtained by connecting the expanded edge points.

image containing the new obtained lesion/lesions after the expansion was achieved.

2.4 Methods to determine lesion expansion coefficients

The lesion expansion coefficient is crucial for lesion expansion processing, which has a great influence on the area of each expanded lesion, thus affecting the subsequent assessment of plant disease severity. In this study, the following three methods were used to determine the lesion expansion coefficients.

2.4.1 Lesion expansion coefficient determination method 1

After the severity classes of the single diseased leaves were determined, the lesion area percentage in the severity grading standard corresponding to the severity class of a single diseased leaf was divided by the actual percentage of the lesion area to the area of the whole single diseased leaf. The calculation result was recorded as r and the value of r corresponding to the single diseased leaf with the highest actual lesion area percentage among the single diseased leaves with the highest severity class (100%) was recorded as Grate and was treated as the lesion expansion coefficient of the single diseased leaves in all the wheat stripe rust severity classes.

2.4.2 Lesion expansion coefficient determination method 2

As in lesion expansion coefficient determination method 1 described above, after the severity classes of the single diseased leaves were determined, the lesion area percentage in the severity grading standard corresponding to the severity class of a single diseased leaf was divided by the actual percentage of the lesion area to the area of the whole single diseased leaf, and the calculation result was recorded as *r*. The value of *r* corresponding to the single diseased leaf with the highest actual lesion area percentage among the single diseased leaves in one of the severity classes was recorded as minrate_S (in which, *S* was the severity class and it could be 1%, 5%, 10%, 20%, 40%, 60%, 80%, or 100%), which was treated as the lesion expansion coefficient of the single diseased leaves at the corresponding severity class (*S*). In particular, when the severity class was 100%, the value of $\text{minrate}_{100\%}$ was the same as the value of Grate obtained using lesion expansion coefficient determination method 1.

2.4.3 Lesion expansion coefficient determination method 3

As in lesion expansion coefficient determination method 1 and lesion expansion coefficient determination method 2 described above, after the severity classes of the single diseased leaves were determined, the lesion area percentage in the severity grading standard corresponding to the severity class of a single diseased leaf was divided by the actual percentage of the lesion area to the area of the whole single diseased leaf, and the calculation result was recorded as *r*. The mean of the *r* values corresponding to the single diseased leaves in one of the severity classes was recorded as meanrate_S (in which, *S* was the severity class and it could be 1%, 5%, 10%, 20%, 40%, 60%, 80%, or 100%), which was treated as the lesion expansion coefficient of the single diseased leaves at the corresponding severity class (*S*).

The lesion expansion coefficients for wheat stripe rust determined using the three lesion expansion coefficient determination methods described above are shown in Table 1. Based on the determined lesion expansion coefficients, lesion expansion processing of the lesion image of each single diseased leaf image was conducted using the lesion expansion methods described above. According to the processing results, the performances for severity assessments of wheat stripe rust using the three lesion expansion coefficient determination methods were compared. In fact, a total of nine specific method combinations for severity assessments were developed by combining the three lesion expansion methods with the three lesion expansion coefficient determination methods in this study.

2.5 Calculation of the actual percentage of the lesion area in the area of each whole single diseased leaf after lesion expansion processing and severity assessment of the single diseased leaves

For each single diseased leaf containing the expanded lesion/lesions obtained by lesion image processing based on a determined lesion expansion coefficient method and a lesion expansion method, the number of the pixels of the whole single diseased leaf and the number of the lesion pixels after lesion expansion were calculated, respectively. The actual percentage of the lesion area in the area of the corresponding whole single diseased leaf after lesion expansion was then calculated according to Formula (2) presented above. Subsequently, the obtained actual lesion area percentage of each

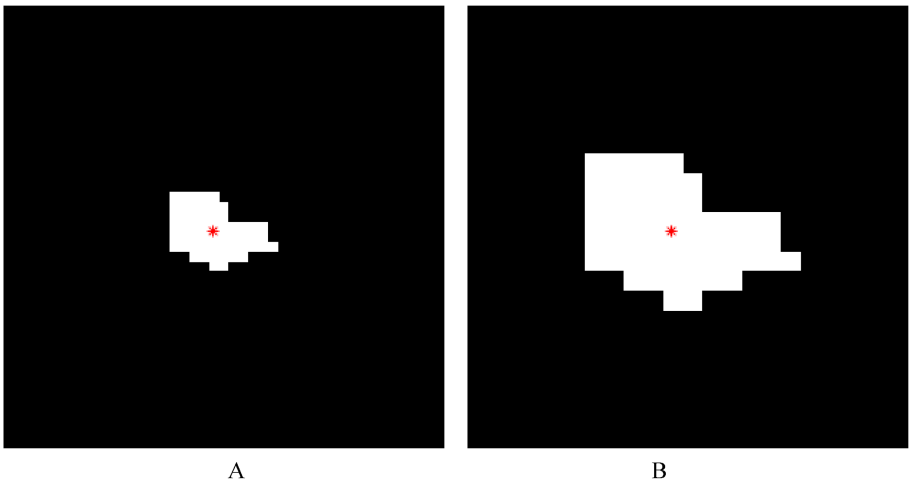


FIGURE 5
Illustrations of lesion image processing using lesion expansion method 3 (the lesion expansion method based on image scaling algorithm). The red asterisk represents the centroid of a lesion. **(A)** The white region represents a lesion and the black region represents the background. **(B)** Based on the centroid of the original binary lesion and a lesion expansion coefficient of 2 (k=2), the original binary lesion in **A** was enlarged. In the enlarged image, each non-zero pixel was considered as the lesion foreground and the corresponding value was set to 1, and each remaining pixel with the value of 0 was unchanged.

10.3389/fpls.2025.1510663

TABLE 1 Lesion expansion coefficients for each severity class of wheat stripe rust determined using the three lesion expansion coefficient determination methods, respectively.

Lesion expansion coefficient determination method	Severity class	Lesion expansion coefficient
Lesion expansion coefficient determination method 1	All the severity classes of wheat stripe rust including 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%	2.74 (Grate)
Lesion expansion coefficient determination method 2	1%	1.28 ($minrate_{1\%}$)
	5%	3.05 ($minrate_{5\%}$)
	10%	3.04 ($minrate_{10\%}$)
	20%	3.17 ($minrate_{20\%}$)
	40%	2.88 ($minrate_{40\%}$)
	60%	3.26 ($minrate_{60\%}$)
	80%	3.31 ($minrate_{80\%}$)
	100%	2.74 ($minrate_{100\%}$)
Lesion expansion coefficient determination method 3	1%	3.52 ($meanrate_{1\%}$)
	5%	4.06 ($meanrate_{5\%}$)
	10%	4.12 ($meanrate_{10\%}$)
	20%	4.17 ($meanrate_{20\%}$)
	40%	4.21 ($meanrate_{40\%}$)
	60%	3.63 ($meanrate_{60\%}$)
	80%	3.78 ($meanrate_{80\%}$)
	100%	3.31 ($meanrate_{100\%}$)

single diseased leaf was compared with the lesion area percentage for each severity class in the disease severity grading standard of wheat stripe rust as described above. Finally, the severity of each single diseased leaf was determined.

2.6 Evaluation of severity assessment performance

In this study, the accuracy of severity assessment, calculated using the maximum error reference method described by Xiao (1997), was utilized to evaluate severity assessment performance. The severity classes of wheat stripe rust of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100% were represented by 1, 2, 3, 4, 5, 6, 7, and 8, respectively, and then the accuracy of the severity assessment of an individual single diseased leaf and the accuracy of severity assessments of multiple single diseased leaves were calculated using Formulas (3) and (4), respectively.

$$s_i = \left(1 - \frac{|\hat{x}_i - x_i|}{x_i \vee (x_{\max} - x_i)}\right) \times 100\,\% \tag{3}$$

where s_i is the accuracy of severity assessment of the ith individual single diseased leaf; $\hat{x}_i$ is the representative value of the severity class assessed for the ith individual single diseased leaf; x_i is the representative value of the severity class determined by an actual survey for the ith individual single diseased leaf; $x_{\max}$ is the representative value of the highest severity class (100%) in the severity grading standard of wheat stripe rust, i.e., 8 in this study; and $x_i \vee (x_{\max} - x_i)$ is the maximum in x_i and $x_{\max}$-x_i, namely, the maximum error of the severity assessment of the ith individual single diseased leaf.

$$s = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{|\hat{x}_i - x_i|}{x_i \vee (x_{\max} - x_i)}\right) \times 100\,\% \tag{4}$$

where s is the accuracy of severity assessments of a total of n single diseased leaves; n is the total number of the single diseased leaves assessed; $\hat{x}_i$ is the representative value of the severity class assessed for the ith individual single diseased leaf; x_i is the representative value of the severity class determined by an actual survey for the ith individual single diseased leaf; $x_{\max}$ is the representative value of the highest severity class (100%) in the severity grading standard of wheat stripe rust, i.e., 8 in this study; and $x_i \vee (x_{\max} - x_i)$ is the maximum in x_i and $x_{\max}$-x_i, namely, the maximum error of the severity assessment of the ith individual single diseased leaf.

The accuracy of severity assessments of multiple single diseased leaves is the mean of the accuracies of severity assessments of multiple individual single diseased leaves. On the premise of obtaining the accuracies of severity assessments of multiple individual single diseased leaves, the accuracy of severity assessments of multiple single diseased leaves can be calculated using Formula (5).

$$s = \frac{1}{n}\sum_{i=1}^{n} s_i \quad (5)$$

where s is the accuracy of severity assessments of a total of n single diseased leaves; n is the total number of the single diseased leaves assessed; s_i is the accuracy of severity assessment of the ith individual single diseased leaf.

3 Results

Based on the lesion expansion coefficients determined using the three lesion expansion coefficient determination methods, the images of the single diseased wheat leaves in each severity class of wheat stripe rust were processed using the three lesion expansion methods described above, respectively. Severity assessments of the single diseased leaves were then performed based on the obtained actual percentages of the lesion areas after lesion expansion of the single diseased leaves in the area of the corresponding whole single diseased leaves, according to the severity grading standard of wheat stripe rust described above, and the results are shown in Table 2. The results (as shown in Figures 6–8) obtained after processing the images of the single diseased wheat leaves in each severity class of wheat stripe rust using the three lesion expansion methods with the lesion expansion coefficient of 2.74 (i.e., Grate), were taken as the examples to illustrate the results achieved after lesion expansion processing. In particular, the results achieved after lesion expansion processing using lesion expansion method 1 (i.e., the external edge-based lesion expansion method), lesion expansion method 2 (i.e., the internal edge-based lesion expansion method), and lesion expansion method 3 (i.e., the lesion expansion method based on image scaling algorithm) are shown in Figure 6, Figure 7, and Figure 8, respectively.

TABLE 2 Severity assessment results of the single diseased wheat leaves in each severity class of wheat stripe rust after lesion expansion processing using the three lesion expansion methods with the lesion expansion coefficients determined using the three lesion expansion coefficient determination methods, respectively.

Severity class	Representative value of severity class	Lesion expansion coefficient	Accuracy (s)		
			Lesion expansion method 1	Lesion expansion method 2	Lesion expansion method 3
1%	1	2.74 (Grate)	86.86%	88.00%	88.00%
1%	1	1.28 ($minrate_{1\%}$)	92.29%	97.14%	98.57%
1%	1	3.52 ($meanrate_{1\%}$)	85.43%	86.57%	86.57%
5%	2	2.74 (Grate)	87.33%	92.67%	95.00%
5%	2	3.05 ($minrate_{5\%}$)	85.67%	90.00%	89.33%
5%	2	4.06 ($meanrate_{5\%}$)	78.00%	82.33%	83.00%
10%	3	2.74 (Grate)	91.60%	99.20%	98.40%
10%	3	3.04 ($minrate_{10\%}$)	88.80%	94.80%	95.60%
10%	3	4.12 ($meanrate_{10\%}$)	82.80%	87.20%	85.60%
20%	4	2.74 (Grate)	94.50%	97.50%	99.00%
20%	4	3.17 ($minrate_{20\%}$)	91.00%	94.00%	95.50%
20%	4	4.17 ($meanrate_{20\%}$)	84.00%	88.00%	89.00%
40%	5	2.74 (Grate)	92.80%	95.60%	96.00%
40%	5	2.88 ($minrate_{40\%}$)	90.80%	94.00%	95.60%
40%	5	4.21 ($meanrate_{40\%}$)	80.00%	83.20%	85.20%
60%	6	2.74 (Grate)	92.33%	98.00%	98.67%
60%	6	3.26 ($minrate_{60\%}$)	90.00%	93.00%	93.67%
60%	6	3.63 ($meanrate_{60\%}$)	89.00%	91.33%	91.33%
80%	7	2.74 (Grate)	95.14%	98.86%	99.43%
80%	7	3.31 ($minrate_{80\%}$)	90.29%	95.14%	95.71%
80%	7	3.78 ($meanrate_{80\%}$)	89.14%	91.43%	92.00%
100%	8	2.74 (Grate/$minrate_{100\%}$)	100.00%	98.75%	94.75%
100%	8	3.31 ($meanrate_{100\%}$)	100.00%	99.75%	99.50%

In the table, each accuracy value is the mean of the accuracies of severity assessments of all the single diseased leaves at a severity class obtained using a combination method of a lesion expansion method and a lesion expansion coefficient determination method.

10.3389/fpls.2025.1510663

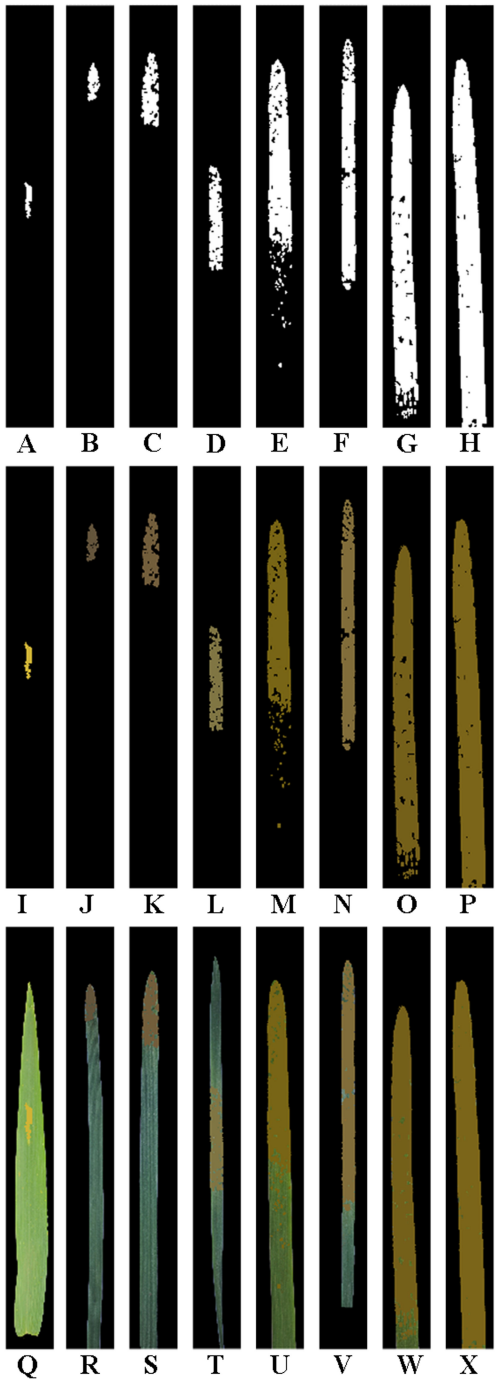


FIGURE 6
Results obtained after lesion expansion of the single wheat leaf images of each severity class of wheat stripe rust shown in Figure 2 using lesion expansion method 1 (i.e., the external edge-based lesion expansion method) with a lesion expansion coefficient of 2.74 (i.e., Grate). To demonstrate more clearly, all the images are shown after being cropped uniformly. **(A–H)** The obtained images after lesion expansion of the segmented lesion images of the single diseased wheat leaf images in the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, respectively. **(I–P)** The obtained RGB images after lesion expansion of the segmented lesion images of the single diseased wheat leaf images in the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, respectively. **(Q–X)** The RGB images obtained after lesion expansion for the single diseased wheat leaves in the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, respectively.

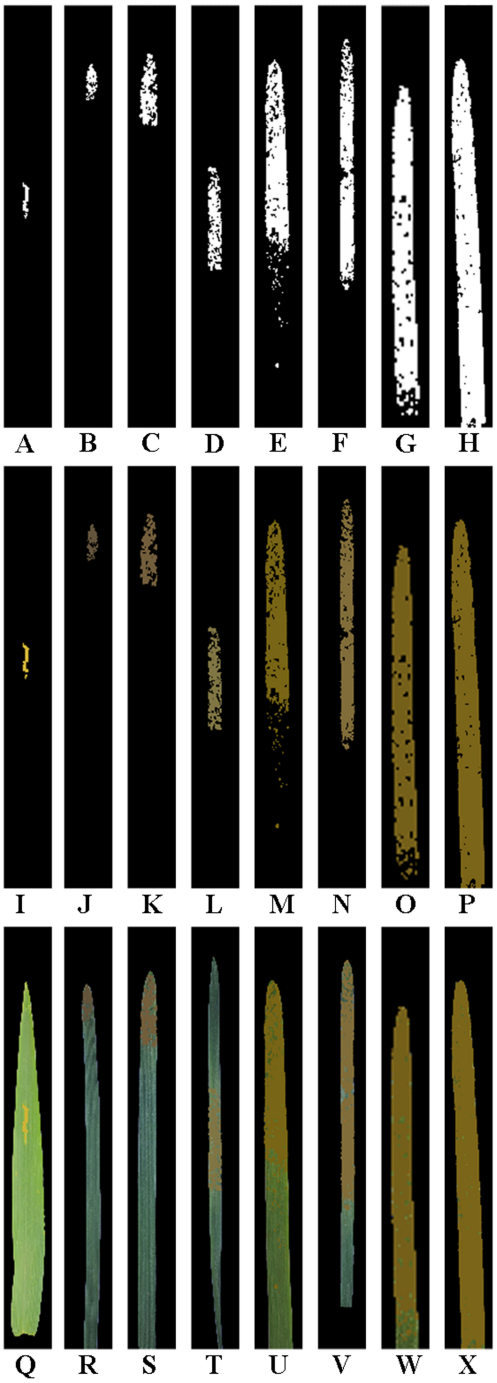


FIGURE 7
Results obtained after lesion expansion of the single wheat leaf images of each severity class of wheat stripe rust shown in Figure 2 using lesion expansion method 2 (i.e., the internal edge-based lesion expansion method) with the lesion expansion coefficient of 2.74 (i.e., Grate). All the images are shown after being cropped uniformly. **(A–H)** The obtained images after lesion expansion of the segmented lesion images of the single diseased wheat leaf images in the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, respectively. **(I–P)** The obtained RGB images after lesion expansion of the segmented lesion images of the single diseased wheat leaf images in the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, respectively. **(Q–X)** The RGB images obtained after lesion expansion for the single diseased wheat leaves in the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, respectively.

10.3389/fpls.2025.1510663

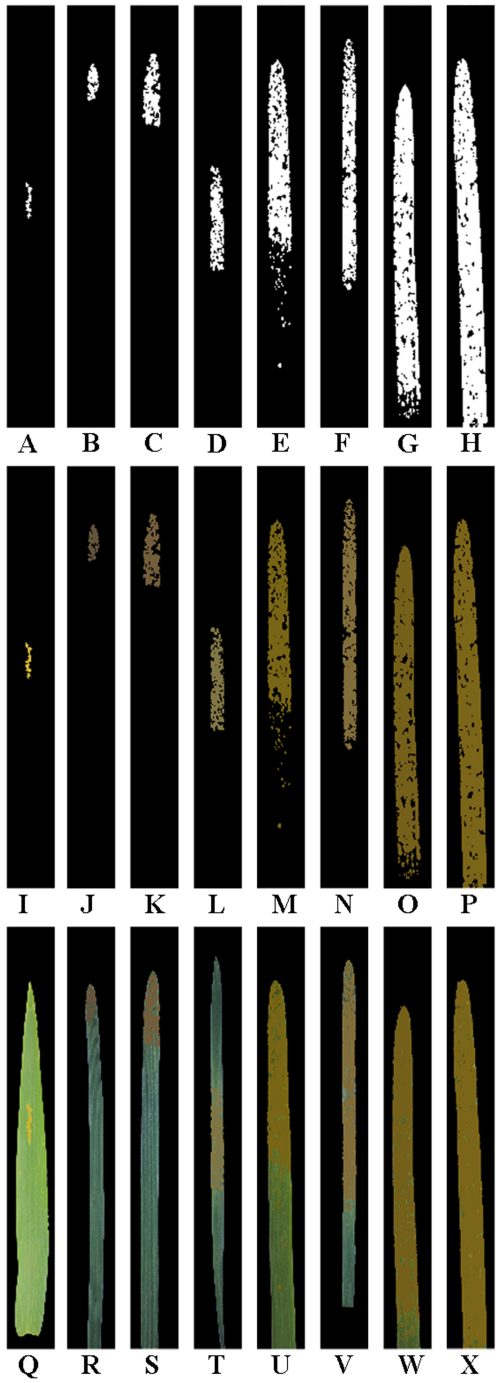


FIGURE 8
Results obtained after lesion expansion in the single wheat leaf images in each severity class of wheat stripe rust shown in Figure 2 using lesion expansion method 3 (i.e., the lesion expansion method based on image scaling algorithm) with a lesion expansion coefficient of 2.74 (i.e., Grate). All the images are shown after being cropped uniformly. **(A–H)** The obtained images after lesion expansion of the segmented lesion images of the single diseased wheat leaf images in the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, respectively. **(I–P)** The obtained RGB images after lesion expansion of the segmented lesion images of the single diseased wheat leaf images in the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, respectively. **(Q–X)** The RGB images obtained after lesion expansion for the single diseased wheat leaves in the severity classes of 1%, 5%, 10%, 20%, 40%, 60%, 80%, and 100%, respectively.

As shown in Table 2, after lesion expansion processing using lesion expansion method 1 with all the determined lesion expansion coefficients, respectively, the accuracy of severity assessments of the single diseased wheat leaves in each severity class of wheat stripe rust, namely, the mean of the accuracies of severity assessments of all the single diseased leaves in each severity class of wheat stripe rust, was in the range of 78.00% to 100.00%. After lesion expansion processing using lesion expansion method 2 with all the determined lesion expansion coefficients, respectively, the accuracy of severity assessments of the single diseased wheat leaves in each severity class of wheat stripe rust, namely, the mean of the accuracies of severity assessments of all the single diseased leaves in each severity class of wheat stripe rust, was in the range of 82.33% to 99.75%. After lesion expansion processing using lesion expansion method 3 with all the determined lesion expansion coefficients, respectively, the accuracy of severity assessments of the single diseased wheat leaves in each severity class of wheat stripe rust, namely, the mean of the accuracies of severity assessments of all the single diseased leaves in each severity class of wheat stripe rust, was in the range of 83.00% to 99.50%. When not considering the lesion expansion coefficients (namely, with all the lesion expansion coefficients determined using the three lesion expansion coefficient determination methods), the accuracies of severity assessments of all the single wheat leaves with wheat stripe rust when the lesion expansion operations were conducted using lesion expansion methods 1, 2, and 3, were 89.91%, 93.13%, and 93.34%, respectively, as shown in Figure 9. The results indicated that, overall, the best severity assessment performance for wheat stripe rust was achieved using lesion expansion method 3 to conduct lesion expansion, the performance of severity assessment of wheat stripe rust achieved using lesion expansion method 2 to conduct lesion expansion ranked second, and the severity assessment performance achieved using lesion expansion method 1 to conduct lesion expansion ranked third.

As shown in Table 2, for the lesion expansion coefficient of 2.74 (Grate) determined using lesion expansion coefficient determination method 1, when lesion expansion processing was conducted using lesion expansion method 1, the accuracy of the severity assessments of the single diseased wheat leaves in each severity class of wheat stripe rust, namely, the mean of the accuracies of severity assessments of all the single diseased leaves in each severity class of wheat stripe rust, was in the range of 86.68% to 100.00%, and the accuracy of severity assessments of all the single wheat leaves infected with wheat stripe rust was 92.57% (as shown in Figure 10). When lesion expansion processing was conducted using lesion expansion method 2, the accuracy of severity assessments of the single diseased wheat leaves in each severity class of wheat stripe rust, namely, the mean of the accuracies of severity assessments of all the single diseased leaves in each severity class of wheat stripe rust, was in the range of 88.00% to 99.20%, and the accuracy of severity assessments of all the single wheat leaves infected with wheat stripe rust was 96.07% (as shown in Figure 10). Finally, when lesion expansion processing was conducted using lesion expansion method 3, the accuracy of severity assessments of the single diseased wheat leaves in each severity class of wheat stripe rust, namely, the mean of the accuracies of severity assessments of

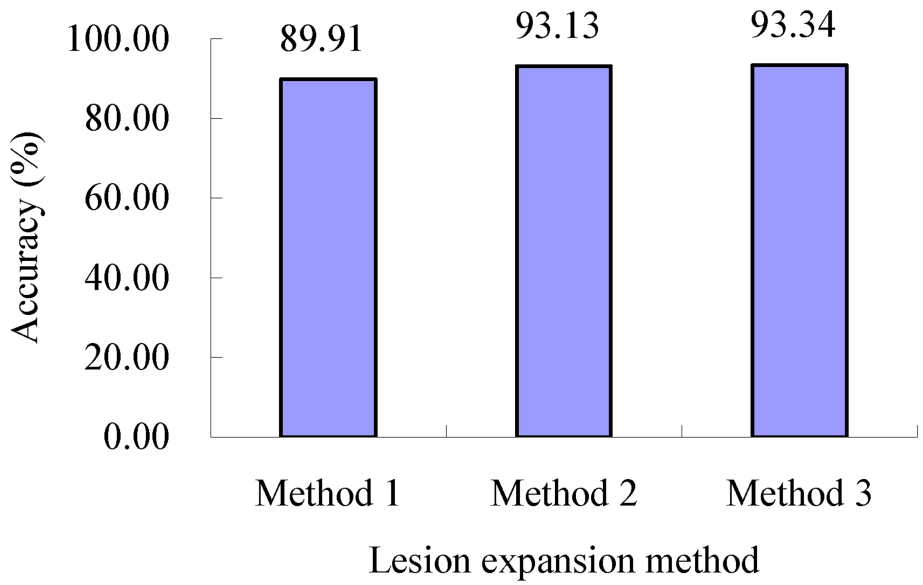


FIGURE 9
Severity assessment accuracies of all the single wheat leaves with wheat stripe rust when the lesion expansion operations were conducted under the condition without considering the lesion expansion coefficients using lesion expansion methods 1, 2, and 3, respectively. The value of 89.91% is the mean of the assessment accuracies of all the single wheat leaves infected with wheat stripe rust achieved when the lesion expansion operations were conducted using the three method combinations of lesion expansion method 1 and the three lesion expansion coefficient determination methods, respectively. The value of 93.13% is the mean of the assessment accuracies of all the single wheat leaves with wheat stripe rust achieved when the lesion expansion operations were conducted using the three method combinations of lesion expansion method 2 and the three lesion expansion coefficient determination methods, respectively. The value of 93.34% is the mean of the assessment accuracies of all the single wheat leaves with wheat stripe rust achieved when the lesion expansion operations were conducted using the three method combinations of lesion expansion method 3 and the three lesion expansion coefficient determination methods, respectively.

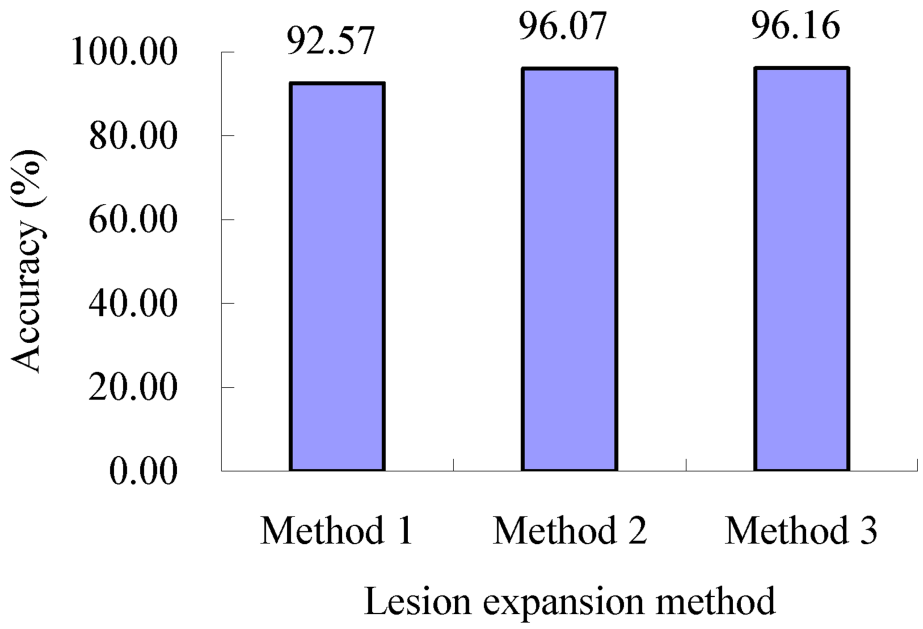


FIGURE 10
Severity assessment accuracies of all the single wheat leaves infected with wheat stripe rust when the lesion expansion operations were conducted with the lesion expansion coefficient of 2.74 (Grate) using lesion expansion methods 1, 2, and 3, respectively. The value of 92.57% is the mean of the assessment accuracies of all the single wheat leaves infected with wheat stripe rust achieved when the lesion expansion operations were conducted using the method combination of lesion expansion method 1 and lesion expansion coefficient determination method 1. The value of 96.07% is the mean of the assessment accuracies of all the single wheat leaves infected with wheat stripe rust achieved when the lesion expansion operations were conducted using the method combination of lesion expansion method 2 and lesion expansion coefficient determination method 1. The value of 96.16% is the mean of the assessment accuracies of all the single wheat leaves infected with wheat stripe rust achieved when the lesion expansion operations were conducted using the method combination of lesion expansion method 3 and lesion expansion coefficient determination method 1.

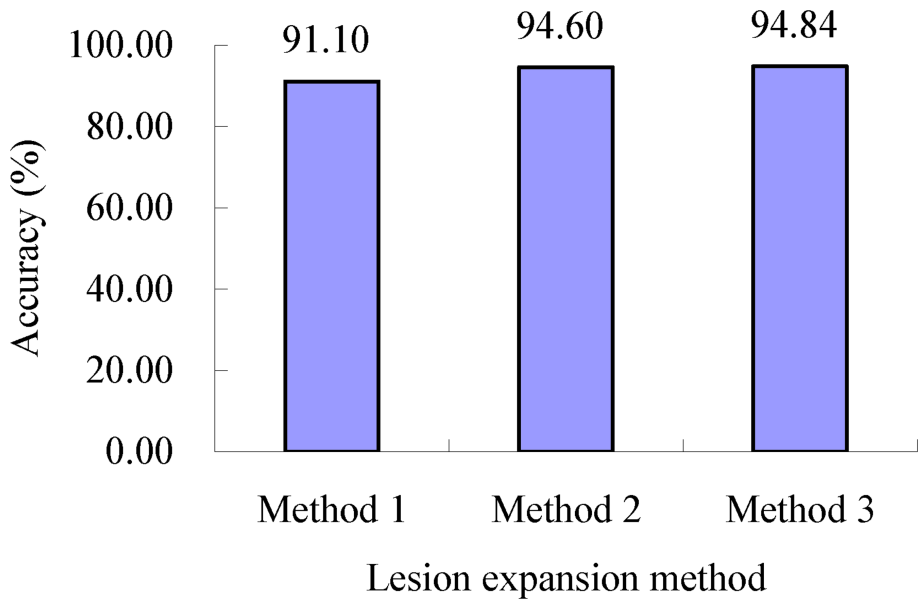


FIGURE 11
Severity assessment accuracies of all the single wheat leaves infected with wheat stripe rust when the lesion expansion operations were conducted with the lesion expansion coefficients $minrate_S$ using lesion expansion methods 1, 2, and 3, respectively. The value of 91.10% is the mean of the assessment accuracies of all the single wheat leaves infected with wheat stripe rust achieved when the lesion expansion operations were conducted using the method combination of lesion expansion method 1 and lesion expansion coefficient determination method 2. The value of 94.60% is the mean of the assessment accuracies of all the single wheat leaves infected with wheat stripe rust achieved when the lesion expansion operations were conducted using the method combination of lesion expansion method 2 and lesion expansion coefficient determination method 2. The value of 94.84% is the mean of the assessment accuracies of all the single wheat leaves infected with wheat stripe rust achieved when the lesion expansion operations were conducted using the method combination of lesion expansion method 3 and lesion expansion coefficient determination method 2.

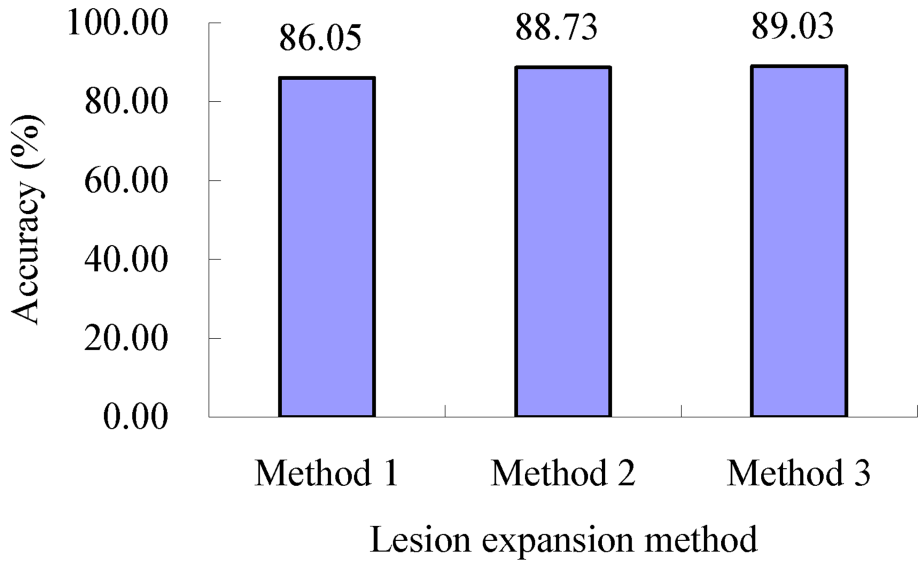


FIGURE 12
Severity assessment accuracies of all the single wheat leaves infected with wheat stripe rust when the lesion expansion operations were conducted with the lesion expansion coefficients $meanrate_S$ using lesion expansion methods 1, 2, and 3, respectively. The value of 86.05% is the mean of the assessment accuracies of all the single wheat leaves infected with wheat stripe rust achieved when the lesion expansion operations were conducted using the method combination of lesion expansion method 1 and lesion expansion coefficient determination method 3. The value of 88.73% is the mean of the assessment accuracies of all the single wheat leaves infected with wheat stripe rust achieved when the lesion expansion operations were conducted using the method combination of lesion expansion method 2 and lesion expansion coefficient determination method 3. The value of 89.03% is the mean of the assessment accuracies of all the single wheat leaves infected with wheat stripe rust achieved when the lesion expansion operations were conducted using the method combination of lesion expansion method 3 and lesion expansion coefficient determination method 3.

all the single diseased leaves in each severity class of wheat stripe rust, was in the range of 88.00% to 99.43%, and the accuracy of severity assessments of all the single wheat leaves infected with wheat stripe rust was 96.16% (as shown in Figure 10). The results indicated that, overall, based on the lesion expansion coefficient of 2.74 (Grate) determined using lesion expansion coefficient determination method 1, the performance for severity assessments of wheat stripe rust achieved using lesion expansion method 3 to conduct lesion expansion was the best, the severity assessment performance achieved using lesion expansion method 2 to conduct lesion expansion ranked second, and the severity assessment performance achieved using lesion expansion method 1 to conduct lesion expansion ranked third.

As shown in Table 2, for the lesion expansion coefficients ($minrate_S$) determined using lesion expansion coefficient determination method 2, when lesion expansion processing was conducted using lesion expansion method 1, the accuracy of severity assessments of the single diseased wheat leaves in each severity class of wheat stripe rust, namely, the mean of the accuracies of severity assessments of all the single diseased leaves in each severity class of wheat stripe rust, was in the range of 85.67% to 100.00%, and the accuracy of severity assessments of all the single wheat leaves infected with wheat stripe rust was 91.10% (as shown in Figure 11). When lesion expansion processing was conducted using lesion expansion method 2, the accuracy of severity assessments of the single diseased wheat leaves in each severity class of wheat stripe rust, namely, the mean of the accuracies of severity assessments of all the single diseased leaves in each severity class of wheat stripe rust, was in the range of 90.00% to 98.75%, and the accuracy of severity assessments of all the single wheat leaves infected with wheat stripe rust was 94.60% (as shown in Figure 11). When lesion expansion processing was conducted using lesion expansion method 3, the accuracy of severity assessments of the single diseased wheat leaves in each severity class of wheat stripe rust, namely, the mean of the accuracies of severity assessments of all the single diseased leaves in each severity class of wheat stripe rust, was in the range of 89.33% to 98.57%, and the accuracy of severity assessments of all the single wheat leaves infected with wheat stripe rust was 94.84% (as shown in Figure 11). The results indicated that, overall, based on the lesion expansion coefficients ($minrate_S$) determined using lesion expansion coefficient determination method 2, the severity assessment performance for wheat stripe rust achieved when using lesion expansion method 3 to conduct lesion expansion was the best, and the severity assessment performances achieved when using lesion expansion method 2 and lesion expansion method 1 to conduct lesion expansion ranked second and third, respectively.

As shown in Table 2, for the lesion expansion coefficients ($meanrate_S$) determined using lesion expansion coefficient determination method 3, when lesion expansion processing was conducted using lesion expansion method 1, the accuracy of severity assessments of the single diseased wheat leaves in each severity class of wheat stripe rust, namely, the mean of the accuracies of severity assessments of all the single diseased leaves in each severity class of wheat stripe rust, was in the range of 78.00% to 100.00%, and the accuracy of severity assessments of all the single wheat leaves infected with wheat stripe rust was 86.05% (as shown in Figure 12). When lesion expansion processing was conducted using lesion expansion method 2, the accuracy of severity assessments of the single diseased wheat leaves in each severity class of wheat stripe rust, namely, the mean of the accuracies of severity assessments of all the single diseased leaves in each severity class of wheat stripe rust, was in the range of 82.33% to 99.75%, and the accuracy of severity assessments of all the single wheat leaves infected with wheat stripe rust was 88.73% (as shown in Figure 12). When lesion expansion processing was conducted using lesion expansion method 3, the accuracy of severity assessments of the single diseased wheat leaves in each severity class of wheat stripe rust, namely, the mean of the accuracies of severity assessments of all the single diseased leaves in each severity class of wheat stripe rust, was in the range of 83.00% to 99.50%, and the accuracy of severity assessments of all the single wheat leaves infected with wheat stripe rust was 89.03% (as shown in Figure 12). The results indicated that, overall, based on the lesion expansion coefficients ($meanrate_S$) determined using lesion expansion coefficient determination method 3, the best performance for severity assessments of wheat stripe rust was achieved using lesion expansion method 3 to conduct lesion expansion, and the severity assessment performances achieved using lesion expansion methods 2 and 1 to conduct lesion expansion ranked second and third, respectively.

The above results demonstrated that the satisfactory severity assessment results of the single diseased wheat leaves infected with *Pst* could be achieved using the proposed lesion expansion methods based on lesion expansion in this study. No matter which method was chosen from the three proposed lesion expansion coefficient determination methods to determine the lesion expansion coefficient/coefficients, satisfactory severity assessment performance for wheat stripe rust could be achieved when the three proposed lesion expansion methods were used for lesion expansion processing, respectively. Among the three lesion expansion methods, the best severity assessment performances were achieved using lesion expansion method 3 to conduct lesion expansion. Among the nine method combinations of the three lesion expansion methods and the three lesion expansion coefficient determination methods, the best severity assessment performance was achieved using the combination of lesion expansion method 3 and lesion expansion coefficient determination method 1. The results indicated that the proposed lesion expansion methods based on lesion expansion in this study can be used for severity assessments of single diseased leaves infected with wheat stripe rust.

4 Discussion

The plant disease severity assessment method proposed in this study is suitable for conducting severity assessments of any plant disease for which the actual ratio of the lesion area to the area of an investigated plant unit is much lower than the lesion area ratio corresponding to the determined severity class in the plant disease severity grading standard. By using the severity assessment method proposed in this study, based on image processing technology, the

lesion expansion coefficient/coefficients can be easily obtained, and the new expanded lesion/lesions can also be easily obtained after lesion expansion processing. In practical applications, the appropriate lesion expansion coefficient/coefficients can be selected and the appropriate lesion expansion method can be used to conduct lesion expansion of the image of a diseased plant unit, and then the area of the new obtained lesion/lesions and the area of the whole diseased plant unit can be calculated. Subsequently, the actual percentage of the lesion area in the area of the whole diseased plant unit can be calculated, and the severity of the diseased plant unit can be determined according to the corresponding severity grading standard. Therefore, once the lesion expansion coefficient/coefficients and the lesion expansion method are determined, the accurate and rapid severity assessments of the diseased plant units can be conveniently realized, and the efficiency and performance of the severity assessments can be improved, which is of great significance for the surveying, monitoring, prediction, and control of plant diseases. Moreover, the methods proposed in this study can be used to expand the lesion/lesions in the image of a diseased plant unit to make the ratio of the area of the expanded lesion/lesions to the area of the investigated plant unit reach the lesion area ratio corresponding to the severity class in the severity grading standard as far as possible. The related image processing operations can be automatically performed using image processing technology. This study provided a method and idea for severity assessments of plant diseases, which is conducive to the realization of accurate and automatic assessments of plant disease severity.

In this study, based on the fact that the actual ratio of the lesion area to the area of an investigated plant unit is much lower than the lesion area ratio corresponding to the determined severity class in the plant disease severity grading standard, a novel severity assessment method of the diseased plant units of the corresponding plant diseases was proposed. The proposed method is conducive to the accurate assessment of the severity of plant diseases and the implementation of automated and intelligent severity assessment of plant diseases. The results obtained in this study indicated that it is feasible to conduct severity assessments of plant diseases after lesion expansion processing and calculating the ratios of the lesion areas to the areas of the corresponding diseased plant units and that the key point is to select the appropriate lesion expansion method and the appropriate lesion expansion coefficient/coefficients. According to the methods provided in this study, the appropriate lesion expansion method and the appropriate lesion expansion coefficient/coefficients can be determined and selected for severity assessments of diseased plant units.

In this study, after lesion expansion processing using the method combinations of the three lesion expansion coefficient determination methods and the three lesion expansion methods, severity assessments of single diseased wheat leaves infected with *Pst* were conducted, and satisfactory severity assessment performance for wheat stripe rust was achieved. In the practical application of the methods proposed in this study for plant disease severity assessments, for convenience, lesion expansion coefficient determination method 1 can be used to determine the lesion expansion coefficient for lesion expansion processing, because the performance of the lesion expansion coefficient determined using this method was the best among the lesion expansion coefficients determined using the three lesion expansion coefficient determination methods. The results obtained in this study indicated that the severity assessment performance decreased with an increase in the lesion expansion coefficient. Because the mean of the *r* values corresponding to the single diseased wheat leaves in all the severity classes was greater than the value of Grate of 2.74, in this study, no further attempt was made to use the mean of the *r* values corresponding to the single diseased wheat leaves in all the severity classes as the lesion expansion coefficient. In further studies, the lesion expansion coefficient for the single diseased leaves in each individual severity class can be set in a certain range, and then, based on the values obtained using the enumeration method with a searching step, lesion expansion processing operations can be performed on the images of the single diseased leaves in the severity class using the lesion expansion methods described above. Subsequently, according to the processing results, the lesion expansion coefficient with the best severity assessment performance can be treated as the lesion expansion coefficient of the corresponding severity class, thus achieving a lesion expansion coefficient for each severity class. In another way, the lesion expansion coefficient can be set in a certain range, and then, based on the values obtained using the enumeration method with a searching step, lesion expansion processing operations can be performed on the images of the single diseased leaves in all the severity classes using the lesion expansion methods described above. Subsequently, the lesion expansion coefficient with the best severity assessment performance can be treated as the lesion expansion coefficient for all the severity classes. In addition, the functional relationship between *r* and the actual lesion area percentage can be established based on machine learning, and then, based on the actual lesion area percentage of each single diseased leaf, the corresponding *r* value can be calculated and can subsequently be used as the lesion expansion coefficient of the corresponding single diseased leaf.

For plant diseases in which the severity classes are divided based on the ratios of the lesion areas to the areas of the investigated plant units, the lesion expansion coefficients should be greater than or equal to 1. When the lesion area ratios corresponding to the severity classes in the severity grading standards of the plant diseases are the actual lesion area ratios, the lesion expansion coefficients can be considered to be equal to 1, and the severity classes of the diseased plant units can be determined by directly comparing the actual lesion area ratios with the lesion area ratios corresponding to the severity classes in the severity grading standards, without requiring any lesion expansion operations. When the lesion area ratios corresponding to the severity classes in the severity grading standards of the plant diseases are not the actual lesion area ratios, the lesion expansion coefficients should be greater than 1, and the disease severity assessments can be conducted according to the lesion expansion-based methods proposed in this study. The maximum values of the lesion expansion coefficients may be different in specific plant diseases.

Disease image identification technology is developing rapidly and is increasingly widely applied in plant disease surveys and

monitoring. There are many applications of deep learning technology, which has become the mainstream in plant disease image identification. However, traditional image identification technology also has its own advantages, especially when the number of plant disease images is not enough. Image segmentation has always been a hotspot and challenge in image processing research, and if the target/targets can be accurately segmented from the images, it will lay a very good foundation for target identification. In recent years, deep learning has been applied in the field of plant disease image segmentation (Zhang et al., 2020; Li et al., 2022). In future studies, image segmentation methods based on deep learning can be used to carry out the lesion image segmentation, and then the number of the segmented lesion pixels and the number of pixels in the corresponding single diseased leaf or the area of segmented lesion/lesions and the area of the corresponding single diseased leaf can be obtained. Subsequently, the severity of the corresponding single diseased leaf can be assessed, according to the corresponding severity grading standard.

The determination of the lesion boundary/boundaries in the image of a single diseased leaf is a very important step in the severity assessment of the single diseased leaf using the severity assessment methods proposed in this study. When the boundary of a lesion is determined, the distances between the centroid of the lesion and the points on the boundary of the lesion can be determined, and then, based on the distances and the determined lesion expansion coefficient, the lesion can be expanded outward. In this study, two methods were used to determine the lesion boundary/boundaries, and thus two lesion expansion methods (i.e., lesion expansion method 1—the external edge-based lesion expansion method and lesion expansion method 2—the internal edge-based lesion expansion method) were developed. The results obtained in this study indicated that satisfactory severity assessment performance can be achieved using the severity assessment methods based on the two lesion expansion methods. The severity assessment method based on lesion expansion method 2 outperformed that based on lesion expansion method 1, i.e., the severity assessment performance obtained after lesion expansion processing using the internal edge-based lesion expansion method was better than that obtained after lesion expansion processing using the external edge-based lesion expansion method.

When using a lesion expansion method to process the lesion images of plant diseases, there may be situations in which the expansion range of a lesion may exceed the boundary of the diseased plant unit, and the expanded lesions are connected together. After lesion expansion, the expanded lesion region/regions cannot exceed the boundary of the corresponding diseased plant unit. However, in practice, after a lesion is processed using a lesion expansion method, the new expanded lesion may exceed the boundary of the corresponding diseased plant unit, especially for the lesions that are originally located at or near the edge of the corresponding diseased plant unit. Therefore, the region range of the new expanded lesion should be limited based on the segmented plant unit image obtained before lesion expansion processing and the new lesion image obtained after lesion expansion processing, thus avoiding the region/regions exceeding the boundary of the plant unit that may be included into the new expanded lesion region when calculating lesion area. After lesion expansion processing, the adjacent lesions may be connected together. For the lesions connected together after lesion expansion, methods such as the watershed segmentation method (Li et al., 2013b) can be used to separate the connected lesions. If the connected lesions cannot be separated, multiple lesions connected together can be treated as one lesion.

There are two main ideas based on image processing technology to conduct plant disease severity assessments (Jiang et al., 2022, 2023, 2024). One idea, as described above, is to obtain the actual ratio of the lesion area to the area of an investigated diseased plant unit and then to compare the actual lesion area ratio with the lesion area ratios corresponding to all the severity classes in the disease severity grading standard to determine the severity class of the investigated diseased plant unit (Li et al., 2011, 2013a; Pethybridge and Nelson, 2015; Shrivastava et al., 2015; Jiang et al., 2021, 2022, 2023, 2024; Olivoto et al., 2022; Yao et al., 2024). The other idea is to obtain the features of the investigated diseased plant units and then build a plant disease severity assessment model based on the obtained features to determine the severity classes of the diseased plant units to be assessed (Bao et al., 2021; Jiang et al., 2022, 2023, 2024). However, when using the former, it is easy to make incorrect severity assessments for the plant diseases with the severity grading standards in which the lesion area ratios corresponding to all the severity classes are not the actual lesion area ratios (Jiang et al., 2022, 2023, 2024). Therefore, to solve this problem, based on the actual lesion area ratios, Jiang et al. (2022, 2023) proposed four severity assessment methods as described above, and a novel method was proposed in this study. In particular, for wheat stripe rust, satisfactory severity assessment results were achieved using the methods proposed by Jiang et al. (2022, 2023) and the novel lesion expansion-based methods proposed in this study, although some parameters used in these methods can be further optimized to obtain better severity assessment results. Rapidly developing deep learning technology can be used to build plant disease severity assessment models with the severity classes as different categories, and it is necessary to explore appropriate deep learning models for application. These methods can be used to develop severity assessment systems based on personal computers, Internet, and mobile terminals to facilitate plant disease severity assessments in practice.

5 Conclusion

In this study, a novel method based on lesion expansion to assess plant disease severity was proposed, and then, by taking wheat stripe rust as an example, the lesion expansion-based severity assessment methods for wheat stripe rust were proposed by combining three lesion expansion methods with three lesion expansion coefficient determination methods. Based on image processing technology, after lesion expansion processing of the single diseased wheat leaves using the nine method combinations of the three lesion expansion methods and the three lesion expansion coefficient determination methods, the area of each whole single diseased leaf and the area of the corresponding new obtained lesion/lesions after lesion expansion

were calculated, respectively, and then, the actual percentage of the lesion area in the area of the corresponding whole single diseased leaf after lesion expansion was calculated. Subsequently, the obtained actual lesion area percentage of each single diseased leaf was compared with the lesion area percentage for each severity class in the disease severity grading standard of wheat stripe rust, and finally, the severity of each single diseased leaf was determined. The results showed that the accuracy of severity assessments of the single diseased wheat leaves in each severity class of wheat stripe rust (namely, the mean of the accuracies of severity assessments of all the single diseased leaves in each severity class of wheat stripe rust) was in the range of 78.00% to 100.00%, indicating satisfactory performance for severity assessments of the single diseased leaves infected with wheat stripe rust were achieved using the proposed methods based on lesion expansion. No matter which lesion expansion coefficient determination method was used, the severity assessment performance for the single diseased leaves infected with wheat stripe rust achieved using lesion expansion method 3 to conduct lesion expansion was the best, among the three lesion expansion methods. An accuracy of 96.16% for severity assessments of all the single wheat leaves infected with wheat stripe rust was achieved using the method combination of lesion expansion method 3 and lesion expansion coefficient determination method 1, indicating that this method combination is optimal among the nine method combinations of the three lesion expansion methods and the three lesion expansion coefficient determination methods. The results demonstrated that it is feasible to assess the severity of single diseased leaves infected with wheat stripe rust using the lesion-expansion-based severity assessment methods proposed in this study. This study provided a feasible solution to solve the problem of severity assessments for all the plant diseases in which the actual ratio of the lesion area to the area of an investigated plant unit is much lower than the lesion area ratio corresponding to the determined severity class in the plant disease severity grading standards, and provided a methodological reference for the accurate assessment of plant diseases.

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

FQ: Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. HGW: Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision. QJ: Investigation, Writing – review & editing. HLW: Investigation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Key Basic Research Program of China (Grant No. 2021YFD1401001) and the National Natural Science Foundation of China (Grant No. 32072357).

Conflict of interest

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

Generative AI statement

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

Publisher's note

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

References

Bao, W. X., Lin, Z., Hu, G. S., Liang, D., Huang, L. S., and Yang, X. J. (2021). Severity estimation of wheat leaf diseases based on RSTCNN. *Trans. Chin. Soc Agric. Mach.* 52, 242–252, 263. doi: 10.6041/j.issn.1000-1298.2021.12.026

Barbedo, J. G. A. (2014). An automatic method to detect and measure leaf disease symptoms using digital image processing. *Plant Dis.* 98, 1709–1716. doi: 10.1094/PDIS-03-14-0290-RE

Bock, C. H., Chiang, K. S., and Del Ponte, E. M. (2022a). Plant disease severity estimated visually: A century of research, best practices, and opportunities for improving methods and practices to maximize accuracy. *Trop. Plant Pathol.* 47, 25–42. doi: 10.1007/s40858-021-00439-z

Bock, C. H., Pethybridge, S. J., Barbedo, J. G. A., Esker, P. D., Mahlein, A. K., and Del Ponte, E. M. (2022b). A phytopathometry glossary for the twenty-first century: Towards consistency and precision in intra- and inter-disciplinary dialogues. *Trop. Plant Pathol.* 47, 14–24. doi: 10.1007/s40858-021-00454-0

Huang, M. Y., Huang, W. J., Liu, L. Y., Huang, Y. D., Wang, J. H., Zhao, C. J., et al. (2004). Spectral reflectance feature of winter wheat single leaf infected with stripe rust

and severity level inversion. *Trans. Chin. Soc Agric. Eng.* 20, 176–180. doi: 10.3321/j.issn:1002-6819.2004.01.042

Jiang, X. M., Feng, H. K., Chang, H., Yang, G. J., and Yang, X. D. (2021). Classification method of wheat stripe rust disease degree based on digital image. *Jiangsu Agric. Sci.* 49, 109–115. doi: 10.15889/j.issn.1002-1302.2021.23.020

Jiang, Q., Wang, H. L., and Wang, H. G. (2022). Two new methods for severity assessment of wheat stripe rust caused by *Puccinia striiformis* f. sp. *tritici*. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1002627

Jiang, Q., Wang, H. L., and Wang, H. G. (2023). Severity assessment of wheat stripe rust based on machine learning. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1150855

Jiang, Q., Wang, H. L., and Wang, H. G. (2024). Development of an automatic grading system of wheat stripe rust severity based on image processing. *Acta Phytopathol. Sin.* 54, 385–397. doi: 10.13926/j.cnki.apps.000910

Li, G. L., Ma, Z. H., and Wang, H. G. (2011). An automatic grading method of severity of single leaf infected with grape downy mildew based on image processing. *J. China Agric. Univ.* 16, 88–93. doi: 10.11841/j.issn.1007-4333.2011.06.014

Li, G. L., Ma, Z. H., and Wang, H. G. (2013a). Development of a single-leaf disease severity automatic grading system based on image processing. *Lect. Notes Elect. Eng.* 212, 665–675. doi: 10.1007/978-3-642-34531-9_70

Li, X. L., Ma, Z. H., Sun, Z. Y., and Wang, H. G. (2013b). Automatic counting for trapped urediospores of Puccinia striiformis f. sp. *tritici* based on image processing. *Trans. Chin. Soc Agric. Eng.* 29, 199–206. doi: 10.3969/j.issn.1002-6819.2013.02.028

Li, X. L., Qin, F., Zhao, L. L., Li, J. H., Ma, Z. H., and Wang, H. G. (2015). Identification and classification of dsease severity of wheat stripe rust using near infrared spectroscopy technology. *Spectrosc. Spect. Anal.* 35, 367–371. doi: 10.3964/j.issn.1000-0593(2015)02-0367-05

Li, Y., Qiao, T. L., Leng, W. B., Jiao, W. R., Luo, J., Lv, Y., et al. (2022). Semantic segmentation of wheat stripe rust images using deep learning. *Agronomy* 12, 2933. doi: 10.3390/agronomy12122933

Nutter, F. W.Jr., Teng, P. S., and Shokes, F. M. (1991). Disease assessment terms and concepts. *Plant Dis.* 75, 1187–1188.

Olivoto, T., Andrade, S. M. P., and Del Ponte, E. M. (2022). Measuring plant disease severity in R: Introducing and evaluating the pliman package. *Trop. Plant Pathol.* 47, 95–104. doi: 10.1007/s40858-021-00487-5

Pethybridge, S. J., and Nelson, S. C. (2015). Leaf Doctor: A new portable application for quantifying plant disease severity. *Plant Dis.* 99, 1310–1316. doi: 10.1094/PDIS-03-15-0319-RE

Shang, H. S., Ren, W. L., and Wang, S. M. (1990). A standard grading for wheat stripe rust disease. *Plant Prot.* 16, 31.

Shrivastava, S., Singh, S. K., and Hooda, D. S. (2015). Color sensing and image processing-based automatic soybean plant foliar disease severity detection and estimation. *Multimed. Tools Appl.* 74, 11467–11484. doi: 10.1007/s11042-014-2239-0

Wang, H. G. (2022). Smart phytoprotection and suggestions for its development. *J. China Agric. Univ.* 27, 1–21. doi: 10.11841/j.issn.1007-4333.2022.10.01

Wang, H. G. (2023). Epidemiology and control of fungal diseases in crop plants. *Agronomy* 13, 2327. doi: 10.3390/agronomy13092327

Wang, H. G., Guo, J. B., and Ma, Z. H. (2012). "Monitoring wheat stripe rust using remote sensing technologies in China," in *Computer and Computing Technologies in Agriculture V. Part III, IFIP Advances in Information and Communication Technology*, vol. 370 . Eds. D. L. Li and Y. Y. Chen (Springer, Berlin, Heidelberg), 163–175. doi: 10.1007/978-3-642-27275-2_18

Wang, H. G., Ma, Z. H., Wang, T., Cai, C. J., An, H., and Zhang, L. D. (2007). Application of hyperspectral sata to the classification and identification of severity of wheat stripe rust. *Spectrosc. Spect. Anal.* 27, 1811–1814. doi: 10.3964/j.issn.1000-0593.2007.09.037

Wang, H., Qin, F., Ruan, L., Wang, R., Liu, Q., Ma, Z. H., et al. (2016). Identification and severity determination of wheat stripe rust and wheat leaf rust based on hyperspectral data acquired using a black-paper-based measuring method. *PLoS One* 11, e0154648. doi: 10.1371/journal.pone.0154648

Xiao, Y. Y. (1997). The studies on the reliability evaluation of pest prediction and forecasting. *Plant Prot. Tech. Ext.* 17, 3–6.

Yao, H., Wang, C. S., Zhang, L. J., Li, J. X., Liu, B., and Liang, F. F. (2024). A cucumber leaf disease severity grading method in natural environment based on the fusion of TRNet and U-Net. *Agronomy* 14, 72. doi: 10.3390/agronomy14010072

Zhang, S. W., Wang, Z., and Wang, Z. L. (2020). Method for image segmentation of cucumber disease leaves based on multi-scale fusion convolutional neural networks. *Trans. Chin. Soc Agric. Eng.* 36, 149–157. doi: 10.11975/j.issn.1002-6819.2020.16.019

Zhao, J. L., Huang, L. S., Huang, W. J., Zhang, D. Y., Yuan, L., Zhang, J. C., et al. (2014). Hyperspectral measurements of severity of stripe rust on individual wheat leaves. *Eur. J. Plant Pathol.* 139, 407–417. doi: 10.1007/s10658-014-0397-6

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 27 March 2025
DOI 10.3389/fpls.2025.1540722

Check for updates

OPEN ACCESS

EDITED BY
Jun Ni,
Nanjing Agricultural University, China

REVIEWED BY
Pei Wang,
Southwest University, China
Sunil Gc,
North Dakota State University, United States
Changkai Wen,
China Agricultural University, China

*CORRESPONDENCE
Shuo Zhang
zhangshuo@nwafu.edu.cn

RECEIVED 06 December 2024
ACCEPTED 06 March 2025
PUBLISHED 27 March 2025

CITATION
Wang H, Chen Y, Zhang S, Guo P, Chen Y, Hu G and Ma Y (2025) Deep learning-based target spraying control of weeds in wheat fields at tillering stage. *Front. Plant Sci.* 16:1540722. doi: 10.3389/fpls.2025.1540722

COPYRIGHT
© 2025 Wang, Chen, Zhang, Guo, Chen, Hu and Ma. 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.

Deep learning-based target spraying control of weeds in wheat fields at tillering stage

Haiying Wang[1], Yu Chen[1], Shuo Zhang[1]*, Peijie Guo[1], Yuxiang Chen[1], Guangrui Hu[2] and Yuxuan Ma[1]

[1]College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling, China, [2]School of Design, Xi'an Technological University, Xian, China

In this study, a target spraying decision and hysteresis algorithm is designed in conjunction with deep learning, which is deployed on a testbed for validation. The overall scheme of the target spraying control system is first proposed. Then YOLOv5s is lightweighted and improved. Based on this, a target spraying decision and hysteresis algorithm is designed, so that the target spraying system can precisely control the solenoid valve and differentiate spraying according to the distribution of weeds in different areas, and at the same time, successfully solve the operation hysteresis problem between the hardware. Finally, the algorithm was deployed on a testbed and simulated weeds and simulated tillering wheat were selected for bench experiments. Experiments on a dataset of realistic scenarios show that the improved model reduces the GFLOPs (computational complexity) and size by 52.2% and 42.4%, respectively, with mAP and F1 of 91.4% and 85.3%, which is an improvement of 0.2% and 0.8%, respectively, compared to the original model. The results of bench experiments showed that the spraying rate under the speed intervals of 0.3-0.4m/s, 0.4-0.5m/s and 0.5-0.6m/s reached 99.8%, 98.2% and 95.7%, respectively. Therefore, the algorithm can provide excellent spraying accuracy performance for the target spraying system, thus laying a theoretical foundation for the practical application of target spraying.

KEYWORDS

weed identification, weed distribution determination, hysteresis property, target spraying, deep learning

1 Introduction

Wheat stands as a globally cultivated cereal crop, encompassing over 22.067 billion hectares planted annually across a spectrum of climatic conditions and diverse geographic locales (Shiferaw et al., 2013), surpassing all other cultivated food crops in scale (Khan et al., 2024). Despite its vast reach, the menace of weeds looms large, imperiling optimal crop yields by vying with crops for vital resources like water, nutrients, and sunlight (Horvath et al., 2023), leading to yield reductions in wheat grains of up to 24% (Jabran et al., 2017). Herbicides have emerged as pivotal agents fostering crop yield proliferation since the latter half of the 20th century (Cassman, 1999).Amidst the efficiency of herbicides in directly combating weeds, farmers frequently resort

to substantial herbicide applications to safeguard against yield decline and compromised agricultural product quality. Nonetheless, the extensive use of agrochemicals exacts a toll on environmental integrity, biodiversity, and human well-being (Sharma et al., 2019). In a departure from indiscriminate broadcast spraying, targeted spraying techniques leverage the discernment of target size and location within fields to regulate the activation of corresponding nozzles, offering a potent solution to challenges like herbicide excess and fluid wastage (Yu et al., 2019).

In recent years, the application of machine vision technology in agricultural production has been widely studied in all aspects of agricultural production (Lv et al., 2024). Traditional machine vision methods usually use manually designed feature extraction methods, which are difficult to capture the complex features of weeds and lack flexibility and generalization. Deep learning (DL) techniques are an emerging artificial intelligence approach that allows for accurate and fast object detection by automatically learning and understanding low-level to high-level image features based on a conceptual hierarchy (LeCun et al., 2015). Recent developments in deep learning have been shown to be effective in a variety of crop management operations (Brown et al., 2008). Modi et al. (2023) used a deep learning model to identify weeds in sugarcane fields and found that the DarkNet53 model outperformed the other models by showing high accuracy and F1 scores in identifying weeds in sugarcane crops. Yuan et al. (2023) constructed a WeedyRice5 target detection model based on the YOLOv5s method for detecting weeds in rice. Shao et al. (2023) proposed an improved deep learning model, GTCBS-YOLOv5s, utilizing Ghost, C3Trans, and CBAM modules that improved weed feature extraction level. Zhu H. et al. (2024) connected a lightweight attention module to the deep network of YOLOX-Darknet, which weakened the channel noise effect of residual computation and made the detection model more efficient. Deng et al. (2024) INet-based model improved the accuracy of YOLOX and YOLOv8 compared to the original model by respectively 1.4% and 3.3%, respectively. Rai et al. (2024) optimized and reconstructed the YOLO-Spot model constructed based on YOLOv7-tiny to identify weeds in crop plants, and the optimized model could use a smaller number of training parameters to identify weeds. Alrowais et al. (2022) combined the HLBODL-WDSA model combined with YOLOv5 for weed detection, where the HLBO algorithm was used as a hyper-parameter optimizer to efficiently classify weeds using a Kernel Extreme Learning Machine (KELM) model. Rahman et al. (2023) used YOLOv5, RetinaNet, EfficientDet, Fast RCNN, and Faster RCNN, to 13 weed detection models were constructed by transferring pre-trained target detection models to the weed dataset, and the data were enhanced by geometric and color transformations. Although the YOLO algorithm has excellent recognition results, it is difficult to be directly deployed on in-vehicle devices with limited computational power due to its high requirements on device arithmetic. Improvements that reduce the computational complexity of the model and increase its accuracy are particularly important given the need for high accuracy in weed detection.

In the area of targeted spraying, Upadhyay et al. (2024) designed and developed a machine vision-based spraying system specifically designed to identify weeds and enable precision spraying. The spraying platform employs a deep-learning YOLOv4 model to accurately recognize a wide range of weeds, thereby facilitating targeted spraying applications. The system is equipped with a FLIR RGB camera for real-time image acquisition, while an Nvidia Jetson AGX Orin is used as an edge device to deploy the deep learning model for weed detection. The Nvidia Jetson pin is used to activate a relay for precise on/off control of a solenoid valve for spot spraying. In the indoor experiment, the spray system achieved an average effective spray rate of 93.33% with 100% accuracy, while the recall rate was 92.8%. In contrast, the field experiment had a slightly lower average effective spray rate of 90.6%, but still maintained an accuracy rate of 95.5% and a recall rate of 89.47%. Sunil et al. (2024) developed a weed spraying robotic platform using the NVIDIA Jetson embedded device as the control and processing unit and tested it under laboratory and field conditions. Fan et al. (2023) developed a deep-learning-based seedling weed detection and on-target spraying robotic system for cotton fields, which is capable of field weed detection and herbicide spraying, laying the foundation for targeted spraying in weed control. However, target spraying robots are expensive and cannot be popularized in large-scale field crops due to their high R&D costs and low operational efficiency. Compared to target spraying robots, target spraying systems mounted on sprayers are less expensive, and target spraying systems can be deployed on existing sprayers relatively easily without major modification or redesign of existing sprayers., the spraying volume of the sprayer equipped with the on-target spraying system was reduced by 40% and the ground deposition was reduced by 41%, and the sprayer of this experiment was operated at a constant speed. Li et al. (2022a) designed a grid decision algorithm for the switching of solenoid valves group to convert the single weed position information in the image into the opening and closing control information of solenoid valves. The current target spraying system mainly focuses on the identification and localization of single plants, while effective treatment solutions in the case of complex plant distribution still need further research. Meanwhile, the hysteresis between target spraying hardware remains a major challenge in target spraying.

The research content and contributions of this paper include: first, the overall scheme of the target spraying control system is proposed. Then, Yolov5s is lightened and improved to make the model balanced in terms of weed recognition accuracy and computational complexity. Based on this, a target spraying decision-making and hysteresis algorithm was designed to control the solenoid valve to differentiate spraying for weeds with different distributions after completing the weed identification, and effectively solved the hysteresis problem existing between spraying hardware. A target spraying control system test bed was developed, and the integrated algorithm was deployed on the test bed for target spraying experiments. The experiment shows that the algorithm has high accuracy in spraying weeds with different distribution and can effectively reduce the waste of herbicides.

2 Materials and methods

2.1 Overall program design

In order to achieve precise weed control in wheat fields at tillering stage, this study developed a pair-target control system that can detect

and spray multiple weeds randomly distributed in wheat fields in real time and on-target, so that the utilization rate of herbicides is improved. A single target spraying control system includes an image acquisition and real-time speed measurement module, a signal conversion module, a spray execution module, and a herbicide supply module. The system control principle is shown in Figure 1.

The image acquisition and real-time speed measurement module consists of an RGB camera, a GNSS antenna and a computer. The RGB camera is responsible for capturing images of weeds in the wheat field. The sprayer operation can use dual GNSS antennas for differential speed measurement, and the frequency of the feedback speed is 10Hz.The computer is responsible for processing the image information from the RGB camera and the speed data from the GNSS antenna, and sends target spraying commands to the Arduino. The RGB camera is a W200 model from Jierui Microcomputer Technology Co., Ltd (Shenzhen, China), with a 2.9MM distortion-free 130-degree wide-angle and 200W pixels. The frame rate is 60 frames per second, the power is 2W, the working voltage is 5V, and the working current is 120mA~220mA.

The signal conversion module consists of an Arduino, a 12V battery and ten MOS electronic switch control boards. The ten pins of the Arduino are connected to the signal terminals of the ten MOS electronic switch control boards. The Arduino is responsible for converting the target spraying commands sent by the computer into the high and low levels of the pins. Since the high level of the Arduino pins is 5V, it cannot directly drive the 12V solenoid valve to work. Therefore, by connecting a 12V battery to the input of the MOS electronic switch control board, when the level of the signal terminal is high, the output of the MOS electronic switch control board will output 12V current to drive the solenoid valve to open; when the signal terminal is low, the solenoid valve is closed.

The spray actuator module consists of 10 solenoid valves and 10 spray nozzles. The positive and negative poles of each solenoid valve are connected to the output of the respective MOS tube electronic switch control board. The solenoid valve model is HL22-02 of Honglian Control Flow Technology Development Co., Ltd (Ningbo, China), with a pressure of 8MPa, G1/4 (2 points) interface, and a voltage of DC12V. A stainless steel fan nozzle, model 3004, with a spray angle of 30 degrees and an aperture of 1.3mm, from Geo Industrial Spray Equipment Co. The flow rate of the nozzle was 1.2 L/min at an operating pressure of 0.8 MPa. The maximum diameter of weeds in a wheat field at the tillering stage is usually no more than 10 cm, and because multiple weeds are often clustered in a wheat field, the diameter of the patches formed can be 20 cm or even larger. Therefore, at a height of 42cm, the spray coverage length using a 30° nozzle is 23cm, with a 1.5cm spray overlap area on each side of the nozzle to ensure that leakage is prevented. The nozzles are fan-shaped and are connected to a solenoid valve.

The herbicide supply module consists of a water tank, a pump, a check valve and a safety valve. The pressure of the safety valve is set to

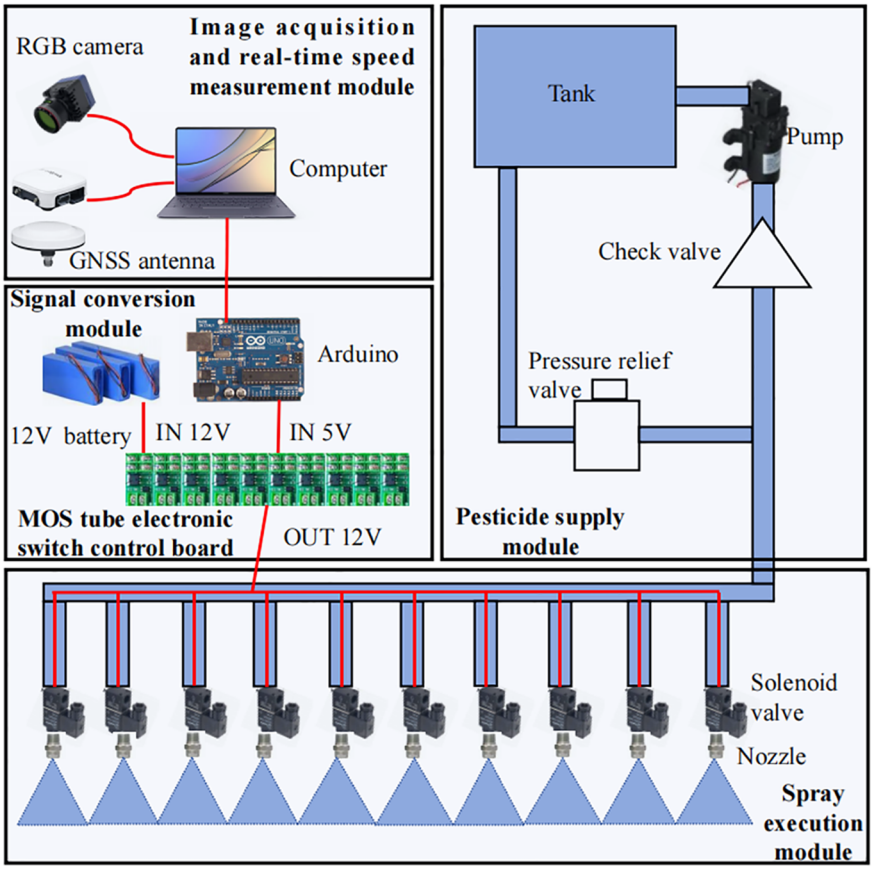


FIGURE 1
Control principle diagram of single pair target spraying system.

0.8Mpa. When the pressure of the pipeline exceeds 0.8Mpa, the safety valve will automatically release the pressure and discharge the excess liquid into the water tank to protect the pipeline. The one-way valve is used to prevent the liquid from flowing back. When expanding multiple pairs of target spraying control systems, it is possible to realize that multiple spraying execution modules share one herbicide supply module according to the working parameters of the pump.

Multiple pairs of target spray control systems can be assembled on the spraying machine, as shown in Figure 2a. The structure of a single counter-target spray system test bed is shown in Figure 2b, including an RGB camera mounted at a height of 1.5 m from the upper surface of the conveyor belt, with a shooting area width of 2 m. The camera is mounted at a height of 1.5 m from the upper surface of the conveyor belt. Ten solenoid valves were mounted on the spray bar below the camera, and the height of the spray nozzles from the upper surface of the conveyor belt was 42 cm, and the spray coverage of each nozzle was about 24 cm. Simulated weeds and simulated wheat were placed on the conveyor belt, and the motor drove the simulated weeds and simulated wheat on the conveyor belt to move in the direction v in Figure 2 to simulate the operation of the sprayer in the field.

2.2 Weed data set collection and processing

Within Shaanxi Province, weeds in wheat fields are mainly dominated by Cruciferae and Gramineae, and some of the more damaging weeds include Silene conoidea, Malcolmia africana, Descurainia sophia, and Capsella bursa-pastoris. In this study, crop and weed growth in wheat field fields at the tillering stage was collected in Yangling District and Heyang County, Shaanxi Province, and the data were collected at the tillering stage of wheat (when the leaves grew to four to six). Part of the dataset is shown in Figure 3.

In order to ensure the diversity of the produced dataset, 1012 images in the environments of complex, dense, uneven light and shade backgrounds were selected as the raw data, respectively. And Matlab software was used to add pretzel noise and Gaussian noise to some of the datasets for image processing, which helps to improve the robustness of the algorithm by introducing noise to reduce the clarity of the image. The dataset images are expanded to 1600 images by image enhancement. The dataset types are shown in Table 1.

2.3 Improved weed detection model based on Yolov5s

In terms of lightweight deep learning models for identifying weeds, there are mainly single-stage networks and two-stage networks, and the average accuracy mAP of single-stage networks such as SSD, RetinaNet, EfficientDet, CenterNe, etc., in identifying weeds needs to be further improved (Saleem et al., 2022), and the RFCN and Faster RCNN as represented by RFCN, Faster RCNN

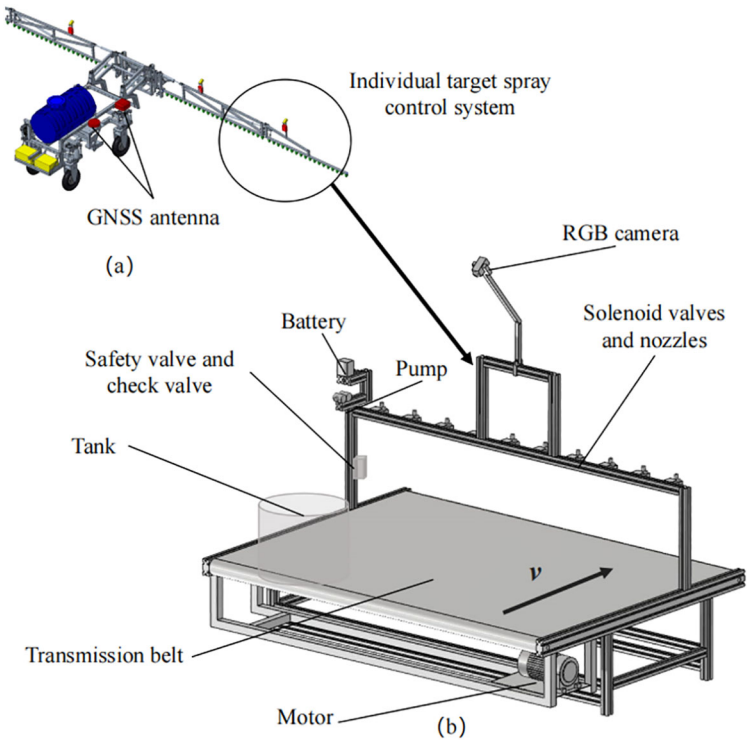


FIGURE 2
Structural schematic diagram: **(a)** a schematic diagram of the structure of a sprayer equipped with a plurality of pairs of target spray control systems, and **(b)** a schematic diagram of the structure of a single target spray control system test bed.

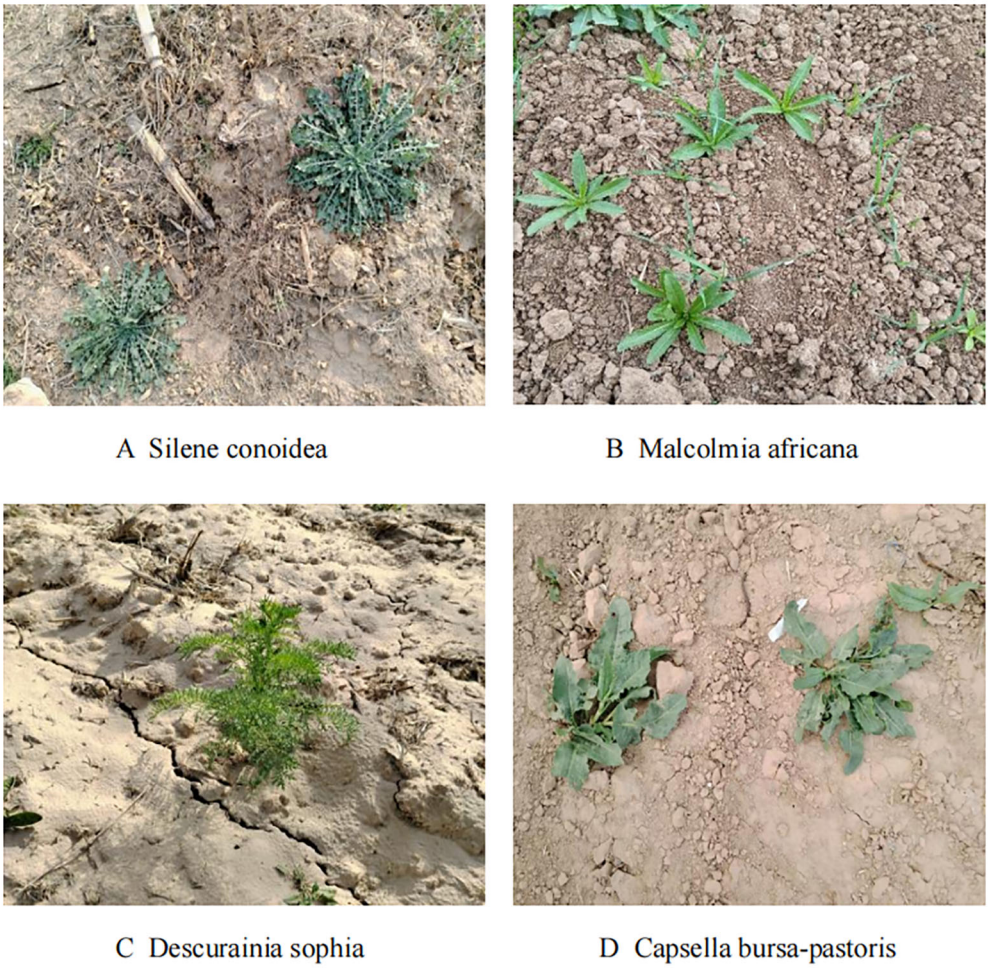


FIGURE 3
Example of weeds in wheat field at tillering stage **(A)** Silene conoidea **(B)** Malcomia africana **(C)** Descurainia sophia **(D)** Capsella bursa-pastoris.

need to be further optimized in terms of computational complexity and inference time (Fan et al., 2023). And Yolo series is an algorithm for fast object recognition using neural networks. The Yolo series algorithms are a fast object recognition method based on neural networks, and their versions are Yolov9, Yolov8, Yolov7, Yolov5, Yolov4, and Yolov3 according to the time of their release. However, the newer is not always the better when choosing the version of a model. Among the models in the same class, Yolov9s (26.7 GFLOPs) and Yolov8s (28.6 GFLOPs) excel in recognition accuracy, but their GFLOPs (computational complexity) are much higher than those of Yolov5s (15.8 GFLOPs). Therefore, when choosing a version of the Yolo algorithm, it is necessary to make a judgment based on a specific recognition scenario.Yolov5 contains five different versions: Yolov5n, Yolov5s, Yolov5m, Yolov5l, and Yolov5x.The main differences between these versions are the depth and width of the model, which is also known as its complexity and performance. The smaller the model, the lower the computational complexity and the faster it runs, but the detection accuracy decreases. For weed target recognition in the field, the model should have the characteristics of high accuracy, low computational complexity and fast recognition.Yolov5s has the advantages of fast speed and high accuracy and shows good processing ability in the case of overlapping targets (Zhu A. et al., 2024), but Yolov5s is not sufficiently accurate in recognizing some of the small weeds. To further reduce the computational complexity of Yolov5s, which can be deployed on vehicle-mounted devices with limited computational power, this study improves Yolov5s by lightweighting and enhancing its accuracy.

TABLE 1 Types of data sets.

Data types	Number of data
Complex environmental context	236
Dense distribution	212
Uneven lighting	296
Occlusion	268
Pepper noise image enhancement	283
Gaussian noise image enhancement	305

2.3.1 Lightweight backbone network construction

Yolov5s uses C3 and Conv as the backbone feature extraction network, while an image after neural network feature extraction generates many feature maps, some of these feature maps are very similar, increasing the number of parameters and making it slower. While Ghost Module (Han et al., 2020) can perform a simple operation on one of the feature maps to generate more similar feature maps. In order to be able to ensure efficient operation and reduce repeated calculations, therefore, this study uses the lightweight model Ghost Module to construct the backbone network.

2.3.2 Neck network construction

In Yolov5s, the neck is the part that connects the backbone to the detection and is responsible for feature fusion and processing. In the scenario of real-time weed detection, traditional large-scale models are difficult to meet the requirements of real-time detection, while lightweight models constructed from a large number of depth-separable convolutional layers are also unable to achieve sufficient accuracy while ensuring real-time performance. Therefore, this study uses a special lightweight convolution technique GSConv (Li et al., 2022b). GSConv first downsamples the input, then uses DWConv deep convolution on the downsampled result, and splices the downsampled result with the deep convolution, and finally performs a shuffle operation and outputs it, which reduces the parameters while as much as possible The connection between the neck and the backbone is preserved, so that the feature map no longer needs to be transformed when it is transmitted to the neck, and there are fewer redundant repetitions. The cross-level network module VoVGSCSP is designed based on GSConv using the one-time aggregation method, and combined with GSConv to form the Slim-Neck architecture.

2.3.3 Self-attention mechanism

Most of the existing attention mechanisms use additional sub-networks to generate the attention weights, which increases the number of parameters of the neural network, so this study uses a self-attention mechanism SimAM (Yang et al., 2021), which does not need to add parameters to the original network. Its input data size is $c \times h \times w$, the 3D weights of the feature maps are inferred by the energy function, and the weights normalized by the Sigmoid function are multiplied with the original feature maps to get the output feature maps of the boosted features, so that the model focuses more on the important parts and improves the performance of detecting the target.

2.3.4 Improved model

In this study, we propose a lightweight Yolov5-SGS model. First, we replace C3 and Conv in Yolov5s feature extraction backbone network with the lightweight C3Ghost and GhostConv. Then, C3 and Conv in the original neck network are replaced with the more lightweight and efficient GSConv and VoVGSCSP, respectively, to speed up the feature fusion and processing. Finally, the SimAM self-attention module is added to each of the three outputs connected to the feature extraction backbone network and the neck network to process the outputs of the forward transfer layer, increase the distinction between weeds and complex backgrounds, and improve the weed feature salience. The improved model network structure is shown in Figure 4.

2.4 Target spraying decision and hysteresis algorithms

In the target spraying system, the key to controlling the spray nozzle for target spraying lies in controlling the opening and closing moments of the solenoid valves and their durations. Li He et al. (2022) designed a grid decision algorithm for the switching of solenoid valves, which converts the position information of a single weed in the image into the opening and closing control information of the solenoid valves. For a single weed, the solenoid valve can be opened and closed normally in the response frequency to complete the target spraying. However, the distribution of weeds in the wheat field is random, for the distance is very close to the multi-plant weeds, the solenoid valve can not respond very quickly, which will lead to the liquid can not spray the neighboring weeds. Therefore, the opening and closing frequency of the solenoid valve determines the minimum distance between multiple weeds that can be sprayed, and the length of the weeds determines the duration of the solenoid valve opening. In the butt-target spraying control system, there is a hardware system delay problem from the completion of the butt-target spraying decision to the control of the solenoid valve spraying, as well as an unavoidable hysteresis distance between the area where the camera recognizes the weeds and the spray nozzle. We have investigated the above problems and designed a target spraying decision and hysteresis algorithm.

2.4.1 Algorithm for decision making on target spraying

After the Yolov5-SGS completed the weed identification, the target spraying decision-making algorithm was used to differentiate the weeds with different distributions. The target spraying decision algorithm is shown in Figure 5.

Here the numbers 1 to 10 represent the serial numbers of the corresponding nozzles and the arrays $[X]$ are the parameters in the hysteresis model.

$$[X] = \begin{bmatrix} A_a & B_b & C_c & D_d & E_e & F_f & G_g & H_h & I_i & J_j \\ M_m & O_o & P_p & Q_q & R_r & S_s & U_u & W_w & Y_y & Z_z \end{bmatrix}$$

First, the image captured by the camera was divided into ten regions, as shown by the grid divided by the black double dashed lines in the figure, with the length of each grid corresponding to the spraying range of its nozzle. When the sprayer works, the image moves from top to bottom and Yolov5-SGS generates a real-time red solid line prediction box for weeds.

Secondly, for the case of a single weed, when the red solid line prediction box touches the grid, the corresponding area of the grid will change from 0 to 1. At the same time, as long as there is a red solid line prediction box in the grid, the judgment value will remain

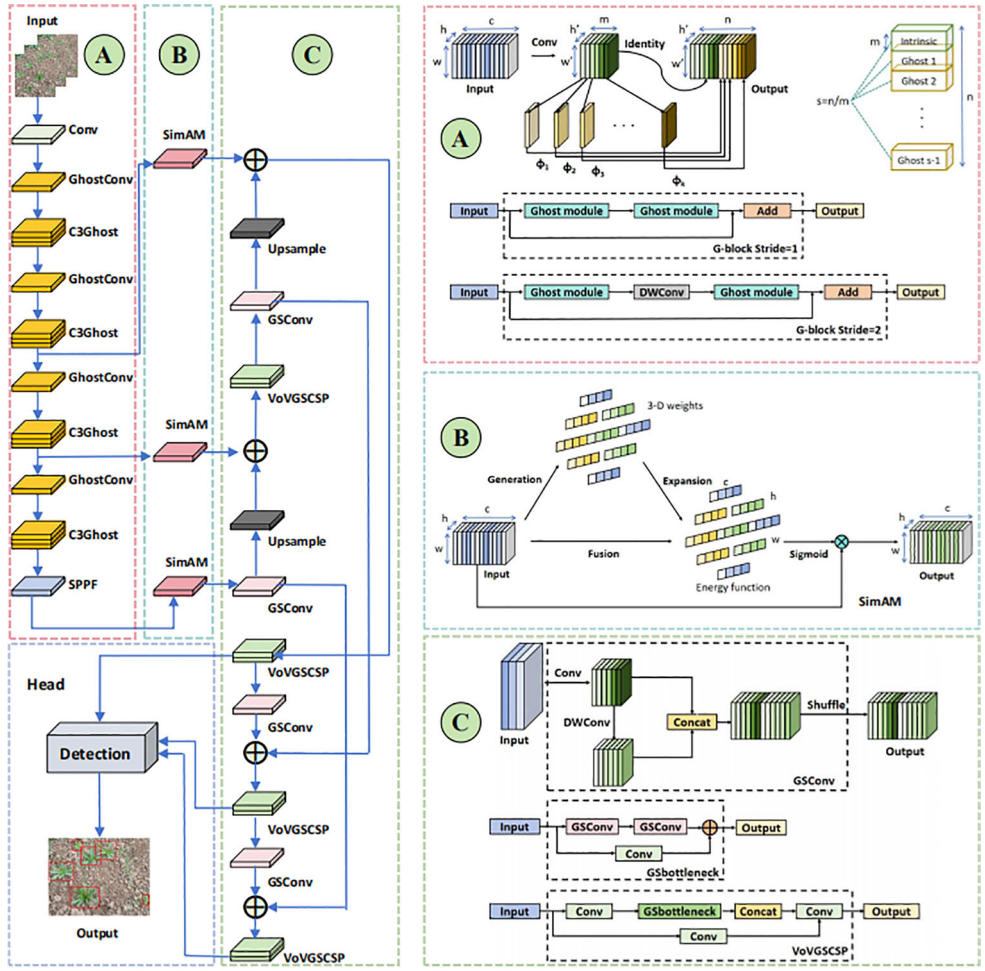


FIGURE 4
(A–D) Yolov5-SGS model network structure.

at 1 until all the red solid line prediction boxes have left the grid, and then the judgment value will change from 1 to 0.

In addition, for the distribution of multiple weeds, there are four scenarios: weeds too close together vertically, weeds too close together horizontally, multiple weeds too close together, and weeds overlapping. The blue dashed box is the minimum range of multi-weed spacing that the solenoid valve can respond to spraying. When the blue dotted line box of multiple weeds intersects, the intersection will become the yellow area shown in the figure, and will automatically generate a red dotted line box on the yellow area according to the minimum distance between multiple weeds that the solenoid valve can respond to spraying. At this time, the system will be generated according to the red solid line box and the red dashed line box for determination, when the red solid line box touches the grid, the corresponding grid will change from 0 to 1, the middle of the red dashed line box will make the grid to keep the determination of the results of 1, until all the red solid line box away from the grid determination will change from 1 to 0. That is to say, there are neighboring multi-weed grid, the algorithm will be for the phenomenon so that the solenoid valve to remain on continuously.

2.4.2 Algorithm for target spraying hysteresis

There is a distance L_1 between the position of the judgment grid for the target decision and the spray nozzle, which requires the sprayer to travel this distance before controlling the solenoid valve for spraying. And in the process of sending the spraying command, there is a delay in the hardware system, and there is a delay time for the liquid to float down to the weeds. The schematic diagram is shown in Figure 6.

Referring to Figures 5 and 6, taking nozzle No. 1 as an example, v is the traveling speed of the sprayer. The delay time t_c includes the equipment processing time t_1, the solenoid valve response time t_2, and the liquid droplet settling time t_3. When the grid determination value changes, the position A_a of the spray nozzle opening and closing includes: the distance between the determination grid position and the spray nozzle L_1. The spray machine traveling distance L_2 at the delay time t_c. In order to ensure the target hit rate and spraying rate, as well as to reduce the error caused by the speed measurement hardware, the distance L_3 is compensated at the above distance.

$$t_c = t_1 + t_2 + t_3$$

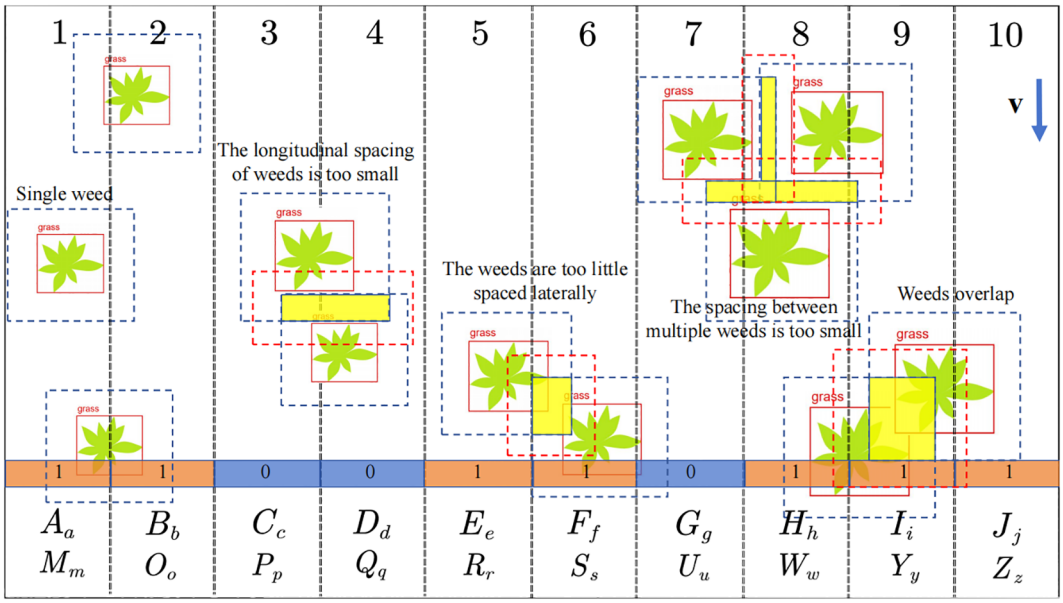


FIGURE 5
Schematic diagram of decision making for target spraying.

$$L_2 = vt_c$$

$$A_a = L_1 + L_2 + L_3$$

The GNSS antenna and the motor speed measuring device feedback the speed at a frequency of 10 Hz, these speed data are discrete, so it is not accurate to calculate them by multiplying the speed by the time and then adding them up, here they are calculated by numerical integration. Numerical integration methods include the Simpson method and the trapezoidal method. In general, the Simpson method is more accurate than the trapezoidal method because it uses a more sophisticated curve approximation method that takes into account the quadratic polynomial relationship between the velocity data points. As a result, the Simpson method typically provides a more accurate estimate of the integral. The trapezoidal method is actually a special case of the Simpson method, which assumes that the relationship between the velocity data points is linear. In order to make the value of distance Mm more accurate, the composite product formula method is used, which consists of the composite Simpson formula and the composite trapezoidal formula. In the case where the speed is essentially

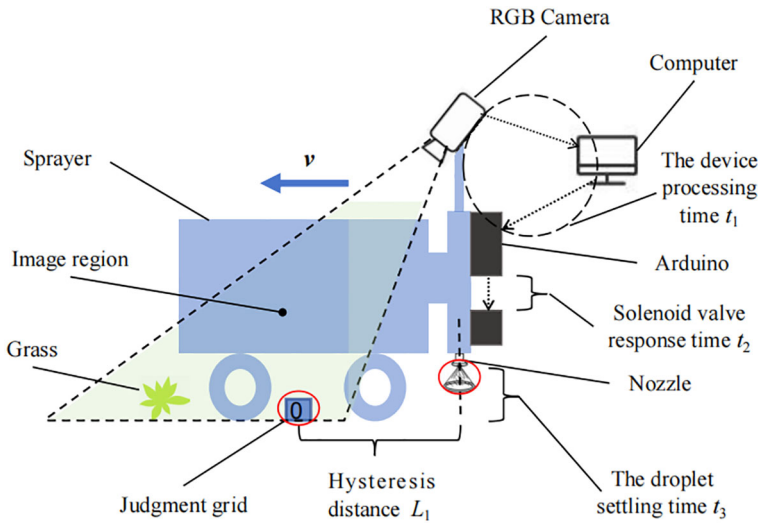


FIGURE 6
Schematic diagram of weed hysteresis to target spraying.

uniform and is measured with a feedback of 0.1 s, the difference between the two methods of integration may not be very significant, and the spray vehicle is essentially moving at a uniform speed during spraying, with a relatively small change in speed, which makes the error between the two methods relatively small. The error of the two methods is relatively small. However, the composite Simpson method requires that the data sub-interval must be even, in the process of the actual working condition of the sprayer, the entire sprayer travel time according to the 0.1 seconds divided by the parity of the interval is uncertain, so the composite trapezoidal method is used for calculation.

$$h = \frac{t}{n} = 0.1$$

$$t_k = kh = 0.1k(k = 0, 1, 2, \ldots, n)$$

$$M_m = \int_0^t v(t)dt \approx \frac{h}{2}\left[v(0) + v(t_k) + 2\sum_{k=1}^{n-1} v(t_k)\right]$$

$$= 0.05\left[v(0) + v(t_k) + 2\sum_{k=1}^{n-1} v(t_k)\right]$$

Here h is the step size. n is the number of times the velocity is recorded, with an initial value of 0. The distance traveled by the sprayer after a change in the grid determination value (when 0 becomes 1 or 1 becomes 0) is M_m. The time is t. The initial values are all 0. t_k is the kth time value, which is a multiple of 0.1. $v(0)$ is the value of the velocity of the sprayer at the time of a change in the grid determination value. $v(t_k)$ is the kth GNSS antenna feedback velocity value.

Here after initialization of the system a=m=1, the controller reads the forward speed of the sprayer in real time, and when the value of the grid determination becomes 1, the controller calculates the positions A_a and M_m where the spraying starts and updates A_1 and M_1 in the array [X]. When $M_1 \geq A_1$, the corresponding nozzle starts spraying, shifts the subscripts of A_a and M_m back by one bit, a=a+1, m=m+1, and resets t_c, t_k, k, A_2 and M_2 to 0. When the value of the grid determination becomes 0, the controller calculates the position A_a and M_m of the end of the spraying, updating them in the array [X], and when $M_2 \geq A_2$, the nozzles end spraying, a=a+1, z=z+1, and reset t_c, t_k, k, A_3 and M_3 to 0.

In array [X], other groups of printhead parameters are updated in the same way as the first group. Simply put, the hysteresis model controls the corresponding printheads to turn on or off by determining whether the downstream parameters in the array [X] are greater than or equal to the upstream parameters of the corresponding position. Combined with the target decision-making algorithm, the entire algorithm workflow is shown in Figure 7.

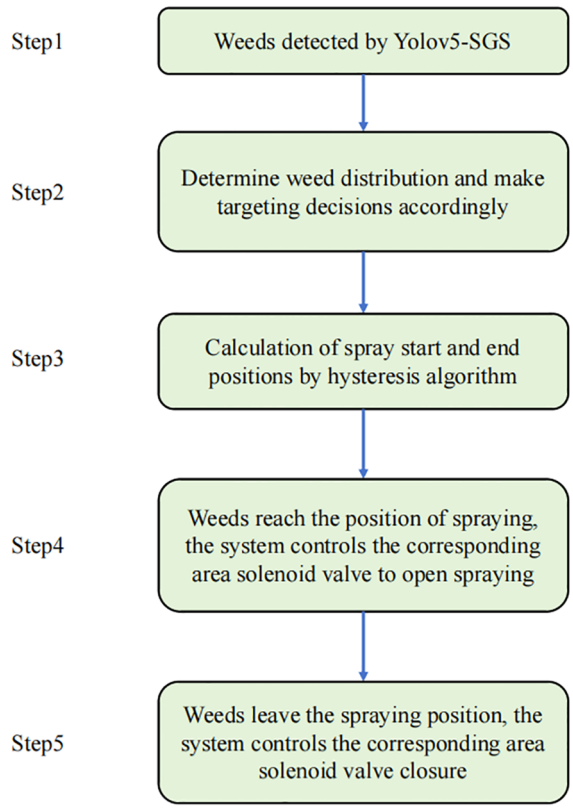


FIGURE 7
Algorithm workflow diagram.

3 Experiments

3.1 Experimental environment for model manipulation

As shown in Table 2, the computer selected for this study is Y9000X manufactured by Lenovo (Beijing, China), based on Intel(R) Core(TM) i7-12700H CPU, 16GB, NVIDIA GeForce RTX 3050Ti Laptop GPU hardware environment for weed model training and testing experiments, running on Python 3.8. The deep learning model is Pytorch 1.11.0+ Torchvision 0.12.0, and the running system is Windows 11.0. During the training process, the initial learning rate is set to 0.001, the ratio between the final learning rate and the initial learning rate is set to 0.2, the momentum of the SGD optimizer is set to 0.937, the number of rounds of the learning rate preheating is set to 3, the momentum at preheating is set to 0.8, and the biased learning at preheating is set to 0.1. rate was set to 0.1, and a cosine annealing learning rate tuning strategy was used. The model was trained for a total of 300 epochs.

3.2 Model detection effectiveness analysis

The improved weed detection results of Yolov5-SGS and Yolov5s are shown in Figure 8, in which the left side is the detection result of Yolov5-SGS and the right side is the detection result of Yolov5s. Through the comparison in the figure, Yolov5-SGS is more accurate in weed target detection, which can effectively solve the problems of small target unrecognizable, low precision of multi-target coexistence detection and inaccurate multi-target cross detection. Due to the wide shooting range, some weeds are smaller in the image and Yolov5s didn't recognize these small weeds, while the improved Yolov5-SGS can detect these small weeds and the

TABLE 2 Training platform configuration details.

Enterprise	Configuration Details
CPU	12th Gen Intel(R) Core(TM) i7-12700H@2.3GHZ
GPU	NVIDIA GeForce RTX 3050 Ti Laptop GPU
Development environment	Python 3.8
Deep learning Module	Pytorch1.11.0+Torchvision 0.12.0
Operating system	Windows 11.0

detection effect of multi-weed coexistence rises. Moreover, Yolov5-SGS has a small detection leakage and improves the accuracy of detection when cross-detecting multiple weeds. This is due to the addition of the SimAM attention mechanism expanding the distinction between weeds and background.

3.3 Ablation experiments

To evaluate the performance of each module, ablation and comparison experiments were conducted. Training and testing were performed on the same dataset and training parameters and the results are shown in Table 3.

From the experimental results, it can be seen that: firstly, after optimizing the backbone feature extraction network using Ghost Module, the GFLOPs (computational complexity) and the model size decreased significantly, by 32.7% and 26.6%, respectively. Secondly, after using Slim -Neck to change the neck fusion network, at this time, F1 and mAP are improved by 1.5% and 0.3% compared with Yolov5s +Ghost Module, while the GFLOPs are decreased by 3.1, which indicates that the improved neck network improves the speed of the model in feature fusion and processing. Finally, introducing SimAM to adjust the attention distribution, F1 and mAP were improved by 0.4% and 0.9% compared to Yolov5s+Ghost Module+Slim-Neck, while

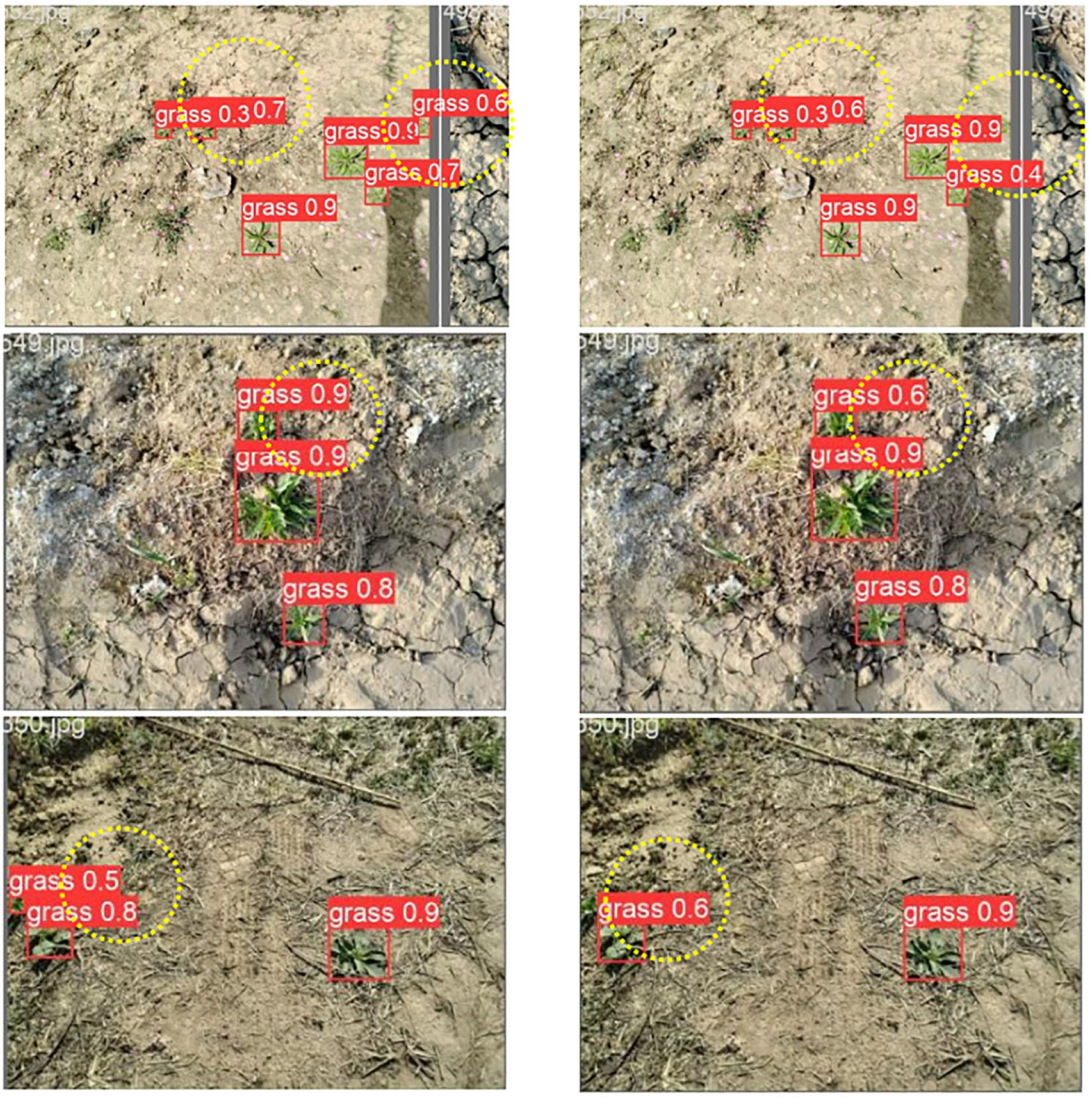


FIGURE 8
Comparison of detection results: the left side is the identification result of Yolov5-SGS, and the right side is the identification result of Yolov5s.

TABLE 3 Results of ablation experiments with different optimization modules.

Modules	Yolov5s	Ghost Module	Slim-Neck	SimAM	GFLOPs	Model size (kB)	mAP (%)	F1 (%)
1	✓				15.9	14059	91.2	84.5
2	✓	✓			10.7	10324	90.2	83.4
3	✓	✓	✓		7.6	8083	90.5	84.9
4	✓	✓	✓	✓	7.6	8084	91.4	85.3

GFLOPs remained unchanged, indicating that the self-attention mechanism improves the detection performance of small targets, widens the distinction between weeds and background, and does not complicate the model. The improved Yolov5-SGS model reduces the computational complexity and model size by half compared to the original model, and the accuracy of recognizing weeds rises.

3.4 Comparison of different target recognition models

The improved Yolov5-SGS network model is subjected to comparative experiments with many other Yolo series lightweight detection models to validate the performance of the proposed model. The experimental results are shown in Table 4.

The comparison shows that the improved model Yolov5-SGS has the smallest GFLOPs (computational complexity) and model size and has the highest mAP and F1 inside all the models compared to Yolov5s, Yolov7tiny, and Yolov8s. Although the GFLOPs and model size of Yolov5-SGS are not as small as Yolov4 tiny, the mAP and F1 of Yolov5-SGS are much higher than that of Yolov4 tiny, which indicates that YOLOv5-SGS has the advantages of low computational complexity and the ability to maintain a very high level of accuracy with a very small model, which are not found in other lightweight YOLO models. accuracy when the model is very small.

4 Experimental results and analysis

4.1 Experimental site

To verify the accuracy of the target spraying decision and hysteresis algorithms and to evaluate the performance of the target spraying control system, a single target spraying control system test bed was built for testing. The test bed is shown in Figure 9. Here, the motor speed of the conveyor belt is sent to the computer in real time and converted into velocity values to replace the GNSS antenna for velocity measurements.

Simulated weeds and simulated wheat similar to those in the field were used in the experiment instead of real plants, and 600 images were taken as a dataset. The modified Yolov5-SGS network model was used in the same operating environment as before for 300 epochs of training. The trained model performed as follows in terms of accuracy metrics: 96.6% for P, 97.7% for R, and 99.1% for mAP.

TABLE 4 Comparison of models.

Models	GFLOPs	Model size (kB)	mAP (%)	F1 (%)
Yolov5-SGS	7.6	8084	91.4	85.3
Yolov4 tiny	5.7	6063	70.2	57.2
Yolov5s	15.9	14059	91.2	84.5
Yolov7 tiny	13.6	11984	91.3	85.2
Yolov8s	28.6	21977	90.0	83.7

4.2 System-to-target spraying effects

The camera and spray nozzle were fixed to the frame. At the beginning of the test, the simulated plants on the conveyor belt moved towards the back of the camera at randomly varying speeds v within the set speed intervals to simulate the actual situation of the sprayer traveling in the field. The speed intervals were set at 0.3-0.4 m/s, 0.4-0.5 m/s and 0.5-0.6 m/s. In each test, a water-sensitive paper was placed next to the simulated weed in order to provide an objective assessment of the spraying effect. This water-sensitive paper would quickly turn red upon contact with water, so that the actual effect of spraying could be clearly observed. In order to ensure the reliability and accuracy of the results, up to 50 trials were conducted by constantly changing the position and type of the simulated weeds on the conveyor belt. The number of weeds used in each test was 20, so that the total number of weeds tested in the 50 tests amounted to 1,000.

Recognition rate, spraying rate and hit rate were calculated based on the reddening of the water sensitive paper color, respectively. Recognition rate is the ratio of the number of weeds detected by the system to the number of all weeds. Spraying rate is the ratio of the number of weeds whose water-sensitive paper turns red to the number of detected weeds. The hit rate is the product of the recognition rate and the spraying rate. Specific test results are shown in Table 5.

The extent of weed coverage by the system spray is recorded by the length of the water-sensitive paper that turns red. Coverage was the ratio of the length of the weed to the length of the water-sensitive paper discoloration. The targeting error is the error between the center position of the weed and the center position of the water-sensitive paper discoloration, and the mean absolute

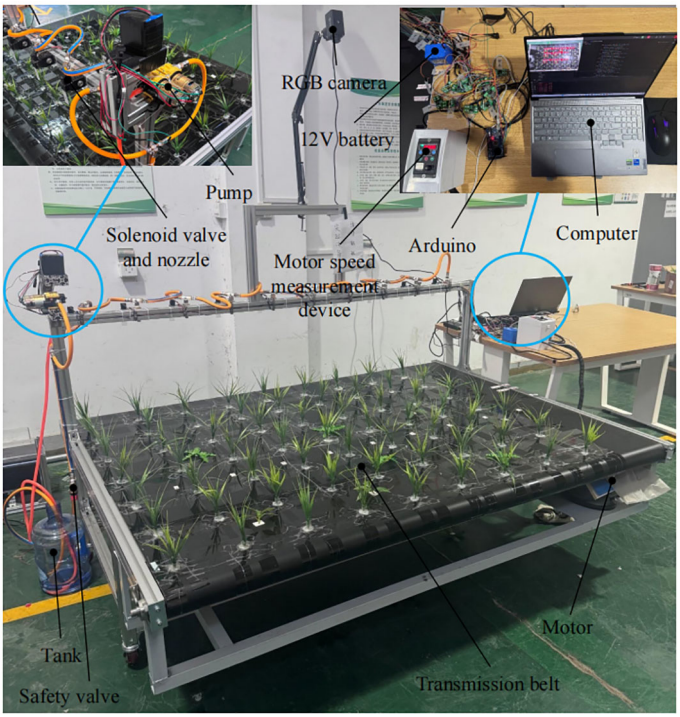


FIGURE 9
Field test of target spraying system.

error and root mean square error were calculated. The results are shown in Table 6.

According to the experimental results, in the velocity interval of 0.3~0.4m/s and 0.4~0.5m/s, the spray system has a very good performance in the rate of target spraying and the coverage of spraying on weeds, especially in the realization of target spraying on multiple weeds in close proximity to each other, both of them can ensure that the mist adheres to the mist, and the spraying rate of the whole experiment reaches 99.8% and 98.2%, respectively, which indicates that the target spraying decision and the hysteresis algorithm is effective. However, there was a substantial decrease in the precision of the target spraying at 0.5 to 0.6 m/s, which was due to the inability of the velocimetry device to feed back the speed at a higher frequency, leading to errors in the calculation of the weed position by the target spraying decision and hysteresis algorithms. Therefore, at faster speeds, a velocimetry hardware device with a higher feedback frequency is needed to ensure the accuracy of the algorithm.

TABLE 5 Spraying results of different speed intervals for the target spraying system.

Speed range (m/s)	Recognition rate (%)	Spraying rate (%)	Hit rate (%)
0.3~0.4	98.5	99.8	98.3
0.4~0.5	98.1	98.2	96.3
0.5~0.6	97.4	95.7	93.2

4.3 Analyses and discussions

Current weed detection and smart spraying systems still face many bottlenecks, and the processing power of computers and the costs associated with system hardware limit the development of target spraying systems (Vijayakumar et al., 2023). Fan et al. (2023) developed a weed detection and target spraying robotic system for seedling stage in a cotton field based on deep learning, which utilized three detection cameras to to acquire cotton and weed images of row plants in real time, each camera controls one spray nozzle each, and the three nozzles are responsible for spraying weeds in the corresponding area.Liu et al. (2021), on the other hand, deployed the DCNN model on a computer and used three webcams to perform weed identification with solenoid valves below each camera with a response time of less than 60 ms.Balabantaray et al. (2024) developed a weeding robot based on YOLOv7. The computer mainframe used by the robot for weed recognition was a Dell Precision 7,670 mobile workstation (12th Gen Intel Core i9-12950HX, 32GB RAM, NVidia RTX A4500 16 GB)), and the spraying system was equipped with a a single spray nozzle to target spray a single row area. These studies set a precedent for deep learning-based weed target spraying and also provide a solid foundation for target weed control in terms of system design. The focus of this research is to achieve higher spraying accuracy on less costly hardware and to allow for a wider coverage of the target spraying work area. Therefore, we lightened YOLOv5s and ensured that the accuracy of the model rose, where the GFOLPs (computational complexity) and model size were reduced by

TABLE 6 Results of spray coverage.

Speed range (m/s)	Coverage rate (%)	Mean absolute error (cm)	Root mean square error (cm)
0.3~0.4	94.3	1.5	2.6
0.4~0.5	93.7	2.1	3.6
0.5~0.6	87.8	3.8	6.6

52.2% and 42.4%, respectively, while the mAP and F1 values were increased by 0.2% and 0.8%, respectively, so that the weed identification model could be deployed on computers with limited processing power. In addition, the choice of a wide-angle distortion-free RGB camera to cover more area reduces the need for multiple cameras, and a rational structural arrangement allows identification of field areas up to 2m in width, thus further reducing the associated costs. In terms of target localization, Li He et al. (2022) combined deep learning algorithms with spraying technology to design a machine vision accurate real-time targeted spraying system for field scenarios, and proposed a solenoid group switching grid decision algorithm for single weeds.Zhao et al. (2022) developed a machine learning-based targeted spraying system equipped with an algorithm can differentiate between cabbages and weeds and accurately spray herbicides on single cabbages to resist the threat of pests and diseases.The algorithm designed in this study is very effective in localizing individual plants with large spacing and canopy and proposes a solution for the hysteresis problem in the operation process. Upadhyay et al. (2024) designed a spraying system based on machine vision by placing cameras in the three sets of grids in the vertical direction of the recognized area, which effectively improved the problems of single weed plant localization and missed detection. The above study provides an excellent reference solution for plant localization, which has a significant impact on the key operational aspects of target spraying. In this study, considering the low response frequency of most solenoid valves on the market, an algorithm for target spraying decision and hysteresis control is designed based on deep learning, which is capable of distinguishing the different distributions of single-plant and multi-plant weeds. After completing weed identification, the system accurately sprays weeds with different distributions by controlling the solenoid valves, thus avoiding the problem that the solenoid valves cannot respond in time when multiple weeds are too close together. This approach eliminates the need for farmers to purchase expensive high-frequency response solenoid valves, thus controlling the cost while performing precise weed control by carrying this weed localization and recognition algorithm. To address the problem of hysteresis characteristics on the hardware side, this study innovatively proposes to utilize the numerical integration method to calculate the speed in order to cope with the challenge that the data fed back from the speed measuring device is discrete, which effectively solves the hysteresis problem between the spraying hardware.

It is important to note that the targeted spraying system developed in this study is still in the bench stage. The motor-controlled weed movement speed is easier to control on the experimental bench compared to the complex field environment. At the same time, the experimental bench can more easily evaluate the recognition effect of the algorithm and the spray response performance compared to the influence of multiple factors in the field. Therefore, the spraying data we collected on the experimental bench will provide important support for the integration of targeted spraying systems in sprayers. Although there are some limitations in the experimental bench test, its modular design allows multiple targeted spraying systems to be carried in the spraying machine, and based on the shared speed measurement module and herbicide supply module, the image acquisition module and spray execution module are added to meet the actual spraying requirements. In real scenario experiments, problems such as inaccurate weed identification, fog liquid offset and sprayer speed measurement need to be considered. In the future, we plan to integrate dual GNSS antennas for RTK differential speed measurement in the sprayer's target spraying system, and equip it with a high-frame-rate anti-shake RGB camera and a high-performance on-board computer. These high-performance hardware will significantly reduce the weed recognition time and system latency, and reduce the image jitter due to uneven road surface, thus improving the weed recognition rate.

5 Conclusion

In this paper, a deep learning-based weed detection and target spraying control method for wheat field at tillering stage is designed and deployed on the target spraying control system testbed. The study draws the following conclusions: firstly, a target spraying control system is proposed, including an image acquisition and real-time speed measurement module, a signal conversion module, a spray execution module, and a herbicide supply module. Second, in order to reduce the computational complexity and model size while guaranteeing the recognition accuracy, the backbone and neck networks of Yolov5s are lightweighted and improved, and a self-attention mechanism is introduced. These improvements resulted in a reduction of 52.2% and 42.4% in the model's GFOLPs (computational complexity) and model size, respectively, and an increase of 0.2% in the mAP value and and 0.8% in the F1 value, which provided the best performance in comparison with other YOLO series lightweight models. Finally, to address the problem of inaccurate spraying due to the close proximity of different weeds and the hysteresis characteristics of the target spraying hardware, we designed an target spraying decision and hysteresis algorithm and deployed the integrated algorithm on a test bed for experimental validation. The experimental results show that the spraying rate reaches 99.8% and 98.2% in the velocity intervals of 0.3~0.4 m/s and 0.4~0.5 m/s, and the coverage rate is 94.3% and 93.7%, respectively. This indicates that the target spraying system differentiated the spraying of weeds with different distributions and the problem of hysteresis between the hardware was effectively solved. However, at faster speed intervals of 0.4 to 0.5 m/s, the spraying and coverage rates decreased to 95.7% and 87.8%, respectively, and thus a velocimetry device with higher feedback

frequency is required to ensure the accuracy of the algorithm at faster speeds.

Data availability statement

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

Author contributions

HW: Conceptualization, Data curation, Investigation, Software, Validation, Writing – original draft. YC: Formal Analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing. SZ: Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing. PG: Conceptualization, Investigation, Software, Visualization, Writing – original draft. YXC: Investigation, Methodology, Validation, Writing – original draft. GH: Conceptualization, Formal Analysis, Project administration, Supervision, Validation, Writing – review & editing. YM: Data curation, Investigation, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study is supported by the "14th Five-Year Plan" National Key Research and Development Program (2022YFD1900304), the Shaanxi Key Research Development Project (2023-ZDLNY-62), the Shaanxi Provincial Key Research and Development Project (2024NC-YBXM-202, 2024NC-YBXM-244).

Acknowledgments

The authors sincerely acknowledge the members of the research team for their help.

Conflict of interest

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

Generative AI statement

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

Publisher's note

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

References

Alrowais, F., Asiri, M. M., Alabdan, R., Marzouk, R., Hilal, A. M., Alkhayyat, A., et al. (2022). Hybrid leader based optimization with deep learning driven weed detection on internet of things enabled smart agriculture environment. *Comput. Electrical Eng.* 104, 108411. doi: 10.1016/j.compeleceng.2022.108411

Balabantaray, A., Behera, S., Liew, C., Chamara, N., Singh, M., Jhala, A. J., et al. (2024). Targeted weed management of Palmer amaranth using robotics and deep learning (YOLOv7). *Front. Robotics AI* 11. doi: 10.3389/frobt.2024.1441371

Brown, D. L., Giles, D. K., Oliver, M. N., and Klassen, P. (2008). Targeted spray technology to reduce pesticide in runoff from dormant orchards. *Crop Prot.* 27, 545–552. doi: 10.1016/j.cropro.2007.08.012

Cassman, K. G. (1999). Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agriculture. *Proc. Natl. Acad. Sci.* 96, 5952–5959. doi: 10.1073/pnas.96.11.5952

Deng, B., Lu, Y., and Xu, J. (2024). Weed database development: An updated survey of public weed datasets and cross-season weed detection adaptation. *Ecol. Inf.* 81, 102546. doi: 10.1016/j.ecoinf.2024.102546

Fan, X., Chai, X., Zhou, J., and Sun, T. (2023). Deep learning based weed detection and target spraying robot system at seedling stage of cotton field. *Comput. Electron. Agric.* 214, 108317. doi: 10.1016/j.compag.2023.108317

Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., and Xu, C. (2020). "Ghostnet: More features from cheap operations," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*., 1580–1589. doi: 10.1109/CVPR42600.2020.00165

Horvath, D. P., Clay, S. A., Swanton, C. J., Anderson, J. V., and Chao, W. S. (2023). Weed-induced crop yield loss: a new paradigm and new challenges. *Trends Plant Science* 28 (5), 567–582. doi: 10.1016/j.tplants.2022.12.014

Jabran, K., Mahmood, K., Melander, B., Bajwa, A. A., and Kudsk, P. (2017). Weed dynamics and management in wheat. *Adv. Agron.* 145, 97–166. doi: 10.1016/bs.agron.2017.05.002

Khan, M. M., Rahman, M. M., Hasan, M. M., Amin, M. F., Matin, M. Q. I., Faruq, G., et al. (2024). Assessment of the salt tolerance of diverse bread wheat (Triticum aestivum L.) genotypes during the early growth stage under hydroponic culture conditions. *Heliyon* 10, e29042. doi: 10.1016/j.heliyon.2024.e29042

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Li, H., Guo, C., Yang, Z., Chai, J., Shi, Y., Liu, J., et al. (2022a). Design of field real-time target spraying system based on improved YOLOv5. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1072631

Li, H., Li, J., Wei, H., Liu, Z., Zhan, Z., and Ren, Q. (2022b). Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. *arXiv*. doi: 10.48550/arXiv.2206.02424

Liu, J., Abbas, I., and Noor, R. S. (2021). Development of deep learning-based variable rate agrochemical spraying system for targeted weeds control in strawberry crop. *Agronomy* 11, 1480. doi: 10.3390/agronomy11081480

Lv, X., Zhang, X., Gao, H., He, T., Lv, Z., and Zhangzhong, L. (2024). When crops meet machine vision: A review and development framework for a low-cost nondestructive online monitoring technology in agricultural production. *Agric. Commun.* 2, 100029. doi: 10.1016/j.agrcom.2024.100029

Modi, R. U., Kancheti, M., Subeesh, A., Raj, C., Singh, A. K., Chandel, N. S., et al. (2023). An automated weed identification framework for sugarcane crop: A deep learning approach. *Crop Prot.* 173, 106360. doi: 10.1016/j.cropro.2023.106360

Rahman, A., Lu, Y., and Wang, H. (2023). Performance evaluation of deep learning object detectors for weed detection for cotton. *Smart Agric. Technol.* 3, 100126. doi: 10.1016/j.atech.2022.100126

Rai, N., Zhang, Y., Villamil, M., Howatt, K., Ostlie, M., and Sun, X. (2024). Agricultural weed identification in images and videos by integrating optimized deep learning architecture on an edge computing technology. *Comput. Electron. Agric.* 216, 108442. doi: 10.1016/j.compag.2023.108442

Saleem, M. H., Velayudhan, K. K., Potgieter, J., and Arif, K. M. (2022). Weed identification by single-Stage and two-Stage neural networks: A study on the impact of image resizers and weights optimization algorithms. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.850666

Shao, Y., Guan, X., Xuan, G., Gao, F., Feng, W., Gao, G., et al. (2023). GTCBS-YOLOv5s: A lightweight model for weed species identification in paddy fields. *Comput. Electron. Agric.* 215, 108461. doi: 10.1016/j.compag.2023.108461

Sharma, A., Kumar, V., Shahzad, B., Tanveer, M., Sidhu, G. P. S., Handa, N., et al. (2019). Worldwide pesticide usage and its impacts on ecosystem. *SN Appl. Sci.* 1. doi: 10.1007/s42452-019-1485-1

Shiferaw, B., Smale, M., Braun, H. J., Duveiller, E., and Muricho, G. (2013). Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. *Food Secur.* 5. doi: 10.1007/s12571-013-0263-y

Sunil, G. C., Upadhyay, A., and Sun, X. (2024). Development of software interface for AI-driven weed control in robotic vehicles, with time-based evaluation in indoor and field settings. *Smart Agric. Technol.* 9, 100678. doi: 10.1016/j.atech.2024.100678

Upadhyay, A., Sunil, G. C., Zhang, Y., Koparan, C., and Sun, X. (2024). Development and evaluation of a machine vision and deep learning-based smart sprayer system for site-specific weed management in row crops: An edge computing approach. *J. Agric. Food Res.* 18, 101331. doi: 10.1016/j.jafr.2024.101331

Vijayakumar, V., Ampatzidis, Y., Schueller, J. K., and Burks, T. (2023). Smart spraying technologies for precision weed management: A review. *Smart Agric. Technol.* 6, 100337. doi: 10.1016/j.atech.2023.100337

Yang, L., Zhang, R.-Y., Li, L., and Xie, X. (2021). "SimAM: A simple, parameter-free attention module for convolutional neural networks," in *Proceedings of the 38th International Conference on Machine Learning*, Vol. 139, Corpus ID: 235825945., 125–135.

Yu, J., Schumann, A. W., Cao, Z., Sharpe, S. M., and Boyd, N. S. (2019). Weed detection in perennial ryegrass with deep learning convolutional neural network. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01422

Yuan, C., Liu, T., Gao, F., Zhang, R., and Seng, X. (2023). YOLOv5s-CBAM-DMLHead: A lightweight identification algorithm for weedy rice (Oryza sativa f. spontanea) based on improved YOLOv5. *Crop Prot.* 172, 106342. doi: 10.1016/j.cropro.2023.106342

Zhao, X., Wang, X., Li, C., Fu, H., Yang, S., and Zhai, C. (2022). Cabbage and weed identification based on machine learning and target spraying system design. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.924973

Zhu, H., Zhang, Y., Mu, D., Bai, L., Wu, X., Zhuang, H., et al. (2024). Research on improved YOLOx weed detection based on lightweight attention module. *Crop Prot.* 177, 106563. doi: 10.1016/j.cropro.2023.106563

Zhu, A., Zhang, R., Zhang, L., Yi, T., Wang, L., Zhang, D., et al. (2024). YOLOv5s-CEDB: A robust and efficiency Camellia oleifera fruit detection algorithm in complex natural scenes. *Comput. Electron. Agric.* 221, 108984. doi: 10.1016/j.compag.2024.108984

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 27 March 2025
DOI 10.3389/fpls.2025.1491170

Check for updates

OPEN ACCESS

EDITED BY
Jun Ni,
Nanjing Agricultural University, China

REVIEWED BY
Bolai Xin,
Wageningen University and Research, Netherlands
Xiangyang Yuan,
Hebei Agricultural University, China
Puyam Singh,
North Eastern Space Applications Centre (NESAC), India

*CORRESPONDENCE
Zhibin Zhang
cszhibin@imu.edu.cn

RECEIVED 04 September 2024
ACCEPTED 06 March 2025
PUBLISHED 27 March 2025

CITATION
Qiao G, Zhang Z, Niu B, Han S and Yang E (2025) Plant stem and leaf segmentation and phenotypic parameter extraction using neural radiance fields and lightweight point cloud segmentation networks. *Front. Plant Sci.* 16:1491170. doi: 10.3389/fpls.2025.1491170

COPYRIGHT
© 2025 Qiao, Zhang, Niu, Han and Yang. 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.

Plant stem and leaf segmentation and phenotypic parameter extraction using neural radiance fields and lightweight point cloud segmentation networks

Gaofei Qiao, Zhibin Zhang*, Bin Niu, Sijia Han and Enhui Yang

Key Laboratory of Wireless Networks and Mobile Computing, School of Computer Science, Inner Mongolia University, Hohhot, China

High-quality 3D reconstruction and accurate 3D organ segmentation of plants are crucial prerequisites for automatically extracting phenotypic traits. In this study, we first extract a dense point cloud from implicit representations, which derives from reconstructing the maize plants in 3D by using the Nerfacto neural radiance field model. Second, we propose a lightweight point cloud segmentation network (PointSegNet) specifically for stem and leaf segmentation. This network includes a Global-Local Set Abstraction (GLSA) module to integrate local and global features and an Edge-Aware Feature Propagation (EAFP) module to enhance edge-awareness. Experimental results show that our PointSegNet achieves impressive performance compared to five other state-of-the-art deep learning networks, reaching 93.73%, 97.25%, 96.21%, and 96.73% in terms of mean Intersection over Union (mIoU), precision, recall, and F1-score, respectively. Even when dealing with tomato and soybean plants, with complex structures, our PointSegNet also achieves the best metrics. Meanwhile, based on the principal component analysis (PCA), we further optimize the method to obtain the parameters such as leaf length and leaf width by using PCA principal vectors. Finally, the maize stem thickness, stem height, leaf length, and leaf width obtained from our measurements are compared with the manual test results, yielding R^2 values of 0.99, 0.84, 0.94, and 0.87, respectively. These results indicate that our method has high accuracy and reliability for phenotypic parameter extraction. This study throughout the entire process from 3D reconstruction of maize plants to point cloud segmentation and phenotypic parameter extraction, provides a reliable and objective method for acquiring plant phenotypic parameters and will boost plant phenotypic development in smart agriculture.

KEYWORDS

three-dimensional point cloud, plant phenotype, neural radiance fields, point cloud segmentation, lightweight network

1 Introduction

Plant phenotyping involves the systematic measurement and evaluation of various observable characteristics exhibited of plants during their growth and development. Corn, as a globally important grain crop, serves as a major food source for humans and livestock and plays a crucial role in industrial and biofuel production. In the process of breeding high-yield, disease-resistant, and high-quality corn varieties, the measurement and analysis of plant traits are essential (Wu et al., 2024). Traditional methods for extracting stem and leaf phenotypic traits primarily rely on manual operations, resulting in high labor costs, low efficiency, and potential sample damage (Tardieu et al., 2017). Additionally, manual operations have limited applicability, and cannot meet the demands of large-scale and diverse researches due to being prone to errors. In recent years, computer vision technology has made significant advances in plant phenotyping. These technologies provide automated and precise methods for data acquisition and analysis, significantly improving the efficiency and quality of plant phenotypic data collection (Rawat et al., 2022).

In previous research, 2D image-based segmentation and detection techniques have been extensively developed. Due to the relatively simple acquisition and processing of 2D images, they are suitable for large-scale data collection and high-throughput analysis and have thus been widely applied in many plant phenotyping studies (Li et al., 2020). For instance, (Zhang et al., 2024) proposed the DSBEAN framework, which combines soybean breeding technology with deep learning algorithms to predict pod counts through primary node detection and pod area identification. (Shi et al., 2022) developed a high-throughput pipeline for maize ear phenotyping, capable of extracting the number of kernels, the number of rows, and the number of kernels per row from images. (Xu et al., 2023) proposed a deep learning-based semantic segmentation model, GlandSegNet, for extracting cotton pigment gland areas. (Du et al., 2020) introduced a convolutional neural network (CNN)-based approach for object detection, semantic segmentation, and phenotypic analysis to extract the geometric and color traits of lettuce varieties. However, 2D image technology is limited by perspective constraints and the lack of depth information, leading to measurement errors, occlusion, and overlapping issues, making it difficult to accurately reflect the three-dimensional structure of plants (Qiao et al., 2023), especially the actual morphology, and characteristics of organs such as corn stems and leaves.

In 3D reconstruction, sensors are crucial for obtaining high-quality three-dimensional data. Various sensors, including monocular cameras, RGB-D cameras, LiDAR, and laser scanners, offer diverse functionalities and advantages to address different needs (Yu et al., 2024). Alongside active reconstruction methods, passive methods use RGB cameras to capture images and extract depth information, such as multiple view stereo algorithm (MVS) (Furukawa and Ponce, 2009) and structure from motion algorithm (SFM) (Moulon et al., 2013). Recent advancements in deep learning have led to the widespread application of neural networks in 3D reconstruction. MVSNet (Yao et al., 2018) enhances reconstruction accuracy by converting traditional MVS problems into deep learning tasks through an end-to-end network that generates depth maps from multi-view images. NeRF (Neural Radiance Fields) (Mildenhall et al., 2021) introduces a novel approach by using a fully connected neural network to learn implicit scene representations from multi-view images, achieving high-fidelity view synthesis and 3D reconstruction. (Thapa et al., 2018) employed a LiDAR-based sensing system to reconstruct the leaf surfaces of corn and sorghum, obtaining features such as leaf area and leaf inclination. (Hu et al., 2018) used the Kinect v2 structured light camera to measure key growth parameters of lettuce. (Wu et al., 2022) applied MVS algorithms to extract wheat point clouds and calculate parameters like leaf count, plant height, and leaf length. (Huang et al., 2024) proposed a neural distance field-based method to extract geometric parameters of walnut shells from multi-view image sequences.

Accurate phenotypic measurement requires high-quality 3D reconstruction and fast, reliable 3D segmentation methods to achieve automated extraction. Traditional point cloud segmentation methods are straightforward to implement, require minimal computational resources, allow real-time processing, and offer strong interpretability. For instance, (Miao et al., 2022) utilized TLS technology to capture point cloud data of bananas and achieved segmentation of individual banana plants using a combination of Euclidean clustering and K-means. (Wang et al., 2020) proposed a dynamic view-based adaptive K-means algorithm for segmenting and measuring the size of wheat spikes. Although these traditional methods perform well in specific scenarios, they typically involve extensive manual operation and complex parameter settings, which limits their scalability for large-scale applications. Recently, deep learning methods have been extensively applied to point cloud processing and analysis. (Yang et al., 2024b) developed a 3D semantic segmentation network for corn ears based on MVS technology to detect infected areas. (Luo et al., 2024) proposed a semantic segmentation model that combines Mask R-CNN with an improved version of PointNet++ to achieve precise segmentation of grape clusters, flower stems, and leaves. (Yan et al., 2024) developed a lightweight 3D deep learning network that accurately extracts the segmentation and phenotypic traits of corn organs.

Through these advanced methods, the potential of 3D reconstruction and point cloud segmentation technologies for agricultural phenotypic analysis has been significantly enhanced. However, practical applications still face challenges including irregular terrain, diverse crop types, and varying lighting conditions. While LiDAR devices offer high precision, they are expensive and affected by weather conditions; on the other hand, Kinect devices are cost-effective but have low point cloud resolution and are influenced by lighting conditions (Yang et al., 2024a). Compared to expensive equipment such as LiDAR, Nerfacto (Tancik et al., 2023) achieves high-quality reconstruction using only images captured by ordinary cameras, significantly reducing hardware costs. In agricultural scenarios, Nerfacto effectively addresses occlusion issues between plant leaves. Furthermore, this method requires only a small number of input images to generate high-fidelity 3D models, further enhancing data utilization

efficiency. To enhance point cloud segmentation accuracy, previous research has developed complex feature extraction modules and deeper network architectures, which inevitably increase model parameters and computational demands. Due to the scarcity of high-quality plant point cloud datasets, data augmentation techniques are necessary to create more diverse training samples. To address these challenges, we first employed the 3D reconstruction method, Nerfacto, which generates high-quality plant models from a limited number of images and adapts to complex environments. Secondly, we proposed a lightweight point cloud segmentation model based on PointNet++ (Qi et al., 2017) that redesigns the encoder and decoder to significantly reduce computational complexity while maintaining accuracy. Finally, we refined the methods for calculating leaf length and width to better accommodate the curvature and complex morphology of leaves.

In summary, the contributions of this paper are as follows:

1. We applied the 3D reconstruction method Nerfacto to plant 3D reconstruction and explored methods for extracting point clouds from implicit neural radiance fields (NeRF). Leveraging Nerfacto technology, we developed a 3D point cloud dataset for maize stalks and leaves, including segmentation and related phenotypic data for 20 plants.
2. We designed a lightweight plant point cloud segmentation model with only 1.33 M parameters. We introduced a novel Global-Local Set Abstraction (GLSA) module to integrate local and global information and an Edge-Aware Feature Propagation (EAFP) module to enhance the handling of deep-edge features. Experimental results show that our proposed model while maintaining the smallest parameter count, achieves competitive accuracy compared to the latest segmentation networks across multiple datasets (maize, tomato, soybean).
3. We proposed an optimized method for better obtaining parameters such as leaf length and leaf width by segmenting the PCA principal vectors. This approach generates curves that accurately reflect leaf curvature, allowing for a more precise measurement of leaf length and width. This method provides robust support for the measurement of plant phenotypic parameters.

2 Materials and methods

2.1 Overview

The overall workflow is illustrated in Figure 1 and is divided into four parts: data acquisition, a 3D reconstruction based on neural radiance fields, a lightweight plant point cloud stem and leaf segmentation model, and plant phenotypic data measurement. Firstly, in part (A), a smartphone is used as the image acquisition device, and it is handheld and slowly moved counterclockwise around the maize plant to ensure the capture of clear and complete coverage videos of the entire plant. Several image frames were extracted from the videos, and COLMAP is used to compute the poses of the camera to determine its position and orientation in space. To accelerate the convergence speed of the neural radiance fields and improve the reconstruction quality, these images and their associated camera poses and parameters are converted into Local Light Field Fusion (LLFF) format (Mildenhall et al., 2019). Secondly, in the part (B), a neural radiance field (NeRF) is trained by using the transformed data, in which NeRF can capture rich 3D structural information by computing depth estimations of the plant surface from each viewpoint. Repeating these steps generates a dense point cloud so that the point cloud of entire maize plant is obtained, and then processed using statistical filtering and farthest point sampling (FPS). Thirdly, in the part (C), the processed point cloud is fed into the proposed lightweight stem-leaf point cloud segmentation model to obtain segmentation results of them. Experiments are also conducted on the tomato and soybean datasets to further validate the proposed segmentation model performance across various plant types. Finally, key four maize phenotypic traits including the stem height, stem diameter, leaf width, and leaf length are extracted from the segmented point clouds in the part (D).

2.2 Experimental materials and data collection

2.2.1 Maize point cloud collection and annotation

We randomly selected 20 maize plants from the experimental field at Inner Mongolia University and transplanted them indoors for data acquisition. During the acquisition process, we ensured adequate lighting and wind-free conditions to minimize the impact of lighting and wind on data quality. A Xiaomi 14 smartphone was handheld and moved counterclockwise around the plants to capture video. The capture time for each maize plant was set to 40-50 seconds, ensuring comprehensive coverage from all angles. Ultimately, 60 to 80 key image frames were extracted from the videos. Next, we used COLMAP to process these image frames and calculate the camera poses, i.e., the position and orientation of the camera in space. To support subsequent 3D reconstruction and analysis, we converted the images and their associated camera poses and parameters into the Local Light Field Fusion (LLFF) format. We then used Nerfacto for 3D reconstruction to obtain dense point cloud data (detailed in section 2.3).

To validate the model's applicability across different datasets and further expand the dataset, we combined the collected point cloud data with a publicly available maize dataset (Yang et al., 2024c), and constructed a new dataset with 448 point cloud samples. To ensure consistency in format, data quality, and annotations between the newly collected maize point cloud data and the public maize dataset, we implemented the following measures:

1. Color Threshold Segmentation: We used color threshold segmentation to distinguish between the ground and maize

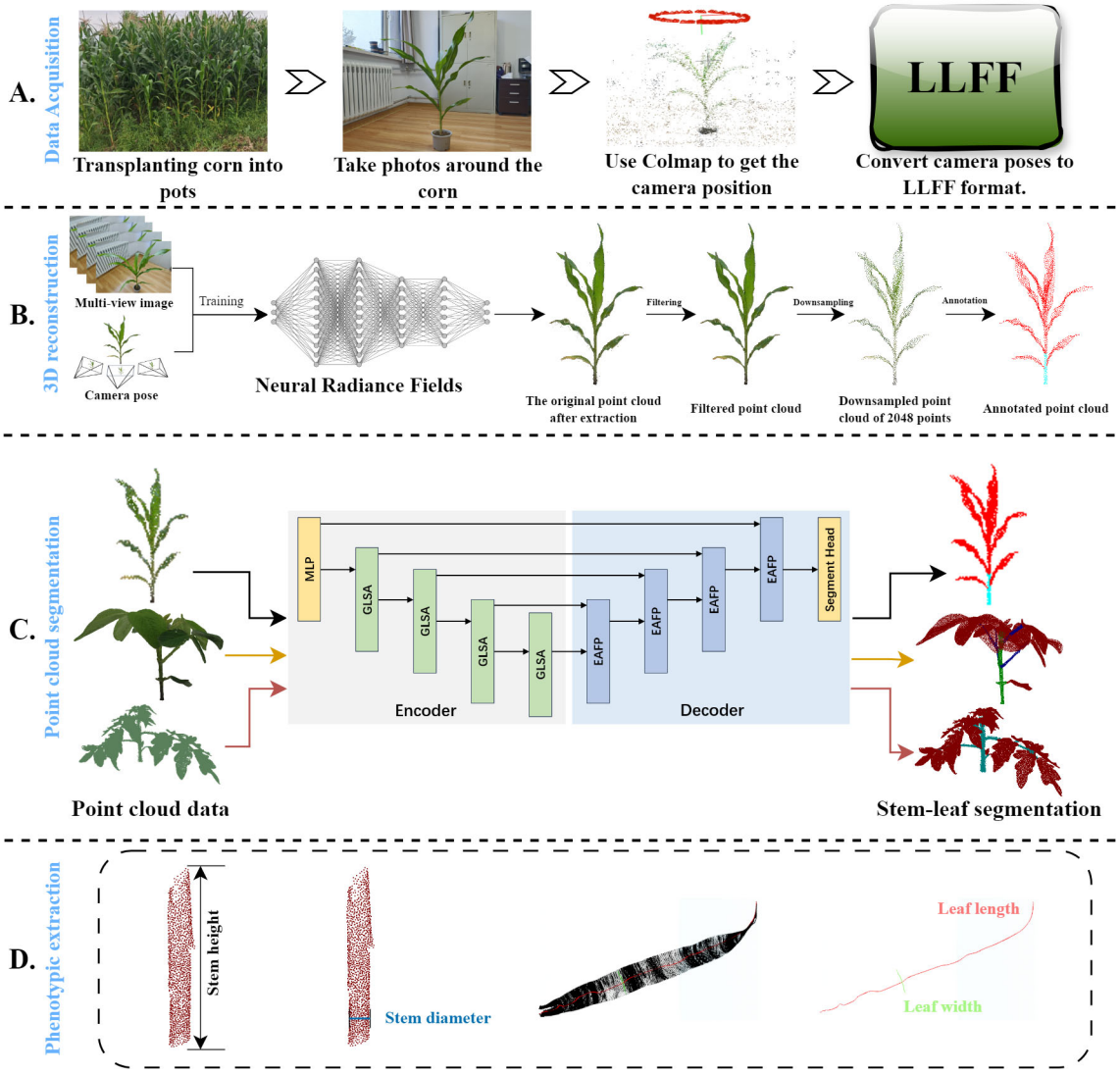


FIGURE 1
Overview of the proposed framework. **(A)** Photographs were taken around a maize plant. The bitmaps of these image frames were computed using COLMAP and then converted to the Local Light Field Fusion (LLFF) format. **(B)** The converted data trains the neural radiance field, which generates dense point clouds. These dense point clouds are then preprocessed. **(C)** The main stems and leaves of the maize are segmented using PointSegNet. The segmentation effectiveness of the model is verified using various complex plants. **(D)** Based on the segmentation results of maize plants, four phenotypic traits (stem height, stem thickness, leaf width, and leaf length) were extracted to validate further the point cloud segmentation effect and shape extraction methods.

plants. First, we calculated the histogram of the color values and then applied Otsu's method to determine the optimal threshold for effectively segmenting the non-ground areas from the ground.

2. Statistical Filtering for Denoising: We used statistical filtering to remove noise by calculating the distance of each point within its neighborhood, then removed the noise points that exceeded this threshold. This method effectively eliminates isolated points and random noise while preserving the main structure.
3. FPS Downsampling: We applied the Farthest Point Sampling (FPS) algorithm to downsample the point cloud. This algorithm iteratively selects the farthest points to ensure that the downsampled point cloud is uniformly distributed while maintaining geometric features. This approach reduces data volume while retaining the overall shape and key details of the point cloud of maize plant.
4. Point Cloud Labeling: We used a point cloud labeling tool to manually label the stem and leaf organs of maize as ground truth for model training.

2.2.2 Tomato and soybean point cloud dataset

Due to the relatively simple structure of maize plants, we further evaluated the performance of our proposed point cloud segmentation model on different plant species using two publicly available plant point cloud datasets with more complex structures: Pheno4D (Schunck et al., 2021) and Soybean-MVS (Sun et al.,

2023). The Pheno4D dataset is a sub-millimeter precision plant point cloud dataset obtained using a laser triangulation scanner. It includes 77 point cloud samples from seven tomato plants collected over 20 days, manually labeled into three categories: soil, stem, and leaves. The Soybean-MVS dataset is a 3D point cloud dataset covering the entire growth period of soybeans, reconstructed using multi-view stereo (MVS) technology. It contains 102 samples of five soybean varieties at 13 growth stages, manually labeled into three categories: leaves, main stem, and branches, including x, y, z coordinates and r, g, b color values. To ensure that the network focuses more on the segmentation of plant organs during training, we preprocessed the point clouds of tomato and soybean plants. Specifically, we removed the "soil" points from the tomato plant dataset, and specific color attributes from the soybean plant dataset. Thus, each point cloud includes x, y, z coordinates and normal vector information N_x, N_y, N_z, calculated through normal vector computation.

2.3 3D reconstruction based on Nerfacto

2.3.1 Implicit neural representations

In traditional 3D representation methods, we typically use point clouds, meshes, or voxels to explicitly define the geometric information of an object, including its vertices, edges, faces, and topological relationships. In contrast, implicit representation methods describe the geometric shape and appearance of a 3D model through a function that uses coordinates as input. This kind of approach does not require the explicit definition of geometric structures but instead predicts surface attributes such as color and density through an implicit function. Neural Radiance Fields (NeRF) (Mildenhall et al., 2021) is an application of implicit neural representation specifically designed for high-quality 3D scene reconstruction and novel view synthesis. NeRF employs a multilayer perceptron (MLP) to process the input 3D coordinates and view directions to predict the color and density of each point, as illustrated in Figure 2. Specifically, NeRF first casts a ray $r(t) = o + td$ from a given camera position $o = (x_0, y_0, z_0)$ along the viewing direction d. Along each ray, a series of sample points are uniformly sampled or using stratified sampling within a predefined depth range to obtain their coordinates. These coordinates, along with the view direction, are input into the MLP, which outputs the color and density for each sampled point. Finally, the volume rendering techniques integrate these color and density values along the ray to obtain the color value of the corresponding pixel in the image, as shown in Equation 1. During NeRF training, the MLP weights are optimized by minimizing the photometric loss, defined as the mean squared error between the rendered image and the actual observed image, to ensure that the generated images are as consistent as possible with the real observations.

$$C(r) = \int_{t_1}^{t_2} T(t) \cdot \sigma(\boldsymbol{r}(t)) \cdot c(\boldsymbol{r}(t), \boldsymbol{d}) \cdot dt \tag{1}$$

The transmittance function $T(t)$ represents the probability that a ray is not occluded by an object at a distance t. In physics, we assume that the scene is composed of a collection of luminous particles, with the density field $\sigma(x)$ describing the probability of a ray encountering particles at position x. The transmittance function $T(t)$ is described in Equation 2 as follows:

$$T(t) = \exp\left(-\int_{t_1}^{t} \sigma(\boldsymbol{r}(u)) \cdot du\right) \tag{2}$$

In the field of agricultural 3D reconstruction, NeRF can capture plant structures and scene details more accurately than traditional methods, providing more realistic and detailed reconstruction results. It is capable of adapting to complex vegetation structures and various farmland environments. However, the training and inference processes of NeRF typically require substantial computational resources and time, which do not fully align with the practical needs of agricultural production. Moreover, NeRF generates implicitly represented 3D scenes, which differ from traditional point cloud or mesh models, leading to challenges in integrating them with existing agricultural analysis tools.

To address the aforementioned issues, we chose an improved version of NeRF, the Nerfacto model, as our 3D reconstruction tool. Nerfacto retains the fine reconstruction capabilities of NeRF while optimizing computational efficiency and resource utilization, making it better suited for the practical needs of agricultural

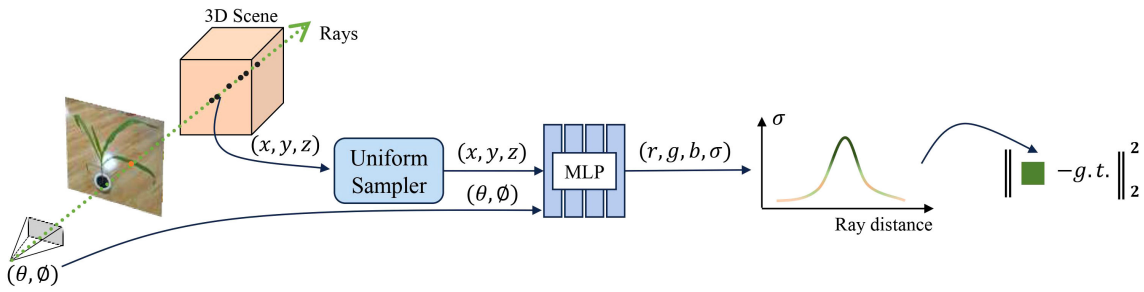


FIGURE 2
NeRF-based 3D representation pipeline.

applications. Next, we will discuss the structure of the Nerfacto model in detail and how to extract point cloud data from its implicit representation to effectively apply it to plant phenotyping tasks.

2.3.2 Network structure of Nerfacto and point-cloud extraction method

To enhance the efficiency and effectiveness of the sampling process, as shown in Figure 3, Nerfacto (Tancik et al., 2023) employs a segmented sampling strategy. This sampler performs uniform sampling within a fixed distance from the camera and applies distributed sampling at greater distances. This approach not only maintains dense sampling for nearby objects but also improves sampling efficiency for distant objects. Additionally, the sampling process emphasizes regions of the scene that contribute most to the final rendering. Unlike NeRF, which uses an MLP model for predictions, Nerfacto processes the input 3D coordinates and view directions by combining hash encoding (HA) and spherical harmonic encoding (SH), and can also predict the scene's density (σ) and color (r, g, b). Moreover, the appearance embedding vectors further enhance the model's adaptability to different viewpoints and lighting conditions. These improvements make Nerfacto significantly superior to NeRF and other NeRF-based models in terms of accuracy, efficiency, and applicability, We thought that this makes it better suited for practical agricultural 3D reconstruction applications.

For each ray, we have a set of sample points, with each point having a known depth d_i. The model, once trained, learns the density and color information of each sample point in the scene. Using the volume rendering equation, it calculates the weight w_i for each sample point. Specifically, as derived from Equation 2, the accumulated transmittance T_i for each sample point represents the probability that the light has not been obstructed from the start of the ray to the current point. For the $i-th$ sample point, the accumulated transmittance is computed as shown in Equations 3, 4, where the transmittance α_i of each sample point is derived from its density σ_i and the sampling interval δ_i. Thereby, from these calculations, the formula for the weight w_i can be determined, as shown in Equation 5.

$$T_i = \prod_{j=1}^{i-1}(1-\alpha_i) \tag{3}$$

$$\alpha_i = 1 - \exp(-\sigma_i \cdot \delta_i) \tag{4}$$

$$w_i = \prod_{j=1}^{i-1}\exp(-\sigma_i \cdot \delta_i)\cdot(1-\exp(-\sigma_i \cdot \delta_i)) \tag{5}$$

Finally, the formula for computing the expected depth $t_{surface}$ is given as shown in Equation 6.

$$t_{surface} = \frac{\sum_{i=1}^{N} d_i w_i}{\sum_{i=1}^{N} w_i} \tag{6}$$

where N is the number of points sampled along the ray.

After obtaining the approximate depth, the next step is to compute the position of the point on the ray $r(t) = o + t\mathbf{d}$ within the world coordinate system, as shown in Equation 7.

$$p = \boldsymbol{o} + t_{surface} \times \mathbf{d} \tag{7}$$

where p represents the position of the 3D point, o is the spatial coordinate of the camera, and d is the direction of the ray.

Once the 3D position is determined, the color value at that location can be directly queried from the trained model. At this stage, the generated point cloud reflects only a single viewpoint of the original scene. By repeating these steps for each pixel across multiple images, and mapping the corresponding coordinates and color information of each sample point in the scene to the world coordinate system, a dense point cloud is generated (Hu et al., 2024). This process can completes the 3D reconstruction of the plant.

2.4 Lightweight plant point cloud segmentation network

2.4.1 Network architecture

PointNet++ (Qi et al., 2017) is a deep learning model designed for processing point cloud data and represents an advancement and refinement of PointNet. It introduces a hierarchical structure for handling point cloud data, including sampling layers, feature extraction layers, and aggregation layers. Each layer is responsible for processing point cloud information at different levels of granularity, progressively extracting higher-level features and thus enhancing the model's ability to learn and analyze point cloud

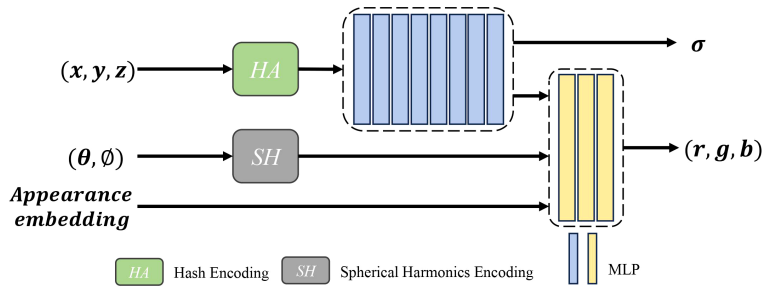


FIGURE 3
Structural components within the Nerfacto model.

features. However, despite the multi-scale processing introduced by PointNet++, the integration of global information may be insufficient, leading to a limited understanding of the overall point cloud structure and impacting the accuracy of analysis and predictions. Especially, for point clouds with sharp edges or complex local shape variations, PointNet++ may struggle to accurately capture and analyze local features, resulting in decreased segmentation and classification precision in these areas. To achieve strong generalization and performance, PointNet++ requires a substantial amount of labeled data for training to enhance model effectiveness and stability.

In recent years, the field of point cloud segmentation has emerged in many advanced feature extraction modules and model architectures. However, these complex structures often result in increased model parameters and computational costs (Guo et al., 2020), in contrasts with the need for lightweight and practical solutions for agricultural applications. To address this, we propose a lightweight plant organ segmentation network called PointSegNet. As illustrated in Figure 4, it follows a conventional encoderdecoder framework. Initially, data is processed through a multi-layer perceptron (MLP) module for preliminary feature extraction, mapping low-dimensional features to a higher-dimensional space. In the encoder section, we have redesigned the structure and introduced the Global-Local Set Abstraction (GLSA) module, which integrates both local and global information. The encoder, consisting of four GLSA modules, enhances the network's ability to capture point cloud features, extract multi-scale information, and reduce information loss during compression. Thereby, they improve the network's generalization and adaptability. For the decoder section, we have restructured the design and proposed the Edge-Aware Feature Propagation (EAFP) module, which is used for edge-preserving sharpening. This module can retains effectively and recovers detail features from the original resolution, thereby improving the model's understanding of semantic information, especially in terms of the detail and edge preservation. The network's final layer employs a segmentation head composed of two MLP layers to predict semantic labels for each point. Additionally, inspired by PointNext (Qian et al., 2022), we incorporate some advanced training strategies, including data augmentation techniques (such as point re-sampling, jittering, random scaling, and rotation) and some optimization strategies (such as loss functions, optimizers, learning rate scheduling, and hyperparameter tuning). These improvements significantly reduce the model's reliance on the quantity of training data and effectively mitigate overfitting issues, particularly in scenarios with limited data samples.

2.4.2 GLSA module for capturing local and global structural information in point clouds

As illustrated in Figure 4, the proposed Global-Local Set Abstraction (GLSA) module comprises four components: the FPS downsampling layer, the ResMLP module, the Relative Spatial Attention (RSA) module, and the Channel Attention Mechanism with Statistical Enhancement (CAM-SE). Initially, the FPS downsampling layer performs subsampling of the input point cloud using the farthest point sampling at a rate of 0.5. Subsequently, the ResMLP module and the RSA module extract the local and global features from the downsampled point cloud,

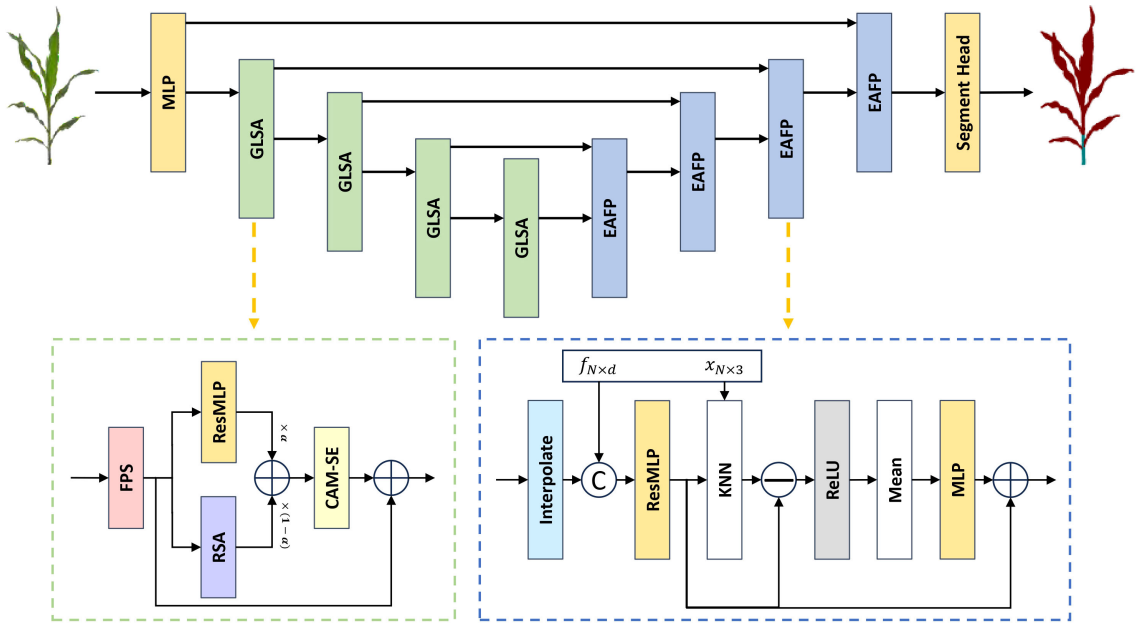


FIGURE 4
The overall architecture of PointSegNet, a U-net style architecture, has a Global-Local Set Abstraction (GLSA) module for downsampling and an Edge-Aware Feature Propagation (EAFP) module for upsampling. The $x_{N\times 3}$ and $f_{N\times d}$ in the Edge-Aware Feature Propagation (EAFP) module are jump connections from the encoder. $x_{N\times 3}$ denotes the spatial location information of the point cloud and $f_{N\times d}$ denotes the feature information of the point cloud.

respectively. Finally, the CAM-SE module is employed to mitigate the impact of noise and irrelevant information.

In the ResMLP module, as illustrated in Figure 5, the process begins with a ball query to obtain the k nearest points for each sampled point, thereby partitioning the point cloud data into multiple local regions. Subsequently, a two-layer MLP structure, implemented using 1×1 convolutions, projects the features of each point into a higher-dimensional space along the channel dimension,

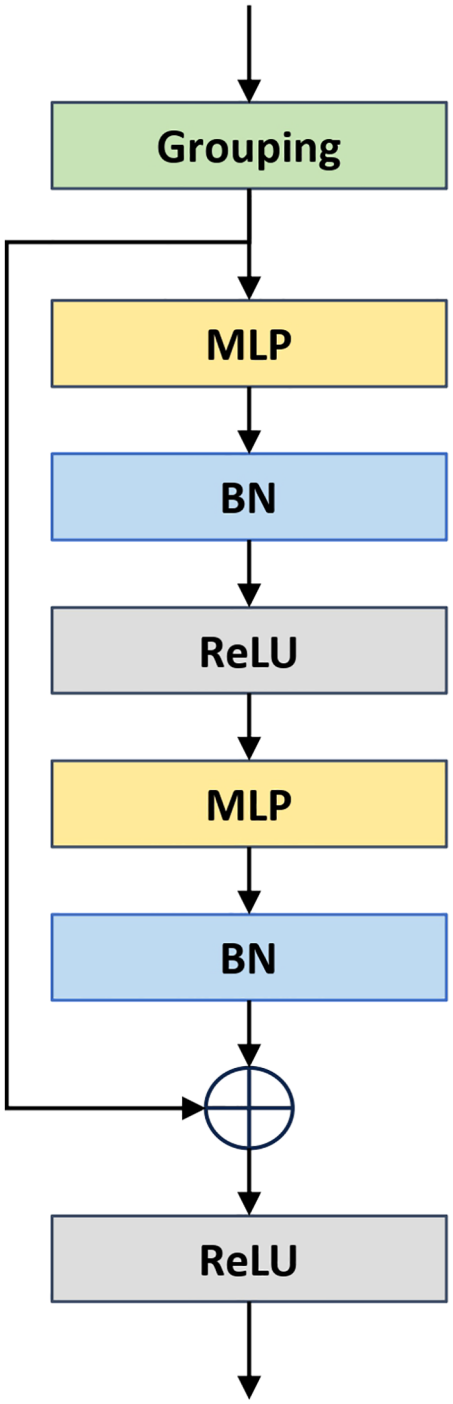


FIGURE 5
The illustration of Residual Multi-Layer (ResMLP) module.

which enhances the network's ability to represent local features. Residual connections enable the model to capture deeper features and help mitigate the issue of gradient vanishing.

As shown in Figure 6, the Relative Spatial Attention (RSA) module differs from standard global attention mechanisms by leveraging spatial and semantic proximity, and approximates the entire shape using only the currently downsampled points, which significantly reduces computational costs. For the input point cloud data p_{in} and f_{in}, the coordinates of all points are first averaged to obtain a global mean point P, serving as the global positional encoding of the point cloud. Concurrently, the features of each point are averaged to obtain a global feature mean F, providing a global perspective for understanding point cloud features. Each point's coordinates are then subtracted from the global mean P, yielding deviation vectors representing each point's position in the global coordinate system. Similarly, each point's features are subtracted from the global feature mean F, yielding deviation features representing local feature variations relative to the global context. These deviation vectors P and deviation features F are concatenated and input into a multi-layer perceptron (MLP). After passing through a sigmoid function, the attention weights for each point are obtained. The original point cloud features are multiplied by these attention weights and processed through an MLP layer again to yield the final output of the global attention, as shown in Equations 8 to 10. By computing the global mean point and feature mean and determining the positional and feature deviations of each point relative to the global context, this module enhances the understanding of each point's global positional relationships and local feature variations, thereby improving the overall expressive capability of the point cloud.

$$P = p_{in} - \frac{1}{N}\sum_{i=1}^{N} p_{in}^{i} \tag{8}$$

$$F = f_{in} - \frac{1}{N}\sum_{i=1}^{N} f_{in}^{i} \tag{9}$$

$$f_{RAS} = MLP_2(f_{in} \cdot Sigmoid(MLP_1(P, F))) \tag{10}$$

Finally, we use a learnable parameter X to fuse local and global features to control their contributions during the fusion process. This allows the model to adjust to maximize the overall effectiveness of feature representation automatically. Given that each point in the point cloud data may have varying importance and contribution, we employ a CAM-SE module to dynamically adjust the significance of each feature channel using statistical information (mean and standard deviation) along with learned parameters. Specifically, for the input point cloud features f, we first calculate the mean and standard deviation of each channel across the entire point cloud and concatenate them into a new tensor t. We then apply a learnable parameter matrix to weight the tensor t, sum the weighted results along the channel dimension, and process them through batch normalization followed by a Sigmoid activation function to obtain the final attention coefficients g. The point cloud features F are then multiplied by the attention coefficients g, and a residual connection is added to the features F, resulting in

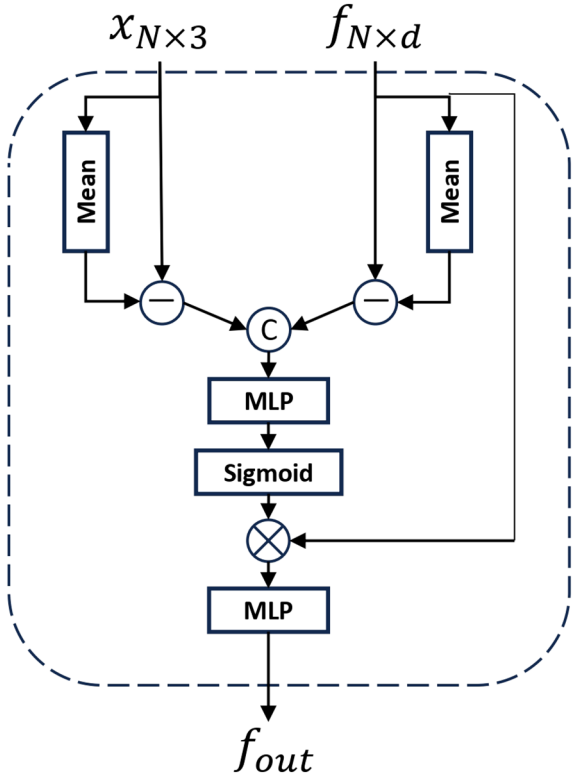


FIGURE 6
The illustration of Relative Spatial Attention (RSA) module. where $x_{N\times3}$ represents the spatial location information of the point cloud and $f_{N\times d}$ represents the feature information of the point cloud.

an enhanced feature representation through weight adjustments, as described in Equations 11–14.

$$u_i = \frac{1}{N}\sum_{j=1}^{N} f_{ij} \tag{11}$$

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(f_{ij} - u_i)^2 + \epsilon} \tag{12}$$

$$g_i = Sigmoid(BN(\sum_{j=1}^{N}([u_i, \sigma_i] \cdot W_i) \tag{13}$$

$$f_{\text{CAM-SE}} = g \cdot f + f \tag{14}$$

2.4.3 EAFP module for edge-aware capability

In point cloud segmentation models such as PointNet++, the decoder module has played a crucial role in the final segmentation results. It is responsible for progressively up-sampling the multiscale features extracted by the encoder, to recover them to the same resolution as the input point cloud, and finally generating predictive labels for each point. The design and implementation of the decoder module directly impact the accuracy and quality of the segmentation results. The process can be summarized as follows: firstly, the lower-resolution features are upsampled to higher resolutions using nearestneighbor interpolation; next, the upsampled features are fused with the skip connection features from the corresponding layers of the encoder; finally, the fused features undergo nonlinear transformation and processing through multilayer perceptrons (MLPs) to map the features to the desired output space (e.g., for classification, segmentation, or generation tasks).

In traditional decoder module designs, nearest-neighbor interpolation easily results in unsmooth interpolation outcomes, and introduce high-frequency noise, so that it makes the network training process difficult to converge, and disrupts the geometric consistency of the point cloud. It is the reason that we propose the Edge-Aware Feature Propagation (EAFP) module. Previous research has shown that edge-aware methods based on EdgeConv (Wang et al., 2019) can better extract local geometric structures through convolution operations. However, when handling deeper edge features, these methods can lead to increased memory consumption during training and inference.

As shown in Figure 4, The structure of our proposed Edge-Aware Feature Propagation (EAFP) module is illustrated as follows: First, upper-layer features are expanded to the current resolution through interpolation and then fused with the skip connection features $f_{N\times b}$ from the corresponding encoder layer to retain high-level semantic information. Next, the features are further processed by ResMLP to enhance their expressive power and non-linear characteristics. Subsequently, by incorporating the positional information $x_{N\times3}$ of the skip connection points, K-Nearest Neighbors (KNN) is used to retrieve the $K = 16$ nearest neighbor features for each point. The original features are then subtracted from these nearest neighbor features to obtain edge features, which are smoothed and enhanced using ReLU activation and Mean operations, improving robustness to outliers (Xiu et al., 2023). Finally, a layer of MLP maps the processed features to the final output space. This module leverages the spatial relationships between points in the point cloud to enhance task-relevant edges while smoothing irrelevant outliers, thereby better preserving and processing useful information and reducing the interference of unrelated elements on the results.

2.4.4 Loss function

During the model training process, we use PolyFocalLoss (Leng et al., 2022) as the loss function, as shown in Equation 15. PolyFocalLoss is an advanced loss function that integrates focal loss with a polynomial weighting mechanism. This combination effectively addresses the class imbalance in point cloud segmentation, significantly enhancing the model's ability to learn from challenging samples and improving segmentation accuracy.

$$L_{Ploy} = L_{FL} + \epsilon \cdot (1 - p_t)^{\gamma+1} \tag{15}$$

In this context, L_{FL} represents the focal loss function, P_t denotes the weighted sum of the predicted probabilities and the true labels, γ is the focus factor, which is used to adjust the weights of the difficult and easy recognition samples, and ϵ is a small constant used to prevent numerical instability or excessively small values. The specific process is detailed in Equations 16 to 18.

$$L_{FL} = L_{CE} \cdot (1 - P_t)^{\gamma} \tag{16}$$

$$L_{CE} = -[labels \cdot \log(p) + (1 - labels) \cdot \log(1 - p)] \tag{17}$$

$$P_t = labels \cdot p + (1 - labels) \cdot (1 - p) \tag{18}$$

Among them, p is the predicted probability after activation by the sigmoid activation function.

2.4.5 Experiment settings and evaluation metrics

We will train and test the point cloud segmentation network on three different plant point cloud datasets. To ensure full utilization of the data, we divided the datasets into an 80% training set and a 20% validation set. Before inputting the model, we selected 2048 points using an FPS downsampling strategy to serve as the number of input data points for the model. We implemented our method on a single NVIDIA GTX 3060Ti GPU using the PyTorch framework. During training, the batch size is set to 4. We choose the AdamW optimizer and set the weight decay to 1×10^{-4} to optimize the parameters in training and effectively suppress overfitting. The training period of the experiment is 300 epochs. To manage the learning rate more accurately, we adopt the multistep scheduler. During the training process, the learning rate is decayed at a rate of 0.1 when 210 and 270 epochs are reached. This decay strategy of multistep learning rate helps to maintain good convergence of the model at different training stages.

For data augmentation, we use a combination of random scaling, normalization, and jittering techniques. Specifically, random scaling adjusts the scale of each point cloud to a range between 0.8 and 1.2, aiding the model in learning features at various scales. Point clouds are also centralized and normalized to ensure consistency in the coordinate system and numerical range across all data before training. To introduce additional randomness and noise, we apply jittering technique to each point by adding independent Gaussian noise, with a standard deviation of 0.001 and a noise threshold of 0.005. This can simulates real-world noise and uncertainty, enhancing the model's robustness to noise and variations. The effectiveness of these data augmentation methods in improving point cloud segmentation accuracy has been validated in the PointNeXt (Qian et al., 2022) model.

This paper evaluates the performance of our proposed point cloud segmentation network, PointSegNet, by using a comprehensive set of metrics, including Intersection over Union (IoU), Precision, Recall, and F1-score, as well as model parameter count and computational complexity. Among them, IoU measures the overlap between the predicted region and the ground truth region, with higher values indicating greater alignment between the prediction and the actual situation. Precision assesses the proportion of true positive samples among all samples predicted by the model as positive class, reflecting the accuracy of the model's predictions. Recall measures the proportion of actual positive samples that are correctly identified by the model, highlighting its ability to detect all positive samples. The F1 score combines Precision and Recall into a single metric, providing a balanced evaluation of the model's performance in predicting both positive and negative samples. Specifically, the formulas for calculating these evaluation metrics are as shown in Equations 19–22.

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

$$F1 = \frac{2Precision \times Recall}{Precision + Recall} \tag{21}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{22}$$

Where TP (True Positive) denotes the number of positive class samples correctly predicted by the model, FP (False Positive) denotes the number of samples where the model incorrectly predicted a negative class as a positive class, and FN (False Negative) denotes the number of samples where the model failed to predict a positive class as a negative class.

2.4.6 Phenotypic trait extraction and evaluation

We extracted four phenotypic parameters—stem height, stem diameter, leaf length, and leaf width—from the segmented maize stems and leaves to evaluate the effectiveness of our proposed point cloud segmentation model (PointSegNet), in plant phenotyping tasks. The detailed parameter extraction process is as follows:

1. Stem height: Before measurement, the point cloud rotation correction is necessary. The principal direction of the stem is determined using PCA (Principal Component Analysis), and a rotation matrix aligning this direction with the Z-axis is computed. The entire point cloud is then rotated to be parallel to the Z-axis. Finally, the stem height parameter is obtained by subtracting the minimum z-value from the maximum z-value within the stem point cloud.
2. Stem diameter: First, the point cloud data is segmented along the Z-axis into four equal-height parts. The lowest segment is selected for analysis because the base of the stem is usually the most stable and uniform. Next, the linear fitting is performed on the point cloud data of the lowest z-value segment using the least squares method. Specifically, a linear regression model fits a plane that describes the variations in z-values in the X and Y directions, respectively. Then, we calculate the vertical distance from each point to the fitted plane (i.e., the difference between the actual z-value of the end and the z-value predicted in the plane). The stem diameter is defined as twice the median of these residuals' absolute values, as the median is insensitive to outliers and can more accurately reflect the true diameter of the stem (Miao et al., 2021).
3. Leaf length and width: To extract leaf length and width from the leaf point cloud, PCA (Principal Component Analysis) is first used to determine the primary direction of the point cloud, and the leaf point cloud is projected onto

this principal vector. The point cloud is then segmented based on these projection values. For each segment, the points corresponding to the minimum and maximum projection values are identified, as the extreme positions along the principal direction, i.e., the two most distant points. To accurately fit the leaf surface, multiple points are uniformly inserted between the start and end points, then projected onto the leaf point cloud. By connecting the start point, endpoint, and projection points, the width of the segment is determined. The maximum width among all segments is taken as the leaf's width. For leaf length calculation, the midpoints of the start and end points of each segment are identified and projected onto the leaf point cloud. By connecting all projection points, the total length of the leaf is obtained. Finally, smoothing is applied to remove noise and irregularities in the leaf length and width contours to improve measurement accuracy and data continuity. Later experimental result show that our method can be responsible for the more stable and reliable calculations of leaf width and length.

To assess the accuracy of the phenotypic parameters extracted from the PointSegNet model segmentation, we compared the predictions calculated from the model segmentation results with the actual measurements by measuring the actual values and calculating the R^2 coefficients and the RMSE, as shown in Equations 23, 24.

$$R^2 = 1 - \frac{\sum_{l=1}^{n}(v_l - \widehat{v_l})^2}{\sum_{l=1}^{n}(v_l - \overline{v_l})^2} \tag{23}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{l=1}^{n}(v_l - \hat{v}_l)^2} \tag{24}$$

where n is the number of sample points, v_l is the actual measured data, and $\hat{v}_l$ is the predicted value calculated from the model segmentation results. $\bar{v}_l$ is the average of all actual measurements.

3 Experiments and results

3.1 Nerfacto-based 3D reconstruction validation experiments

To evaluate the effectiveness of the Nerfacto-based 3D reconstruction method in plant phenotypic parameter extraction, we choose to use the point clouds generated by the mainstream Colmap (Schonberger and Frahm, 2016) open-source software and the 3DF Zephyr commercial software as the reference for the geometric extraction results. Colmap, known for its open-source nature and wide range of applications, is capable of providing highly accurate reconstruction results of point cloud data, while 3DF Zephyr, an industry-leading commercial software, is widely recognized for its fast processing and high-quality reconstruction capabilities. By comparing these reference point clouds, we aim to validate the reconstruction performance and applicability of the Nerfacto (Tancik et al., 2023) method in complex plant scenes, thus providing a new solution for the efficient acquisition of point cloud data.

In the experiments, we provide a detailed analysis from the following three aspects: reconstruction quality under limited viewpoints, visualization of 3D point cloud models, and reconstruction time. The experiments are divided into two groups: one using 40 images and the other using 60 images. As shown in Figure 7, Nerfacto is capable of generating high-quality 3D models even under conditions with limited viewpoints. In contrast, 3DF Zephyr and Colmap are significantly more limited in their reconstruction, and the point cloud models may be less complete and less detailed. In terms of reconstruction time, Nerfacto takes longer in the training phase and is less affected by the number of images. Colmap's dense reconstruction is very slow and consumes a lot of time mainly in stereo matching. In contrast, 3DF Zephyr's reconstruction is faster, but the generated dense point cloud is sparse and cannot capture all the details, and there may be voids and breaks in the point cloud.

Therefore, Nerfacto can be taken as a basic tool for 3D reconstruction of plants in agriculture. Later experimental results show that its high-quality reconstruction can help researchers analyze and characterize plant structures more precisely, providing strong support for plant phenotyping and related research.

3.2 Comparisons of different point cloud segmentation methods

3.2.1 Ablation study

In order to fully evaluate the performance and effectiveness of our proposed point cloud segmentation method, we conducted systematic ablation experiments on the Tomato dataset, as shown in Table 1. Our aim is to validate the contribution of each key component of the method to the final segmentation results and provide a basis for model optimization. Our ablation experiments verify the effectiveness of individual components in point cloud segmentation by removing or replacing these modules one by one.

The ablation study results demonstrate that the complete PointSegNet model achieves a mIoU of 92.525%, indicating its high segmentation performance. Removing individual modules leads to a decrease in model performance: removing the Relative Spatial Attention (RSA) module results in a 0.885% drop in mIoU; removing the channel attention module results in a 0.425% drop; replacing the Edge-Aware Feature Propagation (EAFP) module with the FP module from PointNet results in a 1.315% drop; and replacing the ResMLP module with the SA module from PointNet results in a 0.565% drop. Notably, removing both the RSA and EAFP modules simultaneously causes a significant performance drop of 2.795%. These results indicate that the RSA and EAFP modules make significant contributions to the model's performance, while the channel attention and ResMLP modules also provide some

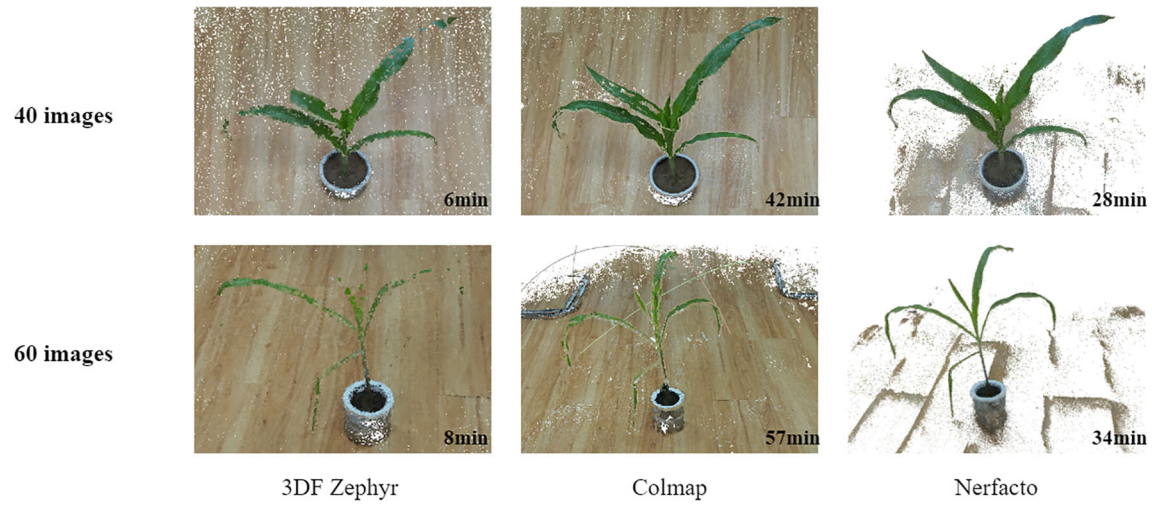


FIGURE 7
Visualization results of 3D reconstruction using Nerfacto, Colmap open source software and 3DF Zephyr commercial software at different numbers of images and the time taken for reconstruction.

performance improvement. Although removing these modules reduces the number of parameters and computational complexity to some extent, the significant performance drop demonstrates that these modules are essential for enhancing the segmentation performance of the PointSegNet model.

3.2.2 Point cloud segmentation results with different datasets

To comprehensively evaluate the performance of PointSegNet, we compared it with several advanced models, including PointNet++ (Qi et al., 2017), PCT (Guo et al., 2021), PointMLP (Ma et al., 2022), SPoTr (Park et al., 2023), and PointVector (Deng et al., 2023). Tables 2, 3 presents the quantitative comparison results for maize organ segmentation, where PointSegNet excels in efficiency metrics such as model parameters and throughput. In terms of segmentation metrics, including Intersection over Union (IoU), F1 score, and recall, PointSegNet surpasses the second-best model by approximately 1%. Although its precision is slightly lower than that of the PointVector model, PointSegNet's model parameter count and computational complexity are only one-third of those of the PointVector model. Table 4 further demonstrates that PointSegNet performs exceptionally well in handling plants with complex structures, such as tomatoes and soybeans, outperforming other models across all four average quantitative metrics. Experimental results indicate that PointSegNet exhibits strong robustness and versatility, reflecting its excellent generalization ability and feature extraction accuracy, which is attributed to its optimized model structure, effective feature fusion, and superior data preprocessing techniques in this paper.

To show the performance of the model in various plant structures more intuitively, we carried out a detailed visualization analysis of its segmentation results. As shown in Figure 8, the segmentation results of maize organs are visualized and analyzed at their different growth stages. PointSegNet has high segmentation accuracy, avoiding obvious mis-segmentation or segmentation omission, and the corresponding segmented stems and leaves have clear and smooth boundaries, and accurately show the actual shapes of them, meanwhile, maintaining a high degree of consistency at different growth stages. Additionally, in Figure 9, we show the visualization results of point cloud organ segmentation for tomato and soybean. For tomato and soybean, due to the complexity of the plant structure, all five models except

TABLE 1 Comparison of results from network model ablation experiments on the tomato point cloud dataset.

Ablate	mIoU (%)	Δ	Params (M)	FLOPS (G)
PointSegNet(ours)	92.525	–	1.33	4.73
– RSA	91.64	0.885	1.068	4.63
– CAM-SE	92.1	0.425	1.32	4.76
– EAFP	91.21	1.315	1.24	4.6
– ResMLP	91.96	0.565	1.152	2.72
– RSA & EAFP	89.73	2.795	0.981	4.5

– denotes removing from baseline.

TABLE 2 Comparison of segmentation performance on maize point cloud datasets.

		Maize		
		Leaf	Stem	Mean
ins. IoU (%)	PointNet++	83.27	98.84	91.055
	PCT	65.92	96.63	81.275
	PointMLP	83.49	98.71	91.1
	SPoTr	85.9	98.93	92.415
	PointVector	86.5	99.08	92.79
	Ours	**88.26**	**99.2**	**93.73**
Precision (%)	PointNet++	94.05	99.22	96.635
	PCT	69.88	**99.54**	84.71
	PointMLP	91.82	99.4	95.61
	SPoTr	94.95	99.37	97.16
	PointVector	**96.06**	99.33	**97.695**
	Ours	94.97	99.52	97.245
F1 score (%)	PointNet++	90.97	99.42	95.195
	PCT	79.99	98.29	89.14
	PointMLP	91.25	99.35	95.3
	SPoTr	92.61	99.47	96.04
	PointVector	92.84	99.54	96.19
	Ours	93.85	**99.6**	**96.725**
Recall (%)	PointNet++	88.1	99.62	93.86
	PCT	**93.53**	97.07	95.3
	PointMLP	90.69	99.31	95
	SPoTr	90.39	99.57	94.98
	PointVector	89.83	**99.75**	94.79
	Ours	92.75	99.67	**96.21**

The best and second-best results are highlighted in bold and underlined.

PointSegNet show mis-segmentation to varying degrees, which are mainly concentrated at the junction of stems and leaves as well as at the intersection of the main stem and branches. Especially in the dense leaf region, these models often fail to distinguish neighboring leaves due to their occlusion and overlapping correctly. In contrast, PointSegNet can effectively handle the challenges posed by these complex structures, maintains high segmentation accuracy and detail retention, and can accurately segment stems, leaves, main stems, and branches with clearer boundaries.

3.3 Evaluation of extracted traits based on point cloud segmentation

To validate the effectiveness of the proposed leaf length calculation method, we visualized each step of the process. As shown in Figure 10, we first computed the principal direction of the leaf point cloud using PCA and segmented the point cloud based on projection values. Next, we calculated the principal vector for each segment of the point cloud to identify the nearest and farthest

TABLE 3 Comparison of segmentation efficiency using PointSegNet and five other state-of-the-art networks.

	PointNet++	PCT	PointMLP	SPoTr	PointVector	Ours
Params (M)	1.74	2.88	16.71	25.56	4.07	1.32
FLOPs (G)	4.09	2.32	132.85	151.18	14	4.73

10.3389/fpls.2025.1491170

TABLE 4 Comparison of segmentation performance on tomato and soybean point cloud datasets.

		Tomato			Soybean			
		Leaf	Stem	Mean	Leaf	Stem	Branch	Mean
ins. IoU (%)	PointNet++	94.44	82	88.22	93.84	51.4	45.44	63.560
	PCT	93.07	77.19	85.13	92.79	51.61	29.17	57.857
	PointMLP	93.64	79.08	86.36	93.75	49.38	36.13	59.753
	SPoTr	94.51	83.1	88.805	94.79	57.37	43.05	65.070
	PointVector	95.8	87	91.4	94.86	62.86	48.26	68.660
	Ours	**96.45**	**88.6**	**92.525**	**95.53**	**63.14**	**51.12**	**69.930**
Precision (%)	PointNet++	96.75	90.71	93.73	96.13	66.54	71.11	77.927
	PCT	95.34	89.62	92.48	95.07	67.99	52.75	71.937
	PointMLP	95.74	90.87	93.305	96.29	62.75	57.28	72.107
	SPoTr	96.31	92.92	94.615	96.14	75.02	66.2	79.120
	PointVector	97.52	93.55	95.535	**96.79**	**75.61**	67.58	79.993
	Ours	**97.93**	**94.53**	**96.23**	96.52	73.74	**75.28**	**81.847**
F1 score (%)	PointNet++	97.14	89.82	93.48	96.83	69.94	66.36	77.710
	PCT	96.4	86.44	91.42	96.2	68.48	47.54	70.740
	PointMLP	96.72	88.06	92.39	96.7	65.84	54.36	72.300
	SPoTr	97.2	90.89	94.045	97.32	74.58	61.31	77.737
	PointVector	97.87	93.04	95.455	97.36	77.09	65.92	80.123
	Ours	**98.2**	**93.96**	**96.08**	**97.7**	**77.3**	**68.41**	**81.137**
Recall (%)	PointNet++	97.53	88.93	93.23	97.54	73.7	62.21	77.817
	PCT	97.48	83.48	90.48	97.36	68.98	43.27	69.870
	PointMLP	97.71	85.42	91.565	97.12	69.24	51.72	72.693
	SPoTr	98.11	88.95	93.53	98.52	74.15	57.1	76.590
	PointVector	98.22	92.52	95.37	97.94	78.62	**64.34**	80.300
	Ours	**98.47**	**93.4**	**95.935**	**98.9**	**81.23**	62.69	**80.940**

The best and second best results are marked in bold and underlined.

points. Due to the curvature of maize leaves, using the distance between the nearest and farthest points directly as the leaf width is inaccurate. Therefore, we propose to generate points uniformly between these two points and projecting them onto the leaf. By connecting the nearest point, projection points, and the farthest point, the total length of the resulting segments is taken as the leaf width, which better fits the shape of the leaf. For leaf length, we project the midpoints of the nearest and farthest points in each segment onto the segment's point cloud. We then connected the nearest points, farthest points, and projection points across the entire leaf point cloud to determine the total length of the leaf. Due to improper setting of the number of segments and the number of uniform point insertions, the resulting curves might exhibit excessive undulations, failing to accurately reflect the leaf parameters. To address this, we apply Gaussian smoothing to these points by using one-dimensional Gaussian filtering along the x, y, and z axes, respectively, to enhance data continuity and smoothness.

To assess the performance of point cloud segmentation results in plant phenotypic parameter extraction, we compared the measurement results obtained using our proposed method with those obtained through manual measurement. As shown in Figure 11a, the validation results about plant stem height produced R^2 and RMSE values of 0.99 and 0.33 cm, respectively. Figure 11b shows that the validation of plant stem diameters resulted have R^2 and RMSE values of 0.84 and 0.12 cm, respectively. The validation results for leaf length and width are presented in Figures 11c, d, where the leaf length has an R^2 value of 0.94 and an RMSE value of 2.95 cm, and the leaf width has an R^2 value of 0.87 and an RMSE value of 0.42 cm. These results indicate that the proposed method in this paper has high accuracy in extracting plant phenotypic parameters, effectively capturing and reflecting the actual characteristics of the plants.

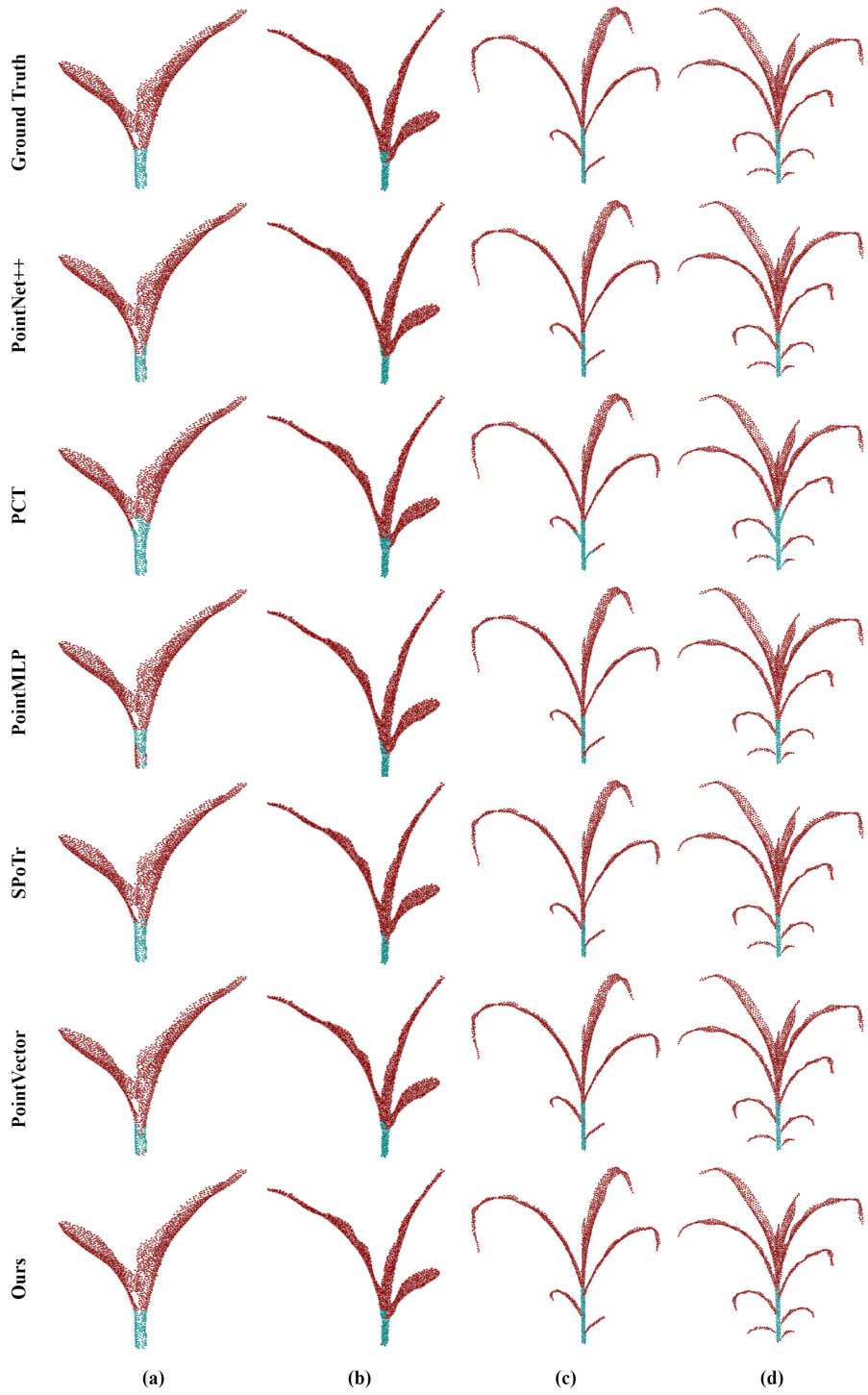


FIGURE 8
Qualitative visual analysis of organ segmentation using PointSegNet on maize point cloud datasets. **(a)**, **(b)**, **(c)**, **(d, e)** show the segmentation results and ground truth of maize point clouds for different growth cycles.

4 Discussion

This paper employs the Nerfacto method for the three-dimensional reconstruction of maize, extracting point clouds from implicit neural radiance fields. The Nerfacto method has low requirements for input data and is capable of generating high-quality 3D models with limited viewpoints and images, making it adaptable to various scenes and backgrounds, and suitable for the reconstruction of different plant types. During the image acquisition process, uneven lighting or strong reflections may lead to issues such as shadows, glare, or overexposure in the images, which could introduce errors in the reconstruction results and affect the final quality of the 3D model. Therefore, we will consider

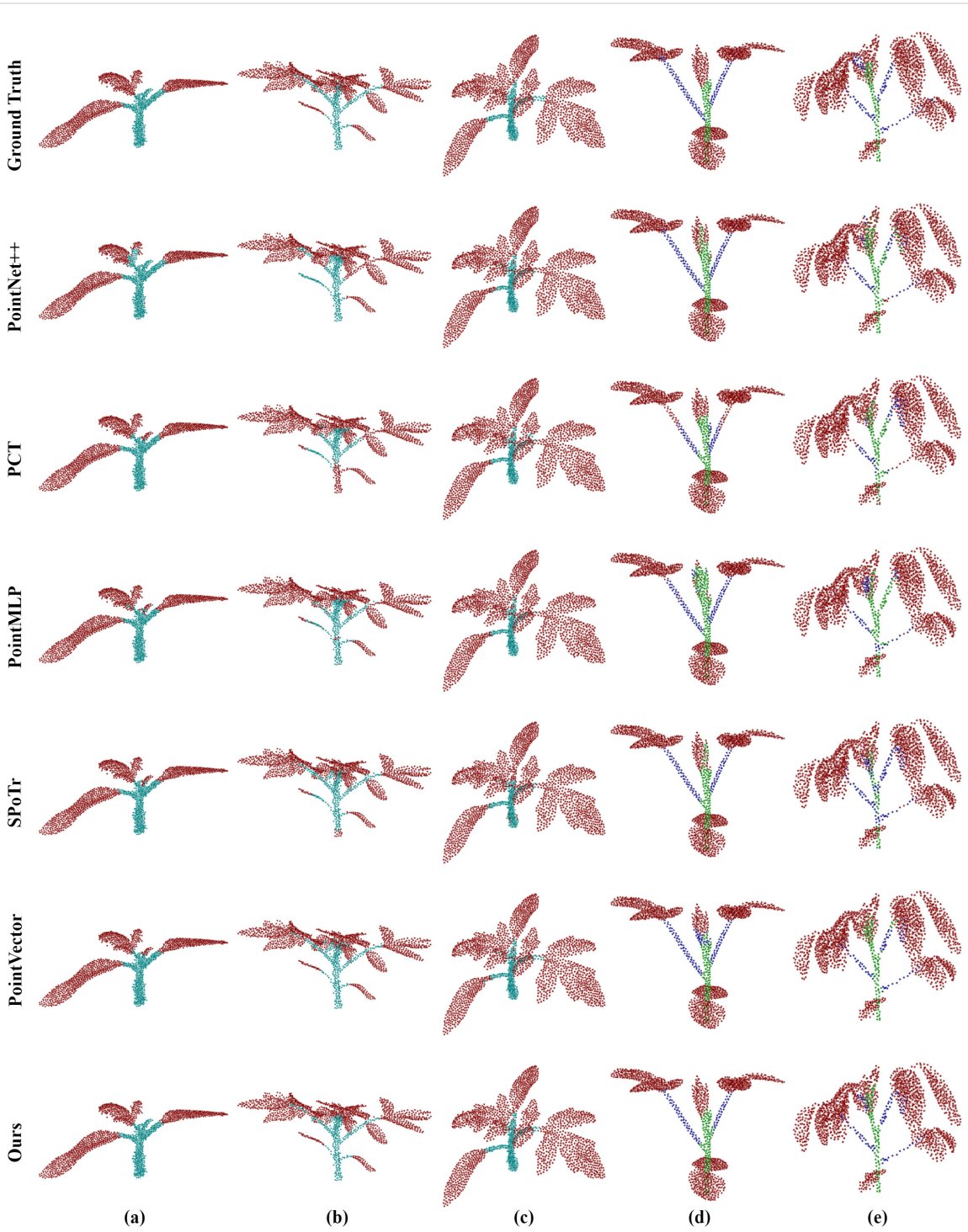


FIGURE 9
Qualitative visual analysis of complex plant organ segmentation using PointSegNet on tomato and soybean point cloud datasets. **(a–c)** show the segmentation results and ground truth for tomato, while **(d)** and **(e)** present the segmentation results and ground truth for soybean.

the impact of environmental lighting during data acquisition and processing, and employ appropriate image preprocessing techniques, such as exposure correction, to mitigate the interference of these factors on phenotyping measurements. Although the training and inference processes of Nerfacto can be completed using a single consumer-grade GPU, the entire procedure typically requires 20 to 30 minutes. Particularly when processing high-resolution images, the computational resource consumption and time required increase significantly. To address this issue, we plan to utilize CUDA programming to implement part or all of the computationally intensive tasks, thereby improving processing efficiency and accelerating the training process.

In our study, the PointSegNet model, which downsamples plant point clouds to 2048 points, has demonstrated excellent segmentation

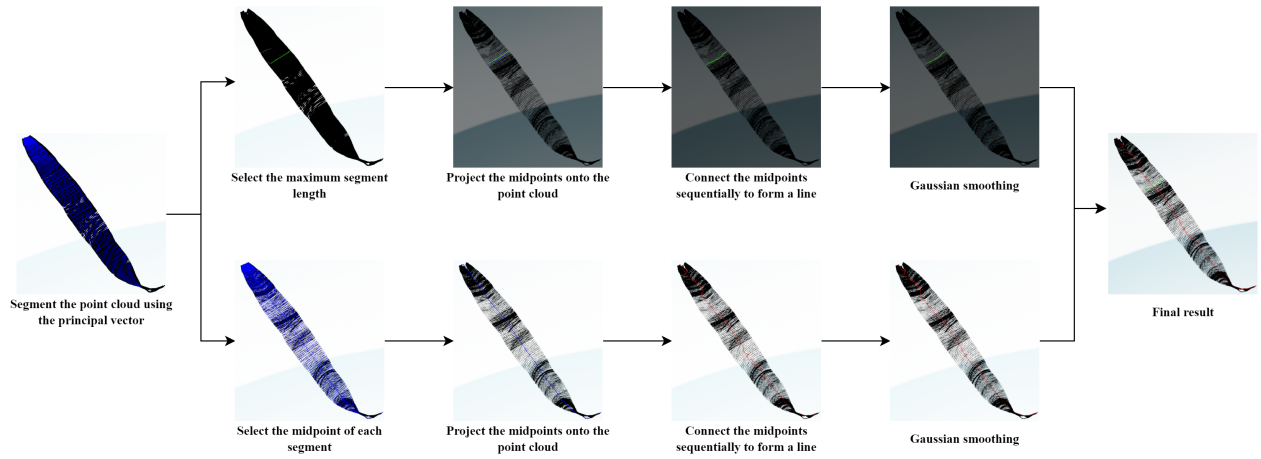


FIGURE 10
Visualization flowchart for leaf length and leaf width parameter extraction.

performance. However, for plants with complex structures, significant downsampling can compromise the original structural details, leading to information loss and makes it challenging for the network to learn useful features. Increasing the number of points in the plant point cloud can better preserve plant detail information, thereby improving segmentation accuracy. Nevertheless, this also introduces new challenges: higher point cloud resolution results in a substantial increase in the network's parameter count and computational load, and requires significant computational resources and time for training and inference. This may potentially lead to issues with insufficient computational resources and reduced efficiency in practical applications. Therefore, finding the optimal balance between model performance and computational resource requirements is a critical direction for future research. The effectiveness of plant point cloud segmentation generally will be improved with the richness of the dataset. However, point-by-point annotation is a time-consuming

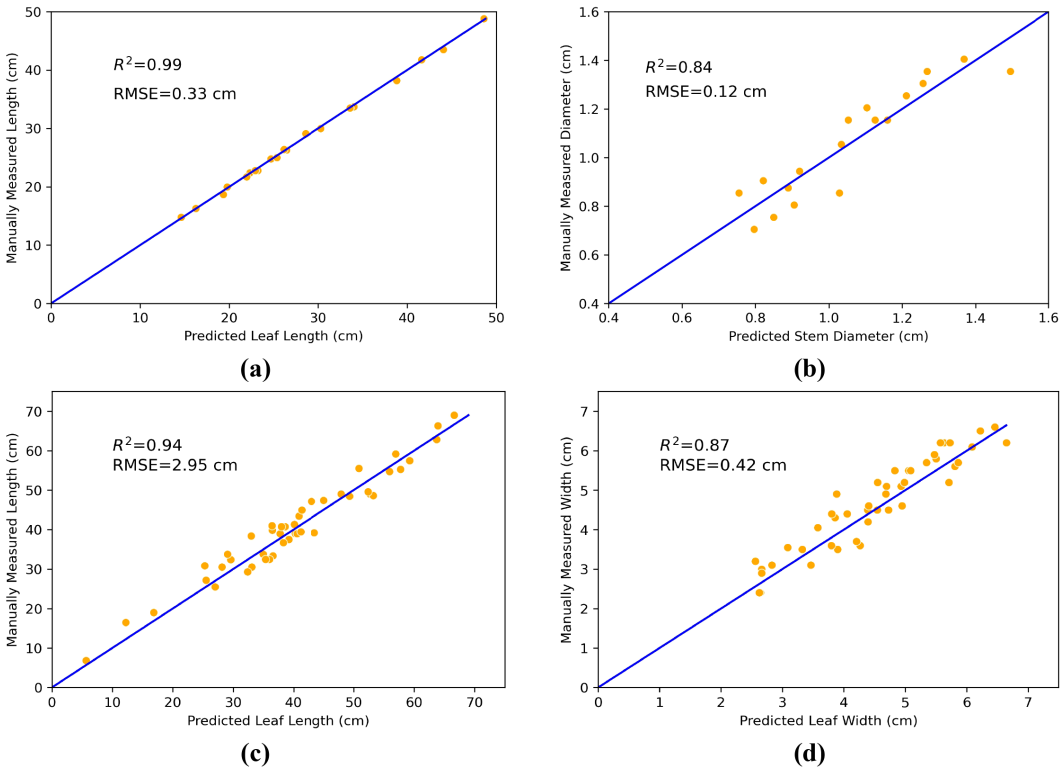


FIGURE 11
Comparison of extracted phenotypic parameters based on maize point cloud segmentation with measured values. **(a)** stem height, **(b)** stem diameter, **(c)** leaf length, and **(d)** leaf width.

and labor-intensive task, which limits the acquisition of large-scale, high-quality datasets. To enhance model performance with limited data, we will explore techniques such as semi-supervised learning, self-supervised learning, and transfer learning in future work. These approaches will not only accelerate model development and application but also reduce data annotation costs and time investment, thus providing substantial support for the advancement of agricultural phenotype measurement technologies.

In the process of extracting plant phenotype parameters, such as stem height, stem diameter, leaf length, and leaf width, the calculation often relies on methods that are dependent on key points. Traditional methods, which rely solely on the distance between the nearest and farthest points, may not accurately fit the curvature of the leaf when it is bent, resulting in an underestimation of leaf length. Our improved approach addresses this issue by segmenting the leaf along its principal direction, thereby enhancing the accuracy of the leaf curve fitting and resulting in more precise measurements of leaf length and width. However, excessive segmentation may lead to an overly curved fit, affecting the final measurement results. Therefore, it is essential to set the number of segments appropriately to ensure that the fitted curve closely approximates the actual shape of the leaf while maintaining measurement accuracy. Additionally, the presence of outliers is a common and significant issue. Outliers can arise from various sources, such as errors in point cloud acquisition equipment, environmental interference, similar situations occur in color between the plant and background, or errors in dataset annotations. These outliers may introduce substantial errors in phenotype parameter calculations, thereby affecting the final measurement results. Future research should focus on methods to prevent outlier interference during phenotype parameter computation to enhance the accuracy and reliability of measurements.

5 Conclusion

In this paper, a plant stem and leaf segmentation and phenotypic parameter extraction is proposed. Firstly, a novel 3D reconstruction method, Nerfacto, is employed to plant 3D reconstruction to explore the extraction of point clouds from implicit neural radiance fields. This approach provides a new avenue for obtaining point cloud data and reduces the reconstruction difficulty associated with complex plant scenes by using traditional methods. Secondly, a lightweight plant organ point cloud segmentation network, PointSegNet, is developed. This network features an efficient encoder-decoder structure that significantly enhances segmentation accuracy while maintaining low parameter counts and computational complexity. Experimental results on maize, tomato, and soybean datasets demonstrate that PointSegNet outperforms existing advanced networks in segmentation performance, with higher accuracy and robustness. Lastly, the PCA-based methods are proposed for calculating leaf length and leaf width parameters by determining the primary direction of the leaf point cloud. Experimental results indicate that the proposed methods achieve high precision in extracting phenotypic parameters such as stem length, stem diameter, leaf length, and leaf width. In the future, we aim to provide stronger support for plant phenotyping research and agricultural applications by continuously optimizing and expanding these technologies to develop automated tools in smart agriculture.

Data availability statement

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

Author contributions

GQ: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. ZZ: Writing – review & editing. BN: Writing – review & editing. SH: Writing – review & editing. EY: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The Inner Mongolia Natural Science Foundation project (No.2024MS06019), the Major Special Project of the Inner Mongolia Autonomous Region (No.2021SZD0043), Inner Mongolia University independent project (No.21500-5247012) and the fund of supporting the reform and development of local universities (Disciplinary construction).

Acknowledgments

We thank Bin Niu, Sijia Han, and Enhui Yang for their discussions and help during data acquisition and research.

Conflict of interest

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

Publisher's note

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

References

Deng, X., Zhang, W., Ding, Q., and Zhang, X. (2023). "Pointvector: a vector representation in point cloud analysis," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada. (New York: IEEE), 9455–9465.

Du, J., Lu, X., Fan, J., Qin, Y., Yang, X., and Guo, X. (2020). Image-based high-throughput detection and phenotype evaluation method for multiple lettuce varieties. *Front. Plant Sci.* 11, 563386. doi: 10.3389/fpls.2020.563386

Furukawa, Y., and Ponce, J. (2009). Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 1362–1376. doi: 10.1109/TPAMI.2009.161

Guo, M.-H., Cai, J.-X., Liu, Z.-N., Mu, T.-J., Martin, R. R., and Hu, S.-M. (2021). Pct: Point cloud transformer. *Comput. Visual Media* 7, 187–199. doi: 10.1007/s41095-021-0229-5

Guo, Q., Jin, S., Li, M., Yang, Q., Xu, K., Ju, Y., et al. (2020). Application of deep learning in ecological resource research: Theories, methods, and challenges. *Sci. China Earth Sci.* 63, 1457–1474. doi: 10.1007/s11430-019-9584-9

Hu, Y., Wang, L., Xiang, L., Wu, Q., and Jiang, H. (2018). Automatic non-destructive growth measurement of leafy vegetables based on kinect. *Sensors* 18, 806. doi: 10.3390/s18030806

Hu, K., Ying, W., Pan, Y., Kang, H., and Chen, C. (2024). High-fidelity 3d reconstruction of plants using neural radiance fields. *Comput. Electron. Agric.* 220, 108848. doi: 10.1016/j.compag.2024.108848

Huang, T., Bian, Y., Niu, Z., Taha, M. F., He, Y., and Qiu, Z. (2024). Fast neural distance field-based three-dimensional reconstruction method for geometrical parameter extraction of walnut shell from multiview images. *Comput. Electron. Agric.* 224, 109189. doi: 10.1016/j.compag.2024.109189

Leng, Z., Tan, M., Liu, C., Cubuk, E. D., Shi, X., Cheng, S., et al. (2022). Polyloss: A polynomial expansion perspective of classification loss functions. *arXiv preprint arXiv:2204.12511*. doi: 10.48550/arXiv.2204.12511

Li, Z., Guo, R., Li, M., Chen, Y., and Li, G. (2020). A review of computer vision technologies for plant phenotyping. *Comput. Electron. Agric.* 176, 105672. doi: 10.1016/j.compag.2020.105672

Luo, J., Zhang, D., Luo, L., and Yi, T. (2024). Pointresnet: A grape bunches point cloud semantic segmentation model based on feature enhancement and improved pointnet++. *Comput. Electron. Agric.* 224, 109132. doi: 10.1016/j.compag.2024.109132

Ma, X., Qin, C., You, H., Ran, H., and Fu, Y. (2022). Rethinking network design and local geometry in point cloud: A simple residual mlp framework. *arXiv preprint arXiv:2202.07123*. doi: 10.48550/arXiv.2202.07123

Miao, Y., Wang, L., Peng, C., Li, H., Li, X., and Zhang, M. (2022). Banana plant counting and morphological parameters measurement based on terrestrial laser scanning. *Plant Methods* 18, 66. doi: 10.1186/s13007-022-00894-y

Miao, T., Zhu, C., Xu, T., Yang, T., Li, N., Zhou, Y., et al. (2021). Automatic stem-leaf segmentation of maize shoots using three-dimensional point cloud. *Comput. Electron. Agric.* 187, 106310. doi: 10.1016/j.compag.2021.106310

Mildenhall, B., Srinivasan, P. P., Ortiz-Cayon, R., Kalantari, N. K., Ramamoorthi, R., Ng, R., et al. (2019). Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Trans. Graphics (ToG)* 38, 1–14. doi: 10.1145/3306346.3322980

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 99–106. doi: 10.1145/3503250

Moulon, P., Monasse, P., and Marlet, R. (2013). "Adaptive structure from motion with a contrario model estimation," in *Lecture Notes in Computer Science*, Daejeon, South Korea, Vol. 11. 257–270 (Berlin, Germany: Springer-Verlag), Revised Selected Papers, Part IV.

Park, J., Lee, S., Kim, S., Xiong, Y., and Kim, H. J. (2023). "Self-positioning point-based transformer for point cloud understanding," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada. (New York: IEEE), 21814–21823.

Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Adv. Neural Inf. Process. Syst.* 30, 5105–5114. doi: 10.48550/arXiv.1706.02413

Qian, G., Li, Y., Peng, H., Mai, J., Hammoud, H., Elhoseiny, M., et al. (2022). Pointnext: Revisiting pointnet++ with improved training and scaling strategies. *Adv. Neural Inf. Process. Syst.* 35, 23192–23204. doi: 10.48550/arXiv.2206.04670

Qiao, Y., Liao, Q., Zhang, M., Han, B., Peng, C., Huang, Z., et al. (2023). Point clouds segmentation of rapeseed siliques based on sparse-dense point clouds mapping. *Front. Plant Sci.* 14, 1188286. doi: 10.3389/fpls.2023.1188286

Rawat, S., Chandra, A. L., Desai, S. V., Balasubramanian, V. N., Ninomiya, S., and Guo, W. (2022). How useful is image-based active learning for plant organ segmentation? *Plant Phenom.* 2002. doi: 10.34133/2022/9795275

Schonberger, J. L., and Frahm, J.-M. (2016). "Structure-from-motion revisited," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA, (New York: IEEE). 4104–4113.

Schunck, D., Magistri, F., Rosu, R. A., Cornelißen, A., Chebrolu, N., Paulus, S., et al. (2021). Pheno4d: A spatio-temporal dataset of maize and tomato plant point clouds for phenotyping and advanced plant analysis. *PloS One* 16, e0256340. doi: 10.1371/journal.pone.0256340

Shi, M., Zhang, S., Lu, H., Zhao, X., Wang, X., and Cao, Z. (2022). Phenotyping multiple maize ear traits from a single image: Kernels per ear, rows per ear, and kernels per row. *Comput. Electron. Agric.* 193, 106681. doi: 10.1016/j.compag.2021.106681

Sun, Y., Zhang, Z., Sun, K., Li, S., Yu, J., Miao, L., et al. (2023). Soybean-mvs: annotated three-dimensional model dataset of whole growth period soybeans for 3d plant organ segmentation. *Agriculture* 13, 1321. doi: 10.3390/agriculture13071321

Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Wang, T., et al. (2023). "Nerfstudio: A modular framework for neural radiance field development," in *SIGGRAPH '23: ACM SIGGRAPH 2023 Conference Proceedings*, Los Angeles, USA. (New York: ACM). Vol. 1–12.

Tardieu, F., Cabrera-Bosquet, L., Pridmore, T., and Bennett, M. (2017). Plant phenomics, from sensors to knowledge. *Curr. Biol.* 27, R770–R783. doi: 10.1016/j.cub.2017.05.055

Thapa, S., Zhu, F., Walia, H., Yu, H., and Ge, Y. (2018). A novel lidar-based instrument for high-throughput, 3d measurement of morphological traits in maize and sorghum. *Sensors* 18, 1187. doi: 10.3390/s18041187

Wang, F., Mohan, V., Thompson, A., and Dudley, R. (2020). "Dimension fitting of wheat spikes in dense 3d point clouds based on the adaptive k-means algorithm with dynamic perspectives," in *2020 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*. Vancouver, BC, Canada (New York: IEEE). 144–148.

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., and Solomon, J. M. (2019). Dynamic graph cnn for learning on point clouds. *ACM Trans. Graphics (tog)* 38, 1–12. doi: 10.1145/3326362

Wu, S., Wen, W., Gou, W., Lu, X., Zhang, W., Zheng, C., et al. (2022). A miniaturized phenotyping platform for individual plants using multi-view stereo 3d reconstruction. *Front. Plant Sci.* 13, 897746. doi: 10.3389/fpls.2022.897746

Wu, S., Zhang, Y., Zhao, Y., Wen, W., Wang, C., Lu, X., et al. (2024). Using high-throughput phenotyping platform mvs-pheno to decipher the genetic architecture of plant spatial geometric 3d phenotypes for maize. *Comput. Electron. Agric.* 225, 109259. doi: 10.1016/j.compag.2024.109259

Xiu, H., Liu, X., Wang, W., Kim, K.-S., Shinohara, T., Chang, Q., et al. (2023). Diffusion unit: Interpretable edge enhancement and suppression learning for 3d point cloud segmentation. *Neurocomputing* 559, 126780. doi: 10.1016/j.neucom.2023.126780

Xu, Y., Wang, G., Shao, L., Wang, N., She, L., Liu, Y., et al. (2023). Glandsegnet: Semantic segmentation model and area detection method for cotton leaf pigment glands. *Comput. Electron. Agric.* 212, 108130. doi: 10.1016/j.compag.2023.108130

Yan, J., Tan, F., Li, C., Jin, S., Zhang, C., Gao, P., et al. (2024). Stem–leaf segmentation and phenotypic trait extraction of individual plant using a precise and efficient point cloud segmentation network. *Comput. Electron. Agric.* 220, 108839. doi: 10.1016/j.compag.2024.108839

Yang, R., He, Y., Lu, X., Zhao, Y., Li, Y., Yang, Y., et al. (2024b). 3d-based precise evaluation pipeline for maize ear rot using multi-view stereo reconstruction and point cloud semantic segmentation. *Comput. Electron. Agric.* 216, 108512. doi: 10.1016/j.compag.2023.108512

Yang, X., Miao, T., Tian, X., Wang, D., Zhao, J., Lin, L., et al. (2024c). Maize stem–leaf segmentation framework based on deformable point clouds. *ISPRS J. Photogram. Remote Sens.* 211, 49–66. doi: 10.1016/j.isprsjprs.2024.03.025

Yang, D., Yang, H., Liu, D., and Wang, X. (2024a). Research on automatic 3d reconstruction of plant phenotype based on multi-view images. *Comput. Electron. Agric.* 220, 108866. doi: 10.1016/j.compag.2024.108866

Yao, Y., Luo, Z., Li, S., Fang, T., and Quan, L. (2018). "Mvsnet: Depth inference for unstructured multi-view stereo," in *Lecture Notes in Computer Science*, Munich, Germany. (Berlin, Germany: Springer). 767–783.

Yu, S., Liu, X., Tan, Q., Wang, Z., and Zhang, B. (2024). Sensors, systems and algorithms of 3d reconstruction for smart agriculture and precision farming: A review. *Comput. Electron. Agric.* 224, 109229. doi: 10.1016/j.compag.2024.109229

Zhang, Z., Jin, X., Rao, Y., Wan, T., Wang, X., Li, J., et al. (2024). Dsbean: An innovative framework for intelligent soybean breeding phenotype analysis based on various main stem structures and deep learning methods. *Comput. Electron. Agric.* 224, 109135. doi: 10.1016/j.compag.2024.109135

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 02 April 2025
DOI 10.3389/fpls.2025.1515021

Check for updates

OPEN ACCESS

EDITED BY
Jun Ni,
Nanjing Agricultural University, China

REVIEWED BY
Nisha Pillai,
Mississippi State University, United States
Pushkar Gole,
University of Delhi, India
Sajjad Saleem,
Lahore Campus, Pakistan

*CORRESPONDENCE
Jianwen Guo
guojw@dgut.edu.cn

RECEIVED 22 October 2024
ACCEPTED 13 March 2025
PUBLISHED 02 April 2025

CITATION
Zhang H, Li S, Xie J, Chen Z, Chen J and Guo J (2025) VMamba for plant leaf disease identification: design and experiment. *Front. Plant Sci.* 16:1515021. doi: 10.3389/fpls.2025.1515021

COPYRIGHT
© 2025 Zhang, Li, Xie, Chen, Chen and Guo. 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.

VMamba for plant leaf disease identification: design and experiment

Hewei Zhang, Shengzhou Li, Jialong Xie, Zihan Chen, Jiyang Chen and Jianwen Guo*

School of Mechanical Engineering, Dongguan University of Technology, Dongguan, China

Introduction: The rapid spread of crop diseases poses a severe threat to agricultural production, significantly reducing both the yield and quality of crops. In recent years, plant disease recognition technologies based on machine vision and artificial intelligence have made significant progress. However, current mainstream deep learning architectures still face numerous challenges in detecting agricultural plant diseases. These include issues such as the complexity of agricultural environments and the reduced accuracy and increased training time caused by small sample sizes of agricultural plant diseases.

Methods: To address these challenges, we introduce the VMamba visual backbone model into the task of detecting agricultural plant diseases. This model effectively reduces computational complexity through a selective scanning mechanism while significantly improving classification accuracy by maintaining a global receptive field and leveraging dynamic weighting advantages. Our study proposes the DDHTLVMamba method, which combines VMamba with diffusion models and transfer learning techniques, and applies it to the detection of plant diseases in small-sample agricultural datasets. This research evaluates the performance of VMamba across different datasets and training methods, conducting comparative analyses with mainstream deep learning architectures.

Results and discussion: Experimental results demonstrate that the VMamba model outperforms mainstream models such as ResNet50, Vision Transformer, and Swin Transformer in disease recognition accuracy, whether on large-scale datasets like PlantVillage or optimized small-sample disease datasets, showcasing superior performance. Compared to Swin Transformer, VMamba achieves a 3.49% increase in accuracy while reducing training time by 80%. Furthermore, the proposed DDHTLVMamba training method demonstrates its effectiveness on small-sample datasets, significantly reducing pre-training time while maintaining recognition accuracy comparable to that achieved with large-sample transfer learning. This study provides an innovative approach for the efficient identification of agricultural diseases and is expected to advance the development of intelligent agricultural disease prevention and control technologies.

KEYWORDS

plant leaf disease, VMamba, transformer, diffusion model, transfer learning, few-shot learning

1 Introduction

Crop diseases pose a serious threat to agricultural production, typically caused by bacteria, fungi, microorganisms, or viruses (Sankaran et al., 2010). Once a disease occurs, it spreads rapidly, often leading to significant decreases in crop yield and quality (Thomas et al., 2018). Therefore, accurate and rapid identification and classification of crop diseases are crucial for effective disease prevention (Shoaib et al., 2023). The physiological state of plants can often be reflected through the leaves (Joshi and Bhavsar, 2020), and changes in leaves accurately indicate the growth conditions of plants and potential diseases. In recent years, significant progress has been made in leaf disease identification using machine vision and artificial intelligence. These technologies reduce the need for operators to have specialized knowledge in agricultural disease analysis, thereby reducing learning costs and labor consumption, while being easy to transfer and widely applicable (Li et al., 2021). In particular, deep learning visual models have excelled in handling complex inputs and classification tasks, having been successfully applied to disease identification in crops such as maize, wheat, citrus, and potatoes (Lee et al., 2020).

However, images of agricultural plant diseases often present challenges like complex backgrounds and unclear diseased areas, especially when sample sizes are insufficient, resulting in a significant drop in recognition performance, which poses a major challenge for computer vision (Mohanty et al., 2016). The introduction of deep residual networks (ResNet) addressed the gradient vanishing problem in deep networks, making it possible to construct deeper learning architectures (He et al., 2016). However, training and inference with ResNet require substantial computational resources, and there is still considerable redundancy within the network. In contrast, the Vision Transformer (ViT) adopts a self-attention mechanism that can process all positions in the input sequence in parallel, significantly improving training speed and inference efficiency (Dosovitskiy et al., 2020). However, with an increase in input scale, the computational complexity of ViT grows quadratically, especially when dealing with high-resolution tasks, leading to significant computational costs, while in agricultural applications, a balance between performance and efficiency is essential.

The VMamba visual backbone model, based on state space models (SSMs) and a selective scanning mechanism, provides an alternative solution to ViT for computer vision (Gu and Dao, 2023). The selective scanning mechanism in the VMamba model effectively reduces the complexity of attention computation. Compared to the ViT model, VMamba maintains the advantages of global effective receptive fields (ERF) and dynamic weights while achieving linear computational complexity and demonstrating excellent classification accuracy (Liu et al., 2024). Therefore, we aim to leverage the strengths of VMamba in the complex task of agricultural plant disease identification to not only maintain high recognition accuracy but also effectively reduce computational resource and data requirements.

The main objective of this paper is to introduce the VMamba visual backbone model into the field of agricultural plant disease identification, investigating its effectiveness in crop disease identification. Specifically, this study focuses on: (1) verifying the performance advantages of the VMamba model in agricultural disease classification; (2) exploring the effectiveness of VMamba in small-sample plant disease identification and its performance in complex crop growth environments; (3) examining the transfer learning performance of VMamba in agricultural disease identification, in combination with diffusion models and transfer learning; and (4) designing the DDHTL-VMamba model to further enhance recognition performance in complex agricultural environments through diffusion models and transfer learning, providing a new approach for agricultural disease prevention.

The remainder of this paper is organized as follows: Section 2 introduces related research work; Section 3 presents the proposed method; Section 4 introduces the experiments; Section 5 presents related discussions; and Section 6 provides the conclusions and future work.

2 Related works

The identification and classification of plant leaf images are considered effective means for disease identification (Shoaib et al., 2023). Hossain et al. divided the machine learning identification process of plant leaves into three steps: image preprocessing, feature extraction, and classification (Hossain et al., 2019). However, plant disease images often feature complex backgrounds and indistinct disease areas, increasing the difficulty of feature extraction and classification. Due to labor-intensive feature extraction, slow data accumulation, and weak generalization ability, traditional machine learning-based plant disease identification methods face significant challenges in dealing with complex disease data (Lee et al., 2020).

With the development of deep learning, Geetharamani et al. proposed a technique with higher accuracy than traditional machine learning methods, providing a new solution for crop disease identification (Geetharamani and Pandian, 2019). Zhang et al. proposed a cucumber leaf disease identification technique based on convolutional neural networks (CNN) (Zhang et al., 2019), and Tang et al. achieved good results in grape disease image classification using CNN (Tang et al., 2020). Kamal et al. widely applied popular deep learning image classification techniques like CNN to crop disease identification, significantly improving the efficiency of disease identification (Kamal et al., 2019). Even in cases with complex backgrounds and indistinct disease targets, deep learning-based models still demonstrate excellent generalization capabilities. However, CNN primarily relies on local receptive fields, making it difficult to model global information, which leads to certain limitations in feature extraction.

ViT is a deep learning-based image classification technology capable of learning high-quality intermediate features while preserving more spatial information compared to ResNet (Dosovitskiy et al., 2020). In the field of leaf disease image recognition, ViT more effectively selects areas with significant features through patch segmentation, allowing its attention mechanism to better identify disease regions and accurately determine disease types (Fu et al., 2024). Thakur et al. applied

ViT to plant leaf disease identification and demonstrated that it can effectively identify various diseases in multiple crops and, due to its lightweight structure, shows significant performance advantages in plant disease classification tasks (Thakur et al., 2023). Borhani et al. combined CNN and ViT for plant disease identification, further improving prediction speed while maintaining high accuracy (Borhani et al., 2022). However, the self-attention mechanism of ViT results in computational complexity that is quadratically related to the length of the input sequence. Specifically, for an image divided into N patches, the computational complexity of the self-attention mechanism is $O(N^2)$, meaning that when the image resolution increases, the computational load rises sharply. Saleem et al. proposed a new hybrid architecture with enhanced vegetative feature isolation, combining deep learning models to improve the performance of multi-crop disease detection, providing important reference for this study (Saleem et al., 2024). Naeem et al. used feature selection and feature concatenation methods to optimize potato leaf disease identification tasks, while our method employs state space models (SSM) for feature extraction to further reduce computational complexity (Naeem et al., 2025). ViT models require an ultra-large-scale dataset for training. When training data is insufficient, its low-level attention mechanism cannot learn local information, leading to reduced model accuracy, limiting ViT's performance on small-sample datasets (Zhao, 2017).

Swin Transformer, through a local window attention mechanism (Liu et al., 2021), associates each position with only pixels within its local window, thereby reducing the complexity of attention computation. Swin Transformer can learn and detect features at different scales, adapting to targets of different scales, thus demonstrating strong performance in tasks such as image classification, object detection, and semantic segmentation. Zhang et al. applied Swin Transformer to field rice disease identification and significantly improved the accuracy of disease detection compared to other popular models (Zhang et al., 2021). However, due to its local window computation approach, Swin Transformer may lead to the loss of information from certain diseased areas, limiting its performance in agricultural disease recognition tasks with complex environmental backgrounds. Additionally, when processing high-resolution images, its self-attention mechanism still causes a sharp increase in computational load.

In recent years, few-shot learning (FSL) has gained significant attention in agricultural disease detection tasks. PDSE-Lite extracts global features through a convolutional autoencoder (Bedi et al., 2024), improving the accuracy of disease severity estimation. However, its ability to model fine-grained lesions (e.g., vein yellowing caused by Huanglongbing) is limited. Saleem et al. proposed a multi-scale feature extraction and fusion method, optimizing few-shot learning models for wheat disease classification problems, providing inspiration for our approach (Saleem et al., 2025). TrIncNet introduced a lightweight ViT architecture optimization scheme to adapt to low-data environments, but its attention mechanism still requires $O(N^2)$ computational complexity, making it difficult to handle high-resolution field images (Gole et al., 2023). Furthermore, Hybrid CNN-Autoencoder combines CNNs and autoencoders, enhancing classification accuracy in scenarios with limited data (Bedi and Gole, 2021). Existing methods predominantly rely on large-scale pre-trained datasets such as ImageNet, while the few-shot generalization capability in agricultural disease scenarios remains insufficiently validated.

VMamba is a visual backbone network based on state space models (SSMs). Unlike traditional attention mechanisms, SSMs allow each element in the sequence to interact with previously scanned samples, reducing the computational complexity of attention from quadratic to linear (Shi et al., 2024). The core advantage of VMamba is incorporating SSMs' global receptive fields, input-dependent weighting parameters, and linear computational complexity into visual processing. In image classification tasks, VMamba outperforms popular CNN and ViT models while maintaining linear computational complexity (Liu et al., 2024). Chen et al. applied VMamba to fine-grained food visual classification tasks, combining it with residual networks, and demonstrated its superiority on high-resolution datasets, outperforming existing state-of-the-art (SOTA) models on ImageNet (Chen C. S. et al., 2024). Dang et al. used VMamba for medical image segmentation tasks, achieving significantly better performance in multiple tasks than CNN- and Transformer-based models (Dang et al., 2024).

3 Methods

3.1 State space models

State space models (SSMs) are mathematical frameworks for characterizing dynamic systems, establishing a mapping relationship between input signals and output signals by modeling the evolution of hidden states. Their core concept originates from linear time-invariant systems (LTI) and can be described using linear ordinary differential equations (ODEs), as shown in Equation 1.

$$\begin{cases} \boldsymbol{h}'(\boldsymbol{t}) = \boldsymbol{A}\boldsymbol{h}(\boldsymbol{t}) + \boldsymbol{B}\boldsymbol{u}(\boldsymbol{t}) \\ \boldsymbol{y}(\boldsymbol{t}) = \boldsymbol{C}\boldsymbol{h}(\boldsymbol{t}) + \boldsymbol{D}\boldsymbol{u}(\boldsymbol{t}) \end{cases} \tag{1}$$

The terminology definitions are as follows:

- $h(t) \in R^N$: The hidden state at time t, which carries the historical information of the system.
- $u(t) \in R$: The input signal that drives the evolution of the state.
- $y(t) \in R$: The output signal, which is jointly determined by the current state and the input.
- $A \in R^{N\times N}$: The state transition matrix, which governs the evolution of the hidden state.
- $B \in R^{N\times 1}$: The input projection matrix, which maps the input to the state space.
- $C \in R^{1\times N}$: The output projection matrix, which maps the state to the output.

$D \in R^1$: The direct transmission term (usually set to 0 to simplify the model) (Liu et al., 2024).

In deep learning, the required state transitions are discrete rather than continuous, so the above linear ordinary differential equations must be converted into discrete functions. Considering an input as a d-dimensional signal stream of length L, this paper employs the Zero-Order Hold (ZOH) method to discretize the aforementioned differential equation. It assumes that the input remains constant over the time interval Δ and solves the ODE through integration, yielding the following discrete recurrence relation, as shown in Equation 2:

$$\boldsymbol{h_k} = \bar{\boldsymbol{A}}\boldsymbol{h_{k-1}} + \bar{\boldsymbol{B}}\boldsymbol{u_k} \tag{2}$$

where the discretization parameters are defined as:

$$\bar{\boldsymbol{A}} = \boldsymbol{e}^{\boldsymbol{A\Delta}}, \ \bar{\boldsymbol{B}} = \left(\int_0^{\Delta} \boldsymbol{e}^{\boldsymbol{A(\Delta-\tau)}}\boldsymbol{d\tau}\right)\boldsymbol{B} \approx \boldsymbol{\Delta B}$$

The Zero-Order Hold (ZOH) method is a standard technique used to approximate continuous-time systems in discrete form. Its purpose is to ensure that the behavior of the discretized system closely resembles that of the original continuous-time system. Specifically, ZOH assumes that the input signal remains constant within the sampling interval, thereby simplifying the computation of hidden state updates. The advantage of this approach lies in its ability to accurately capture the dynamics of the input signal while maintaining linear computational complexity, ensuring numerical stability for high-resolution image inputs.

3.2 Selective scanning mechanism

The static parameters A, B, and C in traditional State Space Models (SSMs) limit their adaptability to context. To address this, the Selective Scanning Mechanism (S6) introduces input-dependent dynamic parameters: B_k , C_k , $\Delta_k = f_\theta(u_k)$. The selective scanning mechanism (called S6) allows the model to selectively process information to focus on or ignore specific inputs. S6 transforms the dimensions of the matrices affecting the input B and the state-influencing C to (B, L, N), where these parameters correspond to batch size, sequence length, and hidden state size, respectively, and changes the size of Δ to (B, L, D). This means that during inference, the model can dynamically calculate the values of matrices B_k, C_k, and step length Δ_k based on different input data, thus achieving the goal of selectively retaining content in the hidden state and ignoring certain content (Gu and Dao, 2023). In S6, for the input sequence u_1 , u_1,…, u_T, the hidden state is updated as follows, as shown in Equation 3:

$$\mathbf{h}_k = \mathbf{w}_k \odot \mathbf{h}_{k-1} + \sum_{i=1}^{k}\left(\prod_{j=i}^{k} e^{-\mathbf{A}_j\Delta_j}\right)\mathbf{B}_i u_i \tag{3}$$

Here, $\mathbf{w}_k = e^{\mathbf{A}_k\Delta_k}$ represents the dynamic decay weight, and ⊙ denotes element-wise multiplication.

For an input feature map $X \in R^{H\times W\times C}$, the output is given by Equation 4:

$$\mathbf{Y} = \sum_{d\in\{Four\ directions\}} \mathrm{Reshape}(\mathrm{S6}_d(\mathrm{Flatten}_d(\mathbf{X}))) \tag{4}$$

In the VMamba architecture, the Selective Scanning Mechanism is applied within the SS2D module. To account for the two-dimensional nature of visual data, a Cross Scan Module (CSM) is used to generate sequential inputs. The input sequence is expanded along multiple scanning paths, extending them into sequences along rows and columns (cross-scanning). Subsequently, the sequences are scanned in four different directions. Cross-Scan: This process divides the input image data into multiple patches and unfolds them along four distinct scanning paths (horizontal, vertical, and two diagonal directions). This allows each patch to interact with others in different orders, thereby collecting contextual information from various directions. Parallel Processing: Parallel processing employs S6 Blocks, which are modules based on State Space Models (SSMs). In each scanning path, the S6 module is responsible for selectively weighting and processing each image patch. The S6 module dynamically adjusts the weight of each patch through an input-dependent selection mechanism, enabling the model to focus on important regions. Cross-Merge: After the four scanning paths complete their respective processing, the Cross-Merge block combines the results from the four scanning paths into a final two-dimensional feature map. This process integrates information from the four different scanning directions, ensuring that each patch receives contextual information from all directions. This is illustrated in Figure 1.

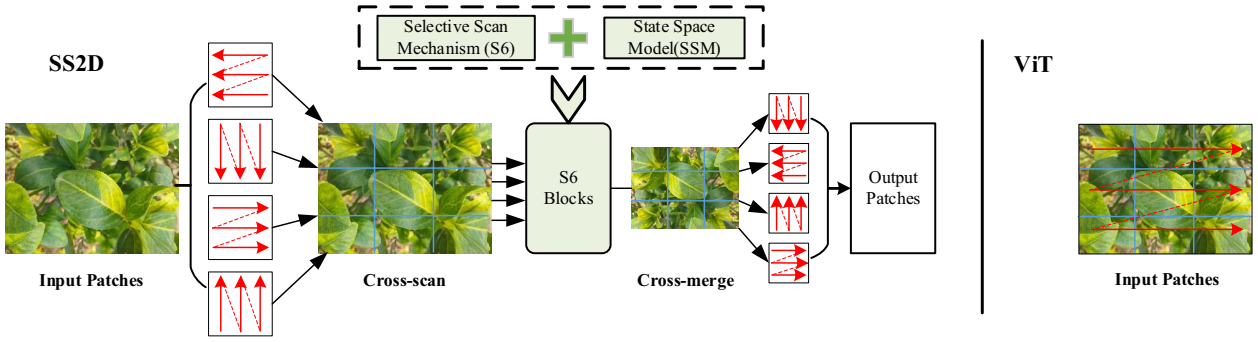


FIGURE 1
SS2D architecture.

Transformer-based models like ViT divide the input image into smaller patches, flatten these patches, and feed them as sequences into the Transformer model (Dosovitskiy et al., 2020), as shown in Figure 1. This approach inevitably leads to a limited receptive field. Moreover, in the self-attention mechanism, each input element interacts computationally with all other elements, resulting in a quadratic computational complexity of $O(N^2)$, where N is the length of the input sequence. Specifically, ViT's self-attention mechanism computes the similarity between every pair of input elements, leading to high computational and memory costs when handling high-resolution images.VMamba adopts the Selective Scan mechanism, allowing each input element to interact only with its neighboring elements, thereby reducing the computational complexity to linear. This approach not only enhances computational efficiency but also maintains a global receptive field, ensuring the model's effectiveness in handling visual tasks.

3.3 VMamba architecture

The VMamba architecture is shown in Figure 2, where all these stages together build a hierarchical representation similar to CNN and ViT models (Liu et al., 2024). The stages are described as follows: Patch Partition: Input images are divided into multiple patches using a Patch Partition similar to ViT (Dosovitskiy et al., 2020), obtaining a feature map of size. **Stage 1**: Several VSS blocks are stacked on the feature map, maintaining the same dimensions. **Stage 2**: The feature map from Stage 1 is downsampled, and more VSS Blocks are stacked to produce an output resolution of. **Stage 3 & Stage 4**: Repeat the process of Stage 2 to create feature maps with resolutions of and, respectively.

The VSS Block is the core of the VMamba architecture. In the VSS Block, the input is passed through Layer Normalization (LN), and the output is split into two streams of information. One stream passes through a 3x3 Depthwise Convolution (DWConv) layer (Chollet, 2017), followed by SiLU activation and entering the core SS2D module. The SS2D output is then passed through a normalization layer and added to the output of the other stream.

SS2D is the key element of the VSS Block, integrating the S6 Block with the Cross Scan Module (CSM), as shown in Figure 1. To address the limitations of LTI SSMs in capturing contextual information, the S6 Block combines S6 with SSMs, allowing each element in a 1D array to interact with any previously scanned sample by compressing the hidden state, effectively reducing quadratic complexity to linear (Liu et al., 2024). VMamba incorporates the selective scanning mechanism (S6) as the core SSMs operator, processing input data causally, thereby capturing information only within the scanned portion of the data. Although the order of operations in the S6 Block aligns with NLP tasks involving temporal data, it faces significant challenges when applied to non-temporal data (e.g., images, graphs, sets) because visual sequences are inherently non-sequential (Liu et al., 2024). To solve this problem, Chen et al. proposed the Cross Scan Module (CSM). CSM selects image patches, expands them into sequences along rows and columns (cross-scan), and then scans in four different directions: top-left to bottom-right, bottom-right to top-left, top-right to bottom-left, and bottom-left to top-right (Chen C. S. et al., 2024). This allows any pixel to integrate information from all other pixels in different directions. Each sequence is independently processed by different S6 blocks, then reshaped into a single image, and all sequences are merged into a new sequence (cross-merge).

In the SS2D Block, input blocks are traversed along four different scanning paths (cross-scan), each sequence is independently processed by different S6 blocks, and the results are subsequently fused to construct a 2D feature map (cross-merge) as the final output. Data forwarding in the SS2D Block is divided into three steps: cross-scan, S6 block selective scanning, and cross-merge. The SS2D Block allows visual data to use 1D selective scanning, processes image blocks in parallel through S6 blocks,

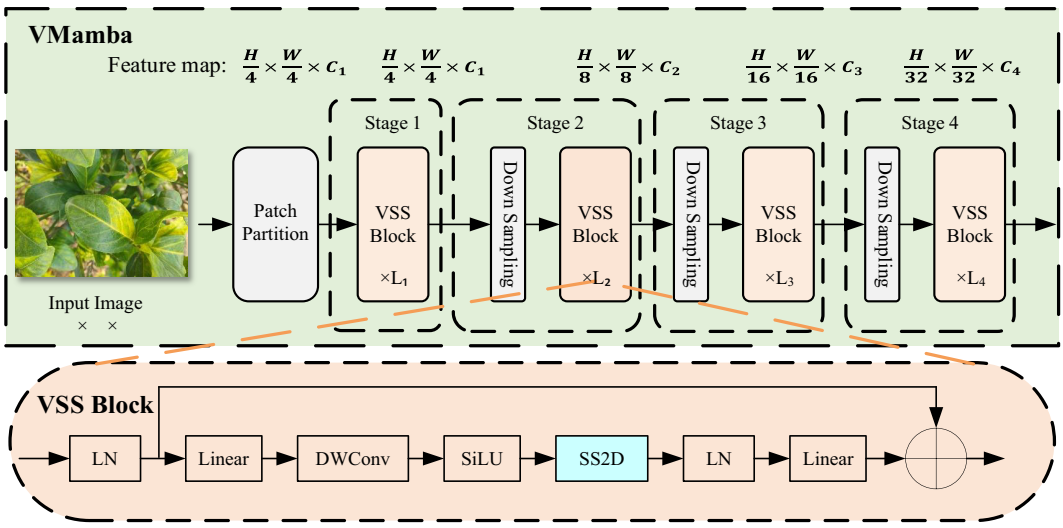


FIGURE 2
VMamba architecture.

and merges the results to form a 2D feature map, thereby effectively utilizing the S6 module. It inherits the linear complexity of the selective scanning mechanism while retaining a global receptive field (Liu et al., 2024).

3.4 Denoising diffusion probabilistic models

In deep learning image classification tasks, data augmentation techniques are widely used to address the issue of insufficient training data. By applying transformations such as rotation, translation, scaling, cropping, flipping, and adding Gaussian or salt-and-pepper noise to the original data, the diversity of the data is increased, thereby enhancing the model's generalization ability and robustness. However, traditional data augmentation methods have certain limitations. These methods typically rely on manually designed rules, which may fail to fully capture the complex features of the data, resulting in limited diversity of the generated samples and an inability to cover the full scope of the data distribution. Moreover, excessive reliance on data augmentation can lead to overfitting of the model to specific transformations, affecting its performance on unseen samples. To overcome these limitations, DDPM offers a new approach to image augmentation (Ho et al., 2020).

Denoising Diffusion Probabilistic Models (DDPM) are latent variable models inspired by non-equilibrium thermodynamics. The structure of DDPM is shown in Figure 3. DDPM defines a Markov chain for the diffusion steps, gradually adding random noise to the data, and then learns the reverse diffusion process to reconstruct the desired data sample from noise, thus generating target data samples from noise (Ho et al., 2020). The diffusion model consists of two main processes: the forward process (diffusion process) and the reverse process. By training a DDPM diffusion model to learn the diffusion of image data, we input randomly generated noisy images into the DDPM diffusion model and perform the reverse diffusion process to ultimately obtain synthesized images similar to real images. A visual comparison between images generated by DDPM and traditional image generation methods is shown in Figure 4.

Unlike other generative models, such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), diffusion models gradually add noise to images during the forward process until the image is completely destroyed, becoming Gaussian noise. During the reverse process, the model learns how to recover the original image from Gaussian noise (Li et al., 2023). This stepwise noise addition and denoising process make diffusion models exhibit superior stability and diversity when generating samples, making them an important focus of recent generative model research.

3.5 Transfer learning

Transfer learning (Zhuang et al., 2020) is a machine learning technique that applies knowledge learned from one domain (source domain) to another related domain (target domain) to improve model performance on a new task. This method is especially suitable for situations with limited data, as it leverages existing knowledge to reduce the amount of training data required for new tasks. Transfer learning offers the advantage of saving time and computational resources, as it avoids the need to train a model from scratch, particularly for complex tasks. Moreover, pretrained models provide better initial weights, accelerating convergence and improving accuracy (Yao et al., 2020). The main steps of transfer learning are as follows: (1) Pretraining: A deep learning model is trained on a large dataset (e.g., ImageNet) to learn general features. For instance, convolutional neural networks (CNNs) can extract edges, textures, and other low-level features from images. (2) Feature Transfer: The weights and structure of the pretrained model are transferred to the target task. The model is fine-tuned on a small amount of data to adjust its parameters. (3) Evaluation and Optimization: The transferred model is evaluated using data from the target domain and further optimized as needed.

3.6 Integration of DDPM, transfer learning, and VMamba for disease identification

To evaluate the performance and characteristics of the VMamba model in recognizing plant leaf diseases, we designed five methods that integrate DDPM, transfer learning, and VMamba.

Method 1: A Plant Disease Recognition Method Using VMamba (Large Image Dataset). This method applies the VMamba model to a large dataset of plant leaf diseases to evaluate its performance on a big dataset.

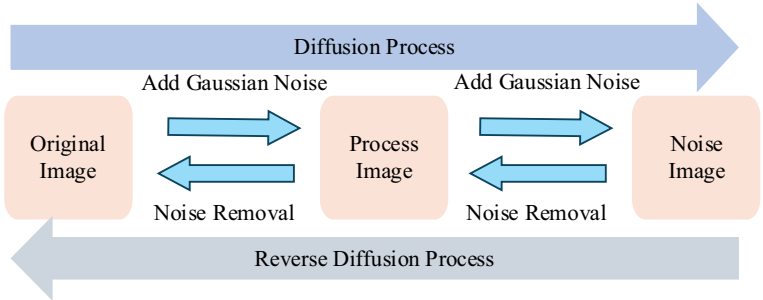


FIGURE 3
Diffusion process.

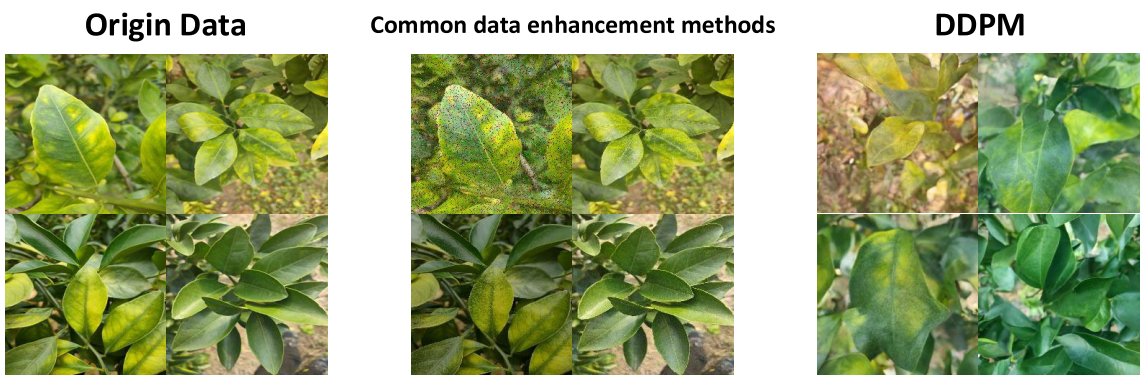


FIGURE 4
Comparison of data enhancement methods.

Method 2: A Plant Disease Recognition Method Using VMamba (Small Image Dataset). This method uses the VMamba model on a small sample dataset of plant leaf diseases, testing its performance with limited data.

Method 3: A Transfer Learning-Based Disease Recognition Method with VMamba. As shown in Figure 5, this method first pre-trains the VMamba model on a large-scale plant disease dataset, generating a pre-trained model. Transfer learning is then used to transfer the learned network architecture and weights to the small sample dataset for recognizing diseases in plant leaf images with limited data.

Method 4: Diffusion-Driven Transfer Learning with VMamba (Small Image Dataset). As shown in Figure 6, this method combines DDPM and transfer learning for VMamba on small datasets. DDPM is used to augment the small plant disease dataset by generating additional images, thereby increasing the training data and enhancing the model's generalization. The model trained on the augmented images is used as a pre-trained model. Transfer learning (Zhuang et al., 2020) is then applied to fine-tune this pre-trained model on the original small sample dataset, aiming to improve the robustness of the VMamba model, reduce sensitivity to image variations, and ultimately improve accuracy.

Method 5: Diffusion-Driven Hybrid Transfer Learning with VMamba (DDHTL-VMamba). As shown in Figure 7, this method consists of two stages:

1. **Pre-training Stage**: In this stage, DDPM (Denoising Diffusion Probabilistic Models) is used to generate an augmented image dataset based on a small-sample dataset (In-field small dataset). This augmented dataset is then combined with a larger dataset (PlantVillage dataset) for pre-training the model. The goal of this stage is to leverage the generative capabilities of DDPM to enrich the diversity of the training data, thereby improving the model's ability to learn robust features.
2. **Transfer Learning Stage**: After pre-training, the model is fine-tuned on a small-sample plant disease dataset using transfer learning techniques. This involves adapting the pre-trained model to the specific characteristics of the target dataset, which typically contains limited labeled data. By doing so, the model can better capture the nuances of the target task while retaining the generalization benefits gained during pre-training.

The purpose of this approach is to enhance the accuracy and performance of the VMamba model in identifying plant diseases from small-sample datasets, as well as to improve its generalization capability in real-world application scenarios. This hybrid strategy effectively addresses the challenges posed by limited data availability, ensuring that the model performs reliably even when trained on smaller, domain-specific datasets.

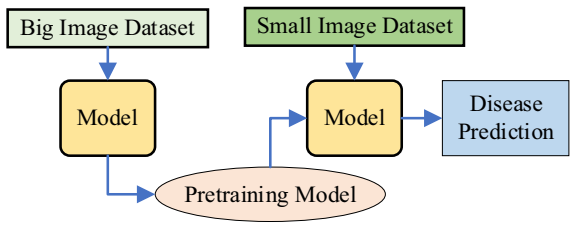


FIGURE 5
Transfer learning-based plant disease recognition method using VMamba.

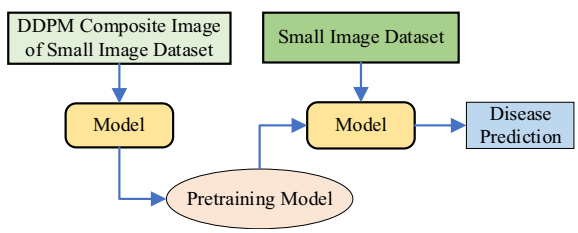


FIGURE 6
The workflow of method 4.

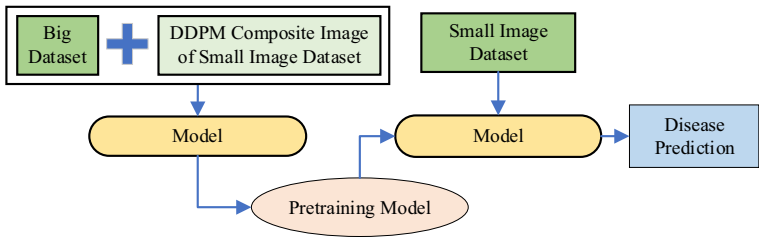


FIGURE 7
The Workflow of Method 5.

4 Experiment

4.1 Experimental datasets and environment

In this study, we utilized the PlantVillage dataset as the large dataset and the In-field small dataset constructed by our team (Li et al., 2023; Chen et al., 2024a; Chen et al., 2024b) as the small sample dataset.

PlantVillage dataset consists of 54,303 images of healthy and diseased plant leaves, covering 38 different species and diseases. This includes crops such as tomatoes, apples, bananas, and their related diseases. This dataset provides a diverse set of images, creating an ideal environment for evaluating the performance of models in plant disease recognition tasks.

In-field small dataset is a small sample dataset focused on citrus plant diseases, collected from real-world field environments. It contains a total of 3,250 high-resolution (4000×3000) color images, divided into three categories: HLB-infected leaves (758 images), magnesium-deficient leaves (739 images), and healthy leaves (1,151 images). This dataset is particularly useful for testing the model's performance in small-sample plant disease recognition tasks.

4.2 Performance evaluation metrics

The performance of the models was evaluated using several key metrics, including Params, Gflops, Time per epoch, AccTop1, AccTop5, Precision, Recall, and F1 Score, as detailed in Table 1. Additionally, Confusion Matrix analysis was employed in some experiments to further assess the classification performance of the models, ensuring accuracy in the results, as shown in Figure 8. The Confusion Matrix is a tool for evaluating classification model performance by visualizing the relationship between predicted and true labels. It helps analyze how well the model classifies different categories. Typically presented as a square matrix, each row represents the true class, and each column represents the predicted class, as shown in Figure 8. In this matrix, True Positive (TP) refers to the correctly predicted positive samples, True Negative (TN) refers to correctly predicted negative samples, False Positive (FP) represents the negative samples incorrectly predicted as positive (false alarms), and False Negative (FN) represents the positive samples incorrectly predicted as negative (missed detections).

TABLE 1 Evaluation metrics.

Evaluation Metrics	Description
Params	The number of parameters in the model, typically used to assess the model's complexity and capacity. A higher number of parameters generally indicates greater model complexity, but it may also lead to overfitting.
Gflops	Gflops (Giga Floating Point Operations Per Second), measures the computational complexity and processing speed of the model, reflecting its computational efficiency.
Time per epoch	The time required for each training epoch, often used to evaluate the efficiency of the training process. Shorter times indicate that the model can iterate through training more quickly.
Acc-Top1	The accuracy of the model in correctly identifying the top predicted class in classification tasks. Higher values indicate better model performance.
Acc-Top5	The accuracy of the model in correctly identifying one of the top five predicted classes in classification tasks, reflecting the model's robustness in multi-class recognition.
Precision	Precision measures the proportion of correctly predicted positive samples out of all samples predicted as positive. A high precision indicates the model's strong accuracy in predicting the positive class.
Recall	Recall measures the proportion of actual positive samples that were correctly predicted as positive. A high recall indicates the model's effectiveness in identifying positive samples.
F1 Score	The F1 Score is the harmonic mean of precision and recall, providing a balanced evaluation of the model's performance, especially in cases of class imbalance. A higher F1 score indicates better overall performance.

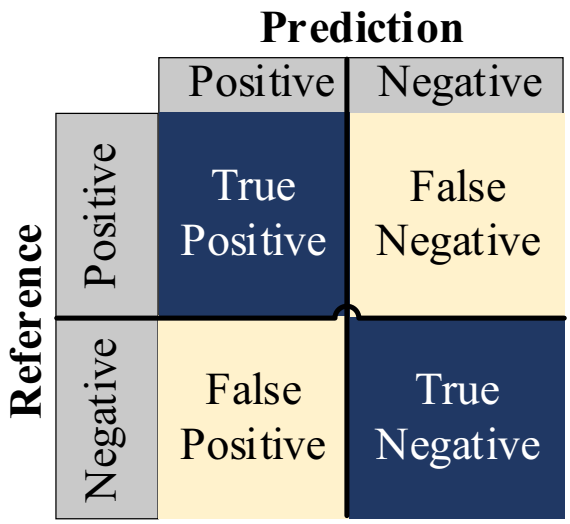


FIGURE 8
Confusion matrix.

4.3 Experiments

The experiments conducted in this study are summarized in Table 2. To ensure data consistency and facilitate model processing, all images were resized to a uniform resolution of 224×224 pixels during the data preprocessing stage of model training. Additionally, random cropping was applied to the input images within a specified range to augment the dataset and improve model robustness.The dataset was split into training, validation, and test sets in a 6:2:2 ratio, ensuring that the training and validation conditions remained consistent across different datasets. All computations were performed on the same hardware and operating environment to ensure controlled variables. The setup includes: CPU: R9 7940H, RAM: 16 GB, GPU: RTX 4060 Max-Q 8 GB, SYSTEM: Ubuntu 22.04, Python: 3.10, PyTorch: 2.0.0, and CUDA: 11.8. In terms of hyperparameter settings, we used a batch size of 16 for the VMamba network. A linear learning rate scaling strategy was employed for scheduling, and the AdamW optimizer was utilized for training. The number of training epochs was uniformly set to 300, and an Early-Stop mechanism was implemented to prevent overfitting and optimize training efficiency. This systematic approach ensures that the experimental results are reliable, reproducible, and reflective of the model's true performance.

4.3.1 Results and analysis of experiment 1

The results of Experiment 1 are shown in Table 3. The findings demonstrate that the VMamba model outperformed other models across various metrics, including Acc-Top1, Acc-Top5, Precision, Recall, and F1 Score. Notably, VMamba achieved an Acc-Top1 accuracy of 99.81%, with the Swin Transformer coming in second at 99.18%. Furthermore, VMamba exhibited significant advantages in training speed, especially when compared to the ViT model, with much shorter training times per epoch.

The results show that although the ViT model has the highest parameter count (85.68M) and computational complexity (16.86 Gflops), its performance did not significantly surpass that of VMamba. Particularly in terms of Acc-Top1 and Acc-Top5 accuracy, ViT lagged behind VMamba, possibly due to its reliance on large datasets to maintain performance in large-scale tasks. The VMamba model, with its efficient selective scanning mechanism (S6), achieved higher accuracy and precision while maintaining lower computational complexity (4.85 Gflops). While the ResNet50 model has fewer parameters (23.59M), its performance was inferior to VMamba, especially in Precision and Recall. The DeiT model showed good results in certain metrics, particularly in terms of faster training times per epoch, but its overall predictive performance was still behind that of VMamba and Swin Transformer.

Conclusion of Experiment: The VMamba model demonstrated outstanding performance in large-scale plant disease recognition tasks, surpassing other models in terms of accuracy and computational efficiency. This indicates that the VMamba model holds great potential for applications in agricultural disease identification, especially in scenarios with limited computational resources.

4.3.2 Results and analysis of experiment 2

The results of Experiment 2 are shown in Table 4. This experiment primarily evaluates the performance of VMamba on small-sample datasets to verify its generalization capabilities in low-data environments. A key challenge in agricultural disease detection is the difficulty of collecting large-scale, high-quality annotated data in real-world applications. Therefore, we used the In-field small

TABLE 2 Experiments.

Experiment	Objective	Method	Dataset
Experiment 1	Evaluating VMamba's Disease Identification Ability	Method 1	Big Image Dataset PlantVillage
Experiment 2	Evaluating VMamba's Disease Identification Ability on Small Image Dataset	Method 2	Small Image Dataset In-field small dataset
Experiment 3	Evaluating the Performance of VMamba Integrated with Transfer Learning	Method 3	same as above
Experiment 4	Evaluating the Performance of VMamba Integrated with DDPM and Transfer Learning	Method 4	Big Image Dataset PlantVillage; Small Image Dataset In-field small dataset
Experiment 5	Evaluating the Performance of DDHTL-VMamba	Method 5	same as above

TABLE 3 Experiment 1 results.

Model	VMamba	ResNet50	DeiT	ViT	Swin Transformer
Params	30.7M	23.59M	5.71M	85.68M	27.53M
Gflops	4.85	4.13	1.08	16.86	4.37
Acc-Top1 (%)	99.81	97.5	99.10	97.28	99.18
Acc-Top5 (%)	99.99	99.91	99.97	99.91	99.98
Precision (%)	99.51	96.06	98.65	96.61	98.88
Recall (%)	99.20	96.50	98.58	96.71	98.76
F1 Score (%)	99.34	96.21	98.60	97.29	98.81
Time per epoch	287s	237s	57s	655s	336s

dataset as the experimental object. This dataset contains 3,250 images, divided into three categories: HLB-infected leaves (758 images), magnesium-deficient leaves (739 images), and healthy leaves (1,151 images).This small-sample dataset reflects the practical challenges of data acquisition in real agricultural environments. The images vary in lighting conditions, background complexity, and leaf angles, simulating the complexities of disease detection tasks in small-sample scenarios.

The findings indicate that VMamba performed exceptionally well in the small-sample plant disease identification task, surpassing models such as ResNet50, ViT, and DeiT in key performance metrics like Acc-Top1, Precision, Recall, and F1 Score. Additionally, VMamba demonstrated superior efficiency in training time. VMamba achieved an Acc-Top1 accuracy of 93.21%, while the Swin Transformer model slightly outperformed VMamba with 93.37%. Although Swin Transformer had a slight advantage in Acc-Top1, its training time per epoch (228 seconds) was 286% longer than that of VMamba (59 seconds), indicating that VMamba has a significant advantage in computational efficiency. To comprehensively evaluate model performance, we conducted statistical significance tests and calculated confidence intervals for metrics such as accuracy, precision, recall, and F1 score.

Compared to ViT, VMamba's performance was particularly strong. ViT's performance dropped significantly on the small-sample dataset, achieving an Acc-Top1 of only 71.59%, much lower than its results on the large-scale PlantVillage dataset. This shows that ViT struggles with small-sample data, whereas VMamba's efficient architecture helps mitigate the performance degradation associated with small-sample datasets. ResNet50 showed results close to VMamba's but fell short in both precision and computational efficiency. VMamba's training time per epoch was 59 seconds, while ResNet50 required 241 seconds, indicating that VMamba is more computationally efficient. Swin Transformer performed well on the small-sample dataset, especially in Acc-Top1 and Precision, slightly outperforming VMamba. This could be attributed to its Window-based Multi-Head Self-Attention (W-MSA) mechanism, which helps extract local features effectively on small datasets (Liu et al., 2021). However, Swin Transformer's training time per epoch was 228 seconds, which was 286% longer than VMamba's. Overall, VMamba strikes a better balance between efficiency and accuracy.

Conclusion of Experiment 2: VMamba outperformed ResNet50, ViT, and DeiT on small-sample agricultural disease datasets in terms of both accuracy and computational efficiency. It also performed comparably or better than Swin Transformer on metrics like Acc-Top1 and Time per epoch. VMamba's efficient selective scanning mechanism and state-space model architecture provide strong adaptability and robustness, maintaining high classification performance while keeping computational complexity low.

4.3.3 Results and analysis of experiment 3

The results of Experiment 3 are shown in Table 5. Transfer learning based on the PlantVillage dataset significantly improved

TABLE 4 Results of Experiment 2.

Model	VMamba	ResNet50	DeiT	ViT	Swin Transformer
Params	30.7M	25.56M	5.71M	85.80M	27.53M
Gflops	4.85	4.13	1.08	68.72	4.37
Acc-Top1 (%)	93.21	91.84	82.26	71.59	93.37
Precision (%)	91.74	91.12	82.57	71.68	92.81
Recall (%)	90.49	89.86	80.75	70.16	92.70
F1 Score (%)	90.94	90.26	81.41	70.74	92.69
Time per epoch	59s	241s	34s	94s	228s

TABLE 5 Results of Experiment 3.

Model	PV-In-field-VMamba	PV-In-field-ViT	PV-In-field-ResNet50	PV-In-field-Swin Transformer
Params	30.7M	85.68M	23.59M	27.53M
Gflops	4.86	16.86	4.13	4.37
Acc-Top1 (%)	99.81	93.06	98.26	96.32
Precision (%)	99.78	91.76	97.79	95.70
Recall (%)	99.78	92.69	98.00	96.17
F1 Score (%)	99.78	92.09	97.89	95.91
Time per epoch	47s	275s	254s	235s

VMamba's performance in the small-sample plant disease identification task. In terms of Acc-Top1, Precision, Recall, and F1 Score, the PV-In-field-VMamba model outperformed PV-In-field-ViT, PV-In-field-ResNet50, and PV-In-field-Swin Transformer. Specifically, VMamba achieved an Acc-Top1 accuracy of 99.81%, significantly higher than ViT's 93.06% and ResNet50's 98.26%. In addition to the accuracy advantage, VMamba was also much more efficient. Due to its large parameter size (85.68M) and high computational complexity (16.86 Gflops), ViT took 275 seconds per epoch, while VMamba required only 47 seconds per epoch, with a lower complexity of 4.86 Gflops. Moreover, VMamba showed more stable overall performance across Precision, Recall, and F1 Score after transfer learning. While Swin Transformer performed closely to VMamba in

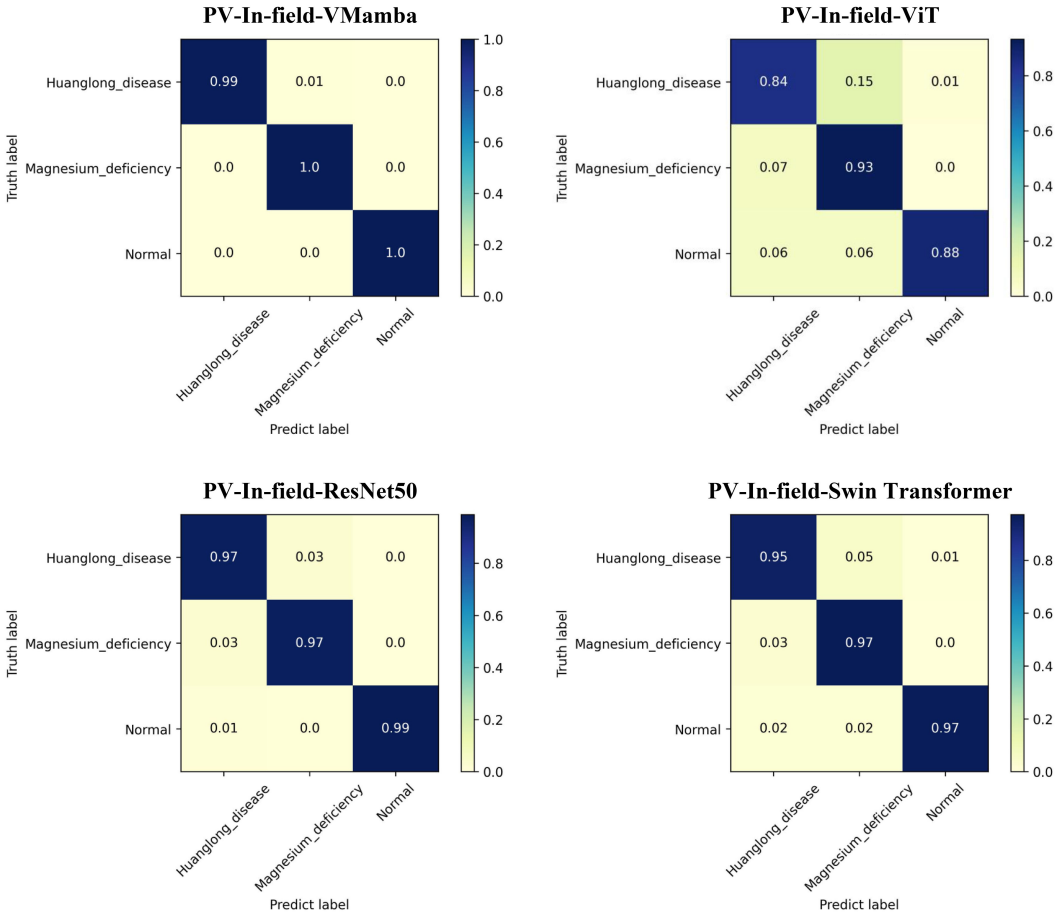


FIGURE 9
Confusion matrix: Experiment 3.

some metrics, it lagged behind in terms of accuracy, especially when handling small-sample disease identification tasks where VMamba's advantages were more pronounced.

The Confusion Matrix for each model is shown in Figure 9, where PV-In-field-VMamba performed best across all categories. In particular, VMamba showed the highest accuracy in identifying Huanglong disease, while ViT and ResNet50 struggled with misclassification and confusion in recognizing Huanglong disease and Magnesium deficiency. VMamba effectively avoided these errors, indicating stronger distinction and robustness in handling complex agricultural disease categories.

Conclusion of Experiment 3: The transfer learning performance of the VMamba model on small-sample agricultural disease datasets was superior to ViT, ResNet50, and Swin Transformer. VMamba not only excelled in accuracy but also exhibited exceptional computational efficiency, particularly in small-sample disease identification tasks. Through transfer learning, VMamba was able to retain useful features from the pre-trained model while adapting to different datasets with outstanding results.

4.3.4 Results and analysis of experiment 4

The results of Experiment 4 are shown in Table 6. The use of the DDPM data augmentation method significantly expanded the small-sample dataset and improved the performance of all models. Specifically, the DDPM-In-field-VMamba model achieved the best results in terms of Acc-Top1, Precision, Recall, and F1 Score, with an Acc-Top1 of 97.93%, far surpassing other models, particularly Swin Transformer (92.83%). In terms of precision and recall, VMamba remained the top performer even after DDPM data augmentation. Although ViT and ResNet50 showed some improvement, they were still unable to outperform VMamba, indicating that VMamba's architecture continues to excel in handling augmented datasets with outstanding generalization and stability. Additionally, the experiments revealed that DDPM-augmented datasets not only improved model accuracy but also shortened training times per epoch. VMamba took only 49 seconds per epoch, compared to ViT's 276 seconds and ResNet50's 253 seconds, achieving a good balance between performance and training efficiency.

The Confusion Matrix for each model is shown in Figure 10, where DDPM-In-field-VMamba outperformed other models in all categories, particularly in Huanglong disease and Magnesium deficiency recognition. ViT and ResNet50 were more prone to misclassification in these categories, showing less stability when handling complex categories compared to VMamba. The data results were subjected to statistical significance tests, indicating that the Acc-Top1 metric of VMamba was significantly higher than that of other models (a p-value less than 0.05 indicates a significant difference).

Conclusion of Experiment 4: The DDPM data augmentation method significantly improved the model's performance when working with small-sample agricultural disease datasets. Particularly in the VMamba model, data augmentation after transfer learning led to the best results across metrics such as Acc-Top1, Precision, Recall, and F1 Score, while also greatly improving training efficiency. Compared to other models, VMamba not only achieved higher accuracy but also significantly reduced training time. The comparison of confusion matrices further shows that VMamba exhibited superior accuracy and robustness across different disease categories.

4.3.5 Results and analysis of experiment 5

The results of Experiment 5 are presented in Table 7. When using the DDHTL-VMamba training method, the model's accuracy showed a slight decrease compared to the PV pre-trained transfer learning method, but the time required for the pre-training phase was significantly reduced by 43.88%, which greatly lowered the training cost. Compared to the transfer learning method using DDPM-augmented datasets, the DDHTL-VMamba method showed improvements in model accuracy. The Confusion Matrix is shown in Figure 11, where we can observe a clear improvement in identifying Huanglong disease compared to the DDPM-augmented pre-training method. The results demonstrate that the DDHTL-VMamba method can more effectively balance model accuracy and training cost.

5 Discussion

Through a series of experiments using the VMamba model on different datasets and training methods, this study found that VMamba exhibits outstanding performance in agricultural disease

TABLE 6 Results of Experiment 4.

Model	DDPM-In-field-VMamba	DDPM-In-field-ViT	DDPM-In-field-ResNet50	DDPM-In-field-Swin Transformer
Params	30.7M	85.68M	23.59M	27.53M
Gflops	4.86	16.86	4.13	4.37
Acc-Top1 (%)	97.93	96.00	96.53	92.83
Precision (%)	97.80	95.14	95.73	92.02
Recall (%)	97.65	95.32	95.92	92.16
F1 Score (%)	97.70	95.23	95.82	92.08
Time per epoch	49s	276s	253s	250s

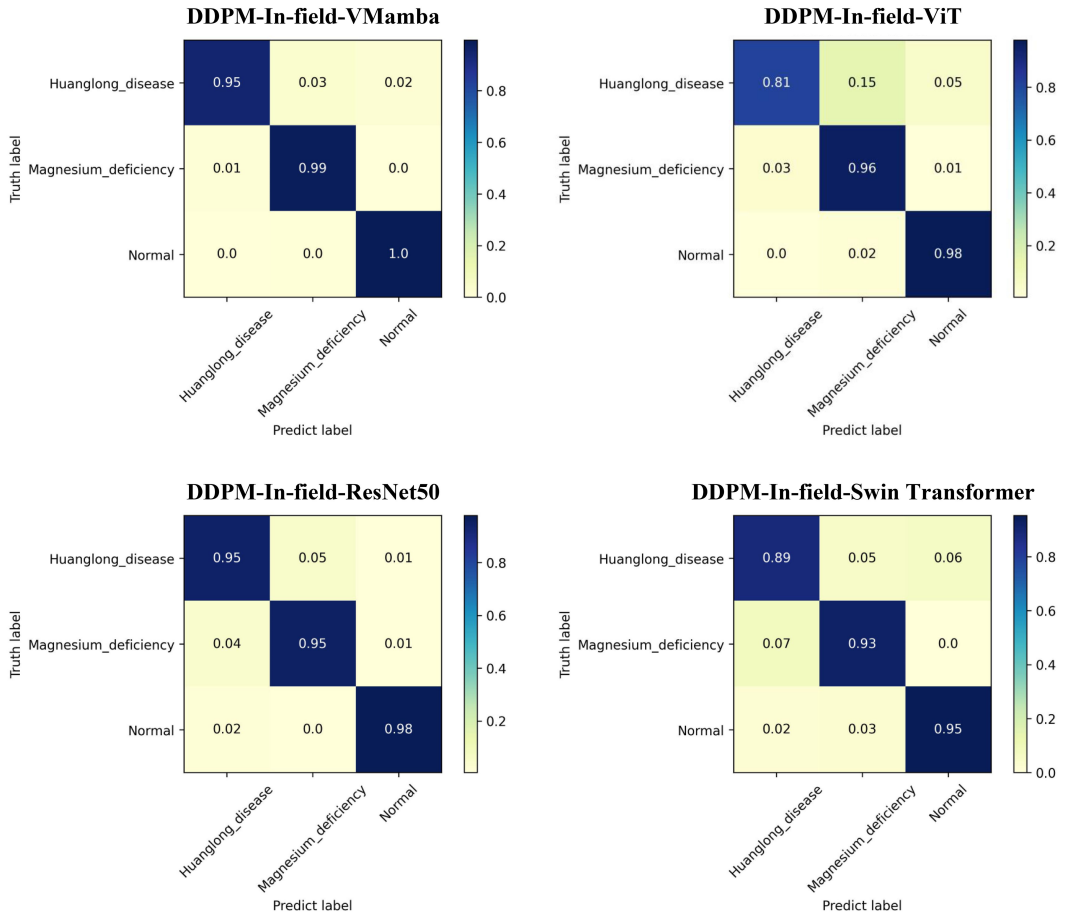


FIGURE 10
Confusion matrix: Experiment 4.

identification, especially on small-sample datasets. Compared to other popular vision models such as ResNet50, ViT, and Swin Transformer, VMamba consistently outperformed them on several key metrics (Acc-Top1, Acc-Top5, Precision, Recall, F1 Score). This was particularly evident in small-sample conditions and when transfer learning and DDPM data augmentation were applied, where VMamba's performance further stood out.

In Experiment 1, the VMamba model demonstrated superior performance on the PlantVillage dataset. Compared to models like ResNet50, ViT and Swin Transformer, VMamba's architecture, based on state-space modeling, was better at handling complex visual features. In Experiment 2, VMamba continued to outperform ResNet50, DeiT, and ViT on the In-field small dataset, showcasing its effectiveness in small-sample data scenarios. This advantage can

TABLE 7 Results of Experiment 5.

Model	DDHTL-VMamba	PV-In-field-VMamba	DDPM-In-field-VMamba	Common data enhancement-VMamba
Params	30.7M	30.7M	30.7M	30.7M
Gflops	4.86	4.86	4.86	4.86
Acc-Top1 (%)	99.434	99.811	97.925	92.830
Precision (%)	99.338	99.776	97.797	92.554
Recall (%)	99.342	99.781	97.650	91.819
F1 Score (%)	99.333	99.778	97.704	92.109
Pretrained time	14,496s	25,830s	12,200s	—

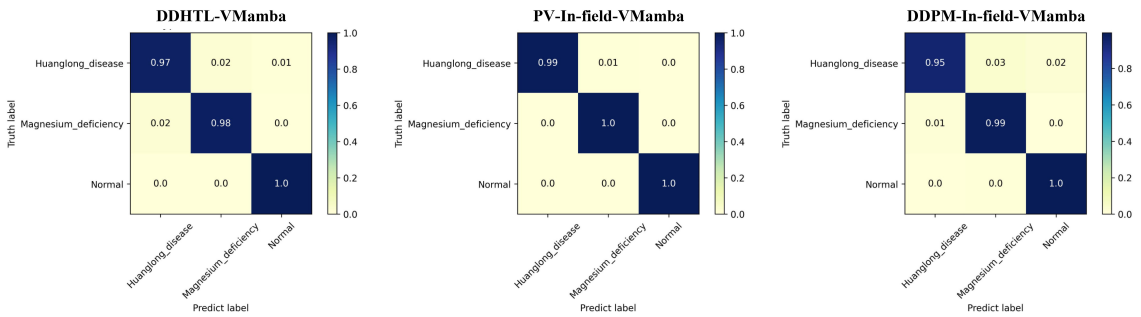


FIGURE 11
Confusion matrix of Experiment 5.

be attributed to VMamba's selective scanning mechanism (S6), which maintains a global receptive field while reducing computational complexity, thereby enhancing the model's generalization ability. Although ResNet50 performs well on large-scale datasets, it struggles with generalization on small-sample datasets. Similarly, the ViT model, due to its high parameter count and reliance on large datasets, performed poorly when data was limited. In Experiment 3, the introduction of transfer learning significantly improved VMamba's performance on small-sample datasets. After pre-training on the PlantVillage dataset, the model could learn general features that maintained high recognition accuracy when transferred to small-sample datasets. This finding aligns with existing literature, which indicates that transfer learning can greatly enhance model performance on small datasets (Zhao, 2017; Zhao et al., 2020). In Experiment 4, the DDPM diffusion model was used to generate additional images, further improving the model's recognition capability. The augmented dataset provided richer features for the small-sample data, increasing the model's robustness. The results showed that VMamba performed the best across multiple metrics after DDPM augmentation, confirming the effectiveness of data augmentation for small-sample challenges. In Experiment 5, the DDHTL-VMamba training method was proposed, which pre-trains on a combination of DDPM-augmented and original datasets and then transfers to the In-field small dataset. The results indicate that this method effectively balances model accuracy with reduced pre-training time, significantly lowering training costs. Compared to traditional data augmentation methods (which include transformations such as rotation, scaling, cropping, flipping, and noise addition like Gaussian and salt-and-pepper noise), both DDPM and DDHTL demonstrated superior performance. This shows that the DDHTL-VMamba training method is ideal for agricultural disease identification tasks where balancing training cost and model performance is critical.

The VMamba model addresses the limitations of traditional The VMamba model addresses the limitations of traditional ViT

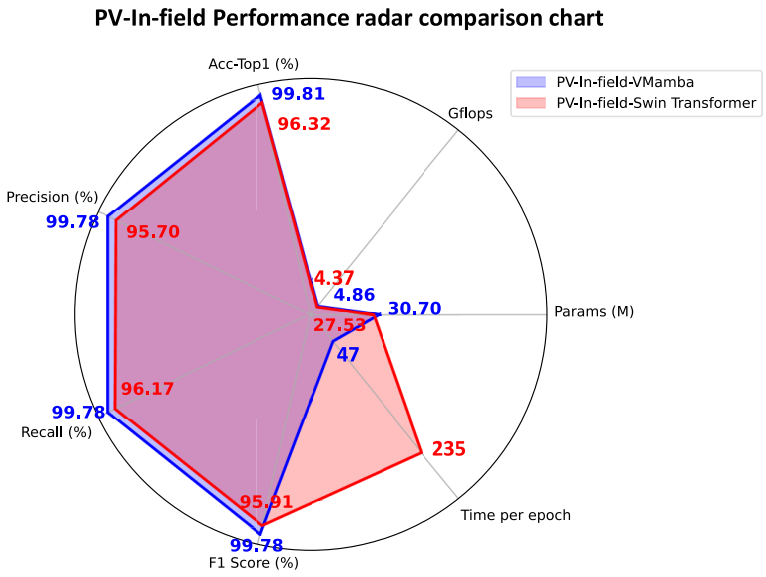


FIGURE 12
PV-In-field performance radar comparison chart.

and ResNet models in handling small-sample data and complex agricultural background data, demonstrating superior performance. Compared to Swin Transformer, proposed by Liu et al. (2021), which is currently the state-of-the-art (SOTA) method, the VMamba model shows better performance on complex agricultural images and small-sample agricultural data, as shown in Figure 12. Swin Transformer is a hierarchical architecture based on Transformers, which significantly reduces computational complexity by introducing local window multi-head self-attention mechanisms (W-MSA) while retaining feature extraction capabilities. Its W-MSA mechanism divides images into fixed-size, non-overlapping windows, making it difficult to capture global dependencies across windows, thus limiting its performance in agricultural disease identification tasks. For example, in citrus leaves infected with Huanglongbing (HLB), lesions may be scattered across the leaf tip, leaf margin, and petiole regions. Window partitioning can disrupt the association between these areas, reducing the model's overall perception of the disease.

The VMamba model, based on state-space models (SSM), uses a selective scanning mechanism (S6) to achieve global context modeling through linear scanning. VMamba processes image sequences independently from four directions and integrates multi-directional context information during the merging stage. This design allows it to capture various distribution patterns of leaf diseases, whereas the single-direction window partitioning of Swin Transformer struggles to adapt to such complex morphologies. Experiment 3 shows that VMamba's Acc-Top1 is 3.49% higher than Swin Transformer (99.81% vs. 96.32%) in complex background interference, validating its advantage in global modeling. In small-sample datasets, the W-MSA mechanism of Swin Transformer initially provides a slight accuracy advantage over VMamba in early training stages. However, after DDPM data augmentation and pre-training optimization, VMamba's global attention-aware capability surpasses that of Swin Transformer, ResNet50, and ViT models.

Although the VMamba model showed excellent performance in this study, there are some limitations. First, while data augmentation and transfer learning significantly improve small-sample dataset performance, these methods rely on high-quality augmented data and pre-trained models. If the original data quality is poor, it may negatively affect the model's performance. Moreover, this study only experimented with disease identification in citrus plants; future research should further evaluate VMamba's applicability in other crops and disease identification tasks.

6 Conclusions and future work

6.1 Limitations

VMamba performs exceptionally well in high-performance computing environments, but its deployment on mobile and edge computing devices presents certain challenges. These devices typically have limited computational power and memory, which may lead to issues such as memory overflow and high computational latency when running VMamba directly. Although VMamba reduces computational complexity through its S6 mechanism, with a computational cost of 4.85 Gflops, which is significantly lower than that of mainstream models like ViT (16.86 Gflops), its real-time performance on edge devices—such as frame rate and power consumption—still requires further testing. In practical applications like farmland monitoring, additional optimizations may be necessary, such as model pruning, to reduce the model size and computational load.

The current experiments were primarily validated on citrus diseases. While VMamba demonstrated excellent transfer learning performance on multi-crop datasets (e.g., PlantVillage) and small-sample datasets (e.g., In-field small dataset), its transferability to other crops such as rice and wheat remains to be further explored. The small-sample dataset used in this study was sourced from real agricultural environments and exhibits certain regional characteristics. Although we adhered to ethical standards during data collection to ensure the legality and accuracy of the data, and took measures to minimize bias, future research should aim to collect image data from different regions and crop types. This will help validate VMamba's performance across diverse conditions and reduce errors.

6.2 Conclutions

In this study, we introduced the emerging VMamba vision backbone model for the task of agricultural plant disease identification, including small-sample plant disease recognition. Experimental results demonstrated that, whether applied to large datasets like PlantVillage or optimized small-sample agricultural disease datasets, VMamba consistently outperformed popular models such as ResNet50, Vision Transformer (ViT), and Swin Transformer in terms of accuracy for plant disease identification. These results verify VMamba's excellent performance in the field, offering a novel approach for agricultural plant disease recognition. Additionally, we proposed the DDHTL-VMamba training method for small-sample agricultural disease datasets. Compared to transfer learning using large datasets, this method largely maintained the accuracy of plant disease identification while significantly reducing the time required for pre-training, thereby providing a more efficient and balanced solution.

6.3 Future work

Future research could focus on evaluating VMamba's performance on larger and more diverse field datasets, particularly in real-world agricultural environments where crop conditions and disease manifestations are often more complex. Further improvements could be made by integrating other data augmentation techniques or optimizing transfer learning strategies

to enhance the model's generalization capabilities on small-sample datasets. Another promising direction would be exploring the application of VMamba in multimodal data (e.g., spectral images and sensor data) to broaden its potential 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 author.

Author contributions

HZ: Data curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. SL: Formal Analysis, Investigation, Validation, Writing – review & editing. JX: Data curation, Formal Analysis, Software, Writing – review & editing. ZC: Validation, Visualization, Writing – review & editing. JC: Conceptualization, Investigation, Writing – review & editing. JG: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work is supported in part by the Dongguan Science and Technology of Social Development Program (20221800905102), and Project of Education Department of Guangdong Province (2022ZDZX4053, pdjh2022b0512).

Conflict of interest

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

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. During the preparation of this work the authors used ChatGPT in order to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

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

Bedi, P., and Gole, P. (2021). Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network. *Artif. Intell. Agric.* 5, 90–101. doi: 10.1016/j.aiia.2021.05.002

Bedi, P., Gole, P., and Marwaha, S. (2024). PDSE-Lite: lightweight framework for plant disease severity estimation based on Convolutional Autoencoder and Few-Shot Learning. *Front. Plant Sci.* 14, 1319894. doi: 10.3389/fpls.2023.1319894

Borhani, Y., Khoramdel, J., and Najafi, E. (2022). A deep learning based approach for automated plant disease classification using vision transformer. *Sci. Rep.* 12, 11554. doi: 10.1038/s41598-022-15163-0

Chen, C.-S., Chen, G.-Y., Zhou, D., Jiang, D., and Chen, D.-S. (2024). Res-VMamba: fine-grained food category visual classification using selective state space models with deep residual learning. *arXiv preprint arXiv:2402.15761*. doi: 10.48550/arXiv.2402.15761

Chen, J., Guo, J., Zhang, H., Liang, Z., and Wang, S. (2024a). Weakly supervised localization model for plant disease based on siamese networks. *Front. Plant Sci.* 15, 1418201. doi: 10.3389/fpls.2024.1418201

Chen, J., Wang, S., Guo, J., Chen, F., Li, Y., and Qiu, H.. (2024b). Improved FasterViT model for citrus disease diagnosis. *Heliyon* 10 (16). doi: 10.1016/j.heliyon.2024.e36092

Chollet, François. (2017). "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. Honolulu, Piscataway, New York, USA. 2017, 1251–1258.

Dang, T. D. Q., Nguyen, H. H., and Tiulpin, A. (2024). LoG-VMamba: local-global vision mamba for medical image segmentation. *arXiv preprint arXiv:2408.14415*. doi: 10.48550/arXiv:2408.14415

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. doi: 10.48550/arXiv

Fu, X., Ma, Q., Yang, F., Zhang, C., Zhao, X., Chang, F., et al. (2024). Crop pest image recognition based on the improved ViT method. *Inf. Process. Agric.* 11, 249–259. doi: 10.1016/j.inpa.2023.02.007

Geetharamani, G., and Pandian, A. (2019). Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Comput. Electrical Eng.* 76, 323–338. doi: 10.1016/j.compeleceng.2019.04.011

Gole, P., Bedi, P., Marwaha, S., Haque, A., and Deb, C. K. (2023). TrIncNet: a lightweight vision transformer network for identification of plant diseases. *Front. Plant Sci.* 14, 12215575. doi: 10.3389/fpls.2023.1221557

Gu, A., and Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*. doi: 10.48550/arXiv:2312.00752

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. Applied sciences, MDPI, Basel, Switzerland, 12(18), 8972.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* 33, 6840–6851. doi: 10.48550/arXiv.2006.11239

Hossain, E., Hossain, M. F., and Rahaman, M. A. (2019). "A color and texture-based approach for the detection and classification of plant leaf disease using KNN classifier," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, IEEE, Piscataway, New York, USA. 1–6.

Joshi, B. M., and Bhavsar, H. (2020). Plant leaf disease detection and control: A survey. *J. Inf. Optimization Sci.* 41, 475–487. doi: 10.1080/02522667.2020.1734295

Kamal, K. C., Yin, Z., Wu, M., and Wu, Z. (2019). Depthwise separable convolution architectures for plant disease classification. *Comput. Electron. Agric.* 165, 104948. doi: 10.1016/j.compag.2019.104948

Lee, S. H., Goau, H., and Joly, A. (2020). New perspectives on plant disease characterization based on deep learning. *Comput. Electron. Agric.* 170, 105220. doi: 10.1016/j.compag.2020.105220

Li, L., Zhang, S., and Wang, B. (2021). Plant disease detection and classification by deep learning—A review. *IEEE Access IEEE* 9, 56683–56698. doi: 10.1109/ACCESS.2021.3069646

Li, Y., Guo, J., Qiu, H., Fengyi, C., and Zhang, J. (2023). Denoising Diffusion Probabilistic Models and Transfer Learning for citrus disease diagnosis. *Front. Plant Sci.* 14, 1267810. doi: 10.3389/fpls.2023.1267810

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*. Piscataway, New York, USA, 10012–10022. doi: 10.48550/arXiv.2401.10166

Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., et al. (2024). Vmamba: Visual state space model. *arXiv preprint arXiv:2401.10166*.

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7, 215232. doi: 10.3389/fpls.2016.01419

Naeem, M. A., Saleem, M. A., Sharif, M. I., Akber, S., Saleem, S., Akhtar, Z., et al. (2025). Deep learning-based approach for identification of potato leaf diseases using wrapper feature selection and feature concatenation. *arXiv preprint arXiv:2502.03370*. doi: 10.48550/arXiv:2502.03370

Saleem, S., Sharif, M. I., Sharif, M. I., Sajid, M. Z., and Marinello, F. (2024). Comparison of deep learning models for multi-crop leaf disease detection with enhanced vegetative feature isolation and definition of a new hybrid architecture. *Agronomy* 14, 2230. doi: 10.3390/agronomy14102230

Saleem, S., Hussain, A., Majeed, N., Akhtar, Z., and Siddique, K. (2025). A multi-scale feature extraction and fusion deep learning method for classification of wheat diseases. *arXiv preprint arXiv:2501.09938*. doi: 10.48550/arXiv

Sankaran, S., Mishra, A., Ehsani, R., and Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Comput. Electron. Agric.* 72, 1–13. doi: 10.1016/j.compag.2010.02.007

Shi, Y., Dong, M., and Xu, C. (2024). Multi-scale VMamba: hierarchy in hierarchy visual state space model. *arXiv preprint arXiv:2405.14174*. doi: 10.48550/arXiv:2405.14174

Shoaib, M., Shah, B., Ei-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., et al. (2023). An advanced deep learning models-based plant disease detection: A review of recent research. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1158933

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

Thakur, P. S., Chaturvedi, S., Khanna, P., Sheorey, T., and Ojha, P. (2023). Vision transformer meets convolutional neural network for plant disease classification. *Ecol. Inf.* 77, 1022455. doi: 10.1016/j.ecoinf.2023.102245

Thomas, S., Kuska, M. T., Bohnenkamp, D., Brugger, A., Alisaac, E., Wahabzada, M., et al. (2018). Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective. *J. Plant Dis. Prot.* 125, 5–20. doi: 10.1007/s41348-017-0124-6

Yao, H., Zhang, C., Wei, Y., Jiang, M., Wang, S., Huang, J., et al. (2020). "Graph few-shot learning via knowledge transfer," in *Proceedings of the AAAI conference on artificial intelligence*, Palo Alto, California, USA. Vol. 34(04), 6656–6663.

Zhang, Z., Gong, Z., Hong, Q., and Jiang, L. (2021). "Swin-transformer based classification for rice diseases recognition," in *2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI). IEEE*. Piscataway, New Jersey, USA. IEEE.

Zhang, S., Zhang, S., Zhang, C., Wang, X., and Shi, Y. (2019). Cucumber leaf disease identification with global pooling dilated convolutional neural network. *Comput. Electron. Agric.* 162, 422–430. doi: 10.1016/j.compag.2019.03.012

Zhao, W. (2017). "Research on the deep learning of the small sample data based on transfer learning," in *AIP conference proceedings*, vol. 1864. (Melville, New York, USA: AIP Publishing).

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., et al. (2020). A comprehensive survey on transfer learning. *Proc. IEEE* 109, 43–76. doi: 10.1109/PROC.5

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 08 May 2025
DOI 10.3389/fpls.2025.1581367

Check for updates

OPEN ACCESS

EDITED BY
Jingcheng Zhang,
Hangzhou Dianzi University, China

REVIEWED BY
Pengchao Chen,
South China Agricultural University, China
Weicheng Xu,
Guangdong Academy of Agricultural Sciences (GDAAS), China

*CORRESPONDENCE
Zhihong Li
lizh@cau.edu.cn
Qing Chen
chqingztq@163.com

RECEIVED 22 February 2025
ACCEPTED 14 April 2025
PUBLISHED 08 May 2025

CITATION
Liu Y, Liang X, Wu C, An X, Wu M, Zhao Z, Li Z and Chen Q (2025) Optimization of spray operation parameters of unmanned aerial vehicle confers adequate levels of control of fall armyworm (*Spodoptera frugiperda*). *Front. Plant Sci.* 16:1581367. doi: 10.3389/fpls.2025.1581367

COPYRIGHT
© 2025 Liu, Liang, Wu, An, Wu, Zhao, Li 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.

Optimization of spray operation parameters of unmanned aerial vehicle confers adequate levels of control of fall armyworm (*Spodoptera frugiperda*)

Ying Liu[1,2], Xiao Liang[1,2], Chunling Wu[1,2], Xingkui An[1,2], Mufeng Wu[1,2], Zihua Zhao[3], Zhihong Li[3]* and Qing Chen[1,2]*

[1]Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences/Key Laboratory of Integrated Pest Management on Tropical Crops, Ministry of Agriculture and Rural Affairs, Haikou, China, [2]Sanya Research Academy, Chinese Academy of Tropical Agriculture Science/Hainan Key Laboratory for Biosafety Monitoring and Molecular Breeding in Off-Season Reproduction Regions, Sanya, China, [3]Sanya Institute of China Agricultural University, Sanya, China

Introduction: The fall armyworm (FAW), *Spodoptera frugiperda*, is a serious threat to maize production. Unmanned aerial vehicles (UAVs) represent a promising method for controlling FAW outbreaks. Given that FAW larvae primarily feed inside the maize whorl, we hypothesized that the efficacy of insecticide application depends on droplet density and coverage rate on the upper maize canopy.

Methods: This study evaluated the effects of spray operation parameters—including three flight heights (2.0, 2.5, and 3.0 m), three spray volumes (30.0, 37.5, and 45.0 L ha^{-1}), and two nozzle types (XR11001VS and XR110015VS)—on droplet deposition distribution in maize canopies. Additionally, the control efficacy of 35% chlorantraniliprole water dispersible granules (WDG) against FAW was assessed over three consecutive years (2019-2021) to determine its correlation with droplet deposition.

Results: Results indicated that flight height, spray volume, and nozzle type significantly influenced droplet deposition distribution. Two treatments—XR110015VS nozzle with 37.5 L ha^{-1} spray volume at 2.5 m flight height and XR110015VS nozzle with 45.0 L ha^{-1} spray volume at 2.5 m flight height—achieved the highest droplet density, optimal coverage rate on the upper maize canopy, and the lowest damage index, confirming our hypothesis. These treatments also demonstrated comparable FAW control efficacy to traditional electric air-pressure knapsack sprayers.

Discussion: The findings provide practical insights for optimizing UAV-based insecticide applications to improve FAW management in maize production systems.

KEYWORDS

unmanned aerial vehicle, fall armyworm, spray operation parameter, droplet deposition, control effect

1 Introduction

Maize (*Zea mays* L.), cultivated across 44.22 million hectares in China (National Bureau of Statistics of China, 2024), faces a significant threat from the Fall armyworm (FAW), *Spodoptera frugiperda* (J. E. Smith, 1797) (Lepidoptera: Noctuidae). Originating in the Americas (Goergen et al., 2016; Bateman et al., 2018; Boaventura et al., 2020), FAW has spread to 22 Chinese provinces (Shan et al., 2022) driven by its adaptability to diverse hosts and environments (Nagoshi et al., 2019; Chinwada et al., 2023), high reproductive capacity (Bateman et al., 2018; Kumar et al., 2022), and long-distance migratory ability (Nagoshi et al., 2012; Westbrook et al., 2016). Early-instar larvae typically hide in the maize whorl and feed on upper leaves, creating small window panes (Bateman et al., 2018), while late-instar larvae bore into maize cobs to feed on kernels (Kansiime et al., 2019), making control challenging.

Traditionally, chemical insecticides applied using knapsack sprayers have been the primary FAW control method in China (Ren et al., 2019; Wang G. et al., 2019). However, these methods are inefficient and pose significant risks to the environment and human health, especially during critical growth stages such as the late whorl and tasseling stages. Therefore, improving the mechanization and precision of pesticide applications is essential for effective FAW control while minimizing environmental and operator risks.

Unmanned Aerial Vehicles (UAVs) have emerged as a promising alternative, offering enhanced mobility, reduced labor costs, and lower operator exposure compared to manual or ground-based spraying (Wang G. et al., 2018; Zhang et al., 2018). UAVs have a working efficiency of 4–10 hectares per hour, which is 30 to 100 times higher than manual spraying and 4 to 33 times higher than ground-based sprayers (Yang et al., 2018). This high efficiency makes UAVs particularly suitable for covering large areas within narrow optimal application windows, crucial for effective FAW control.

The effectiveness of UAV-based pesticide application depends significantly on key operational parameters such as spray volume, nozzle type, and flight height, which influence droplet deposition and control efficacy. For example, higher spray volumes and coarser nozzles improve deposition and control efficacy for wheat aphids and powdery mildew (Wang G. et al., 2019), while larger droplet sizes enhance penetration and uniformity in rice canopies (Chen et al., 2020). Similarly, higher flight heights increase droplet coverage and deposition in cotton (Lou et al., 2018). Taking into account the specific infestation patterns and behaviors of pests is also crucial when applying insecticides via UAV. Qin et al. (2016) obtained a better control efficacy against the brown planthopper by optimizing operation height and velocity, with the highest droplet deposition at a spraying height of 1.5 m and a speed of 3 $m \cdot s^{-1}$ on the lower canopy of rice, where planthoppers generally occur. Given that FAW larvae primarily feed in the maize whorl, we hypothesize that optimizing UAV spray parameters to maximize droplet deposition on the upper canopy will significantly enhance FAW control.

This study systematically evaluates droplet deposition and distribution on maize canopies under varying UAV parameters, including spray volume, nozzle type, and flight height. By optimizing these parameters, we aim to maximize deposition on the maize whorl, improving FAW control. The findings will guide aerial applications of chlorantraniliprole and support sustainable pest management in maize cultivation.

2 Materials and methods

2.1 Insecticide, maize and field site

The insecticide used in this study was 35% chlorantraniliprole Water Dispersible Granules (WDG) applied at a dose of 52.5 g active ingredient (a.i.) per hectare (Bayer Crop Science Co., Ltd., Beijing, China). The maize cultivar selected for the experiment was 'Meilan Huangguan' (Hainan Lvchuan Seed Co., Ltd., Haikou, China). The experimental field was arranged with row spacing of approximately 65 cm, plant spacing of 35 cm, and an average plant height of 65 cm during the application period. The study was conducted at the Chinese Academy of Tropical Agricultural Sciences in Danzhou, Hainan Province, China. This location was chosen due to its favorable climate conditions for maize cultivation and its history of FAW infestations, making it an ideal setting for evaluating the effectiveness of UAV-based pesticide applications.

2.2 Record of climatic conditions

Weather parameters, including wind speed, were monitored and recorded using the Smart Sensor AR866A Anemometer (Dongguan Wanchuang Electronic Products Co., Ltd., Dongguan, China). Air temperature and relative humidity were measured with the Smart Sensor AR837 Temperature and Humidity Gauge (Dongguan Wanchuang Electronic Products Co., Ltd., Dongguan, China). The air temperature ranged from 24.2°C to 35.5°C, while the relative humidity varied between 46.2% and 54.6%. Wind speeds in all test treatments remained below 1.5 m/s. Detailed weather data for each test treatment are provided in Supplementary Table S1.

2.3 Spraying platform and spraying systems

The eight-rotor agricultural UAV (MG-1P, SZ DJI Technology Co., Ltd., Shenzhen, China) employed in this study featured a modular spraying system comprising a 10 L chemical tank, dual customized diaphragm pumps, four vertically mounted flat-fan nozzles, and an integrated radar altimeter for real-time altitude adjustment and obstacle avoidance (Figure 1A). Nozzles were symmetrically affixed to four spray booms aligned with the UAV arms, oriented vertically downward to optimize droplet trajectory within the rotor-induced airflow. An intelligent planning operation mode was utilized, allowing dynamic adjustments to flow rate, spraying width, spray volume, and flight speed based on site-specific conditions. For comparative analysis, a commercial electric air-pressure (EAP) knapsack sprayer (SX-MD18DA, Zhejiang Xixia Sprayer Co., Ltd., Taizhou, China) was deployed (Figure 1B).

FIGURE 1
The DJI MG-1P UAV sprayer and the operator conducting the spray operation **(A)**; electric air-pressure knapsack (EAP) sprayer (SX-MD18DA) used in the experiment **(B)**.

2.4 Experimental design

In 2019, a study was conducted to investigate the influence of various spray operation parameter combinations (i.e., nozzle type, flight height, and spray volume) on insecticide droplet deposition on maize canopies and the control efficacy against FAW. The objective was to determine the optimal spray operation parameters. From 2020 to 2021, two consecutive years of field trials were carried out to further validate the relationship between control efficiency and droplet deposition distribution. Maize sowing dates were June 15, 2019; August 15, 2020; and July 25, 2021. Throughout these experiments, maize was consistently at the whorl stage of growth. Figure 2 illustrated the technical workflow of this study.

2.5 Treatments and spray operation parameters

Twenty treatments were employed in this study, including eighteen UAV treatments, one EAP treatment and an untreated control. The spray operation parameters for each treatment are summarized

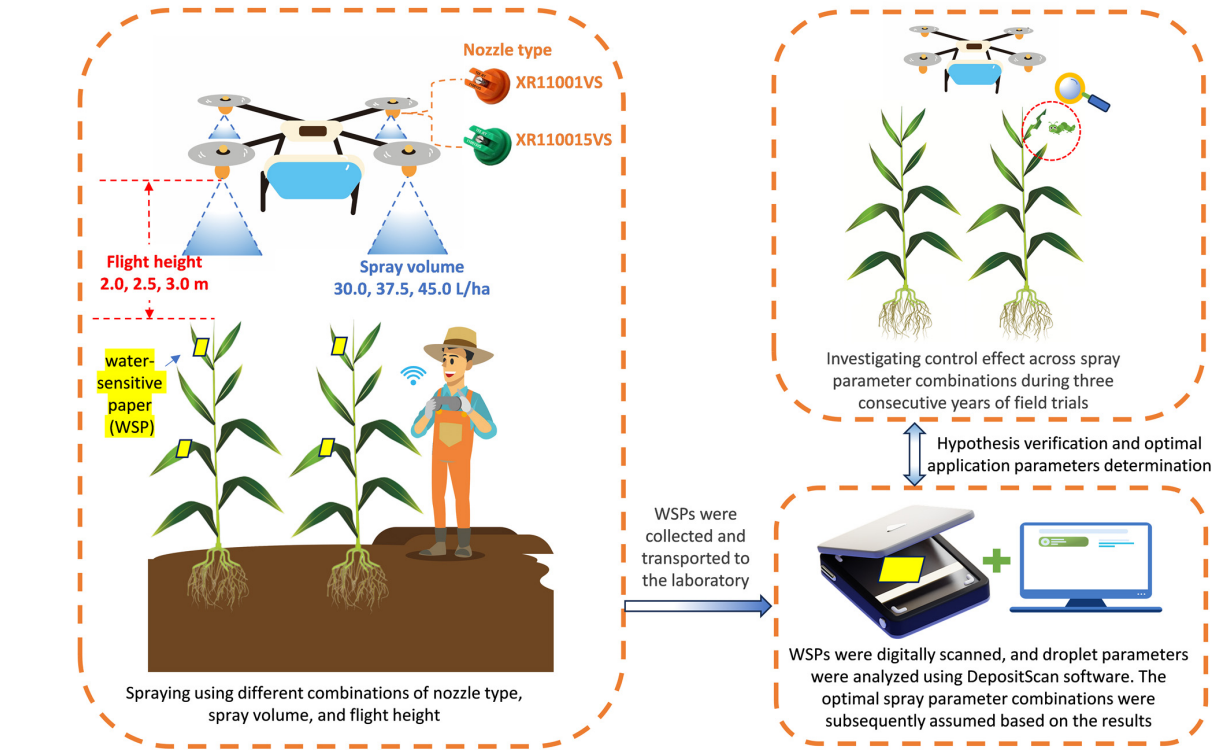


FIGURE 2
Technical workflow of the study.

in Table 1, with each treatment replicated three times. The experimental design followed a randomized block design, as detailed in Supplementary Table S2. Each replicate consisted of a 10 m × 9 m (90 m^2) plot of maize arranged in a rectangular shape (Figure 3A). To minimize drift pollution, treatment plots were separated by a 10 m buffer zone. For the UAV treatments, we evaluated the influence of flight height (2.0, 2.5, and 3.0 m), nozzle type (XR11001VS and XR110015VS), and spray volume (30.0, 37.5, or 45.0 L·ha^{-1}) on droplet density, coverage rate, and droplet size across both the upper and lower canopies of maize plants. The UAV flight route is illustrated in Figure 3A. The red arrow indicates the flight route. The UAV turned around approximately 10 m from the starting point and completed four round trips before reaching the terminal point. After setting the spraying width, spray volume and nozzle discharge, the flight velocity was kept at an automatic setting. The spray volume, spraying width, nozzle discharge and flight velocity were converted using the following equation (Tang et al., 2018).

$$d = \frac{v \times w \times V}{166.67}$$

where d is the nozzle discharge (L·min^{-1}), V is the spray volume (L·ha^{-1}), w is the spraying swath (m), and v is the flight velocity of the UAV (m·s^{-1}). In this study, we selected two commonly used nozzles, TEEJET XR11001VS and TEEJET XR110015VS, for the DJI MG-1P sprayer. These nozzles were chosen due to their ability to accurately calibrate discharge when used interchangeably (Guo et al., 2020). The XR110015VS nozzle has a higher flow rate of 0.15 gallons per minute and larger droplet size of 202 μm compared to the XR11001VS nozzle, which has a flow rate of 0.10 gallons per minute and droplet size of 174 μm, both measured at a pressure of 0.276 MPa (Yu et al., 2020). Additionally, for the XR11001VS nozzle, the droplet diameter range is 130–250 μm, while for the XR110015VS nozzle, the range is 170–265 μm (available at https://www.dji.com/cn/products/compare-agriculture). In this study, the working pressure for both nozzles was set to 0.4 MPa. The UAV and EAP sprayer were operated by well-trained operators. The EAP application was conducted at a spray volume of 450 L·ha^{-1}, a working pressure of 0.4 MPa, and a travelling speed of approximately 0.3 m/s.

TABLE 1 Spray operation parameters in each treatment.

Treatment	Spraying equipment	Flight height (m)	Nozzle type	Spray volume (L·ha^{-1})	Travelling speed (m·s^{-1})	Spraying width (m)	Nozzle discharge (L·min^{-1})
1	UAV	2.0	XR11001VS	30.0	3.3	3	1.8
2				37.5	2.7	3	1.8
3				45.0	2.2	3	1.8
4			XR110015VS	30.0	4.6	3	2.5
5				37.5	3.7	3	2.5
6				45.0	3.1	3	2.5
7		2.5	XR11001VS	30.0	3.3	3	1.8
8				37.5	2.7	3	1.8
9				45.0	2.2	3	1.8
10			XR110015VS	30.0	4.6	3	2.5
11				37.5	3.7	3	2.5
12				45.0	3.1	3	2.5
13		3.0	XR11001VS	30.0	3.3	3	1.8
14				37.5	2.7	3	1.8
15				45.0	2.2	3	1.8
16			XR110015VS	30.0	4.6	3	2.5
17				37.5	3.7	3	2.5
18				45.0	3.1	3	2.5
19	EAP	-[a]	Hollow cone nozzle	450.0	0.3	-[b]	0.8
20	Control	–	–	–	–	–	–

[a]Nozzle moved from the top to the bottom of the canopy and the droplets distributed over the canopy layer as soon as possible.
[b]Sprayed along the maize row and the row spacing was 0.65 m.

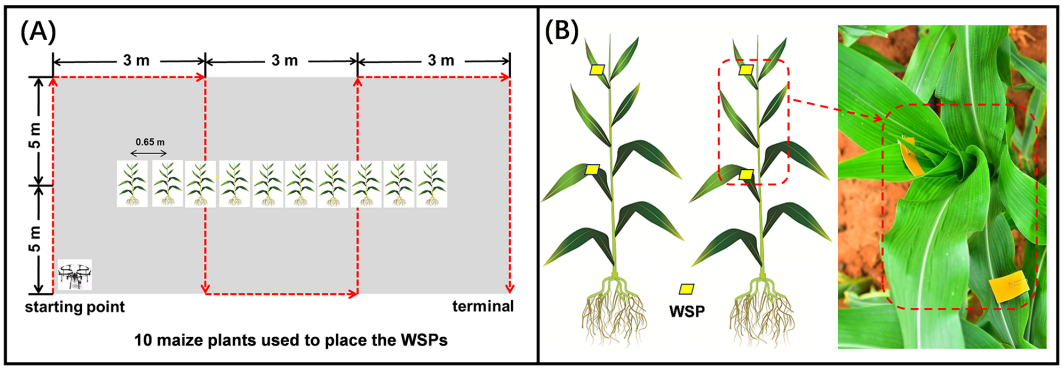


FIGURE 3
Schematic diagram of the site layout and flight route **(A)**; placement of WSPs at each sampling position within the maize canopy **(B)**.

2.6 Sample scheme and characterization of droplet parameters

In this study, water-sensitive paper (WSP, 30 mm×80 mm, Chongqing Liuliu Shanxia Plant Protection Technology Co., Ltd., Chongqing, China) was utilized to characterize droplet characteristics and canopy distribution. Prior to each application treatment, two WSPs were placed in the upper canopy (the seventh leaf, near the maize whorl) and the lower canopy (the third leaf) of the maize plants (Figure 3B). Two WSPs were placed on each of ten maize plants per plot, with each plant separated by 0.65 m (Figure 3A). The objectives of using the WSPs were to assess droplet parameters, including droplet density, coverage rate and droplet size at different canopy levels. Specifically, the WSP placed on the seventh leaf (located in the maize whorl) was crucial because the droplet deposition on this site can best represent the FAW control efficiency. After each spraying treatment, WSPs were collected and bagged in zip-lock plastic bags labeled with the treatment and replicate information. The WSPs were scanned as digital images at a resolution of 600 dpi. Droplet parameters (droplet density, coverage rate, and droplet size) were analyzed using the DepositScan software (Zhu et al., 2011). The analysis of droplet size followed the method described by Tang et al. (2018). The influence of operation parameters on droplet density, coverage rate and droplet size on both the upper and lower canopies of maize plants was analyzed for UAV treatments. However, droplet parameters for EAP treatments were not shown due to the excessively large spray volume, which precluded meaningful droplet distribution analysis on the WSP.

2.7 Control effect against FAW

The damage index was used to evaluate the control effect of each treatment against FAW. The damage index for each treatment plot was calculated according to the damage index equation.

$$Damage\ \ index = \frac{\sum (Number\ \ of\ \ damage\ \ leaves\ \ at\ \ each\ \ level \times Corresponding\ \ level\ \ value)}{Total\ \ number\ \ of\ \ investigation \times 9} \times 100$$

This evaluation was conducted on new leaves of 50 maize plants in each plot at 3 days and 7 days after treatment (DAT), using a modified scale of 0–9 developed by Davis et al (Toepfer et al., 2021). The specific field investigation method was as follows: a five-point sampling method (W-pattern approach) was employed in each plot, and at each point, the scout assessed 10 plants for signs of FAW feeding damage (refer to the damage index scale). During the initial assessment at 3 DAT, all maize plants were marked with red string to ensure consistent monitoring of the same plants at 7 DAT (Shan et al., 2022).

2.8 Statistical analysis

Data were analyzed using SPSS v. 19.0 (SPSS Inc., Chicago, IL, USA). Prior to analysis, data normality was assessed using the Shapiro-Wilk test ($P > 0.05$) and homogeneity of variance was evaluated using Levene's test ($P > 0.05$). To stabilize variances and meet normality assumptions, the coverage rate was transformed using $y = \arcsin\sqrt{X}/100$, while droplet density and size were log (X + 1) transformed. One-way ANOVA was used to detect significant differences in droplet parameters (droplet density, coverage rate, and droplet size) or damage index among treatments, followed by the Tukey's *post hoc* test with $P< 0.05$. Student's t-test was used to compare significant differences in droplet parameters between the two nozzle types.

3 Results

3.1 Effect of spray volume and nozzle type on droplet deposition characteristics in maize canopies at different flight heights

Effect of spray volume and nozzle type on droplet deposition characteristics in maize canopies at the flight height of 2.0 m are shown in Figure 4. At a flight height of 2.0 m using the nozzle XR11001VS, spray volume significantly influenced droplet density and coverage rate in both the upper and lower canopies (Figures 4A,

10.3389/fpls.2025.1581367

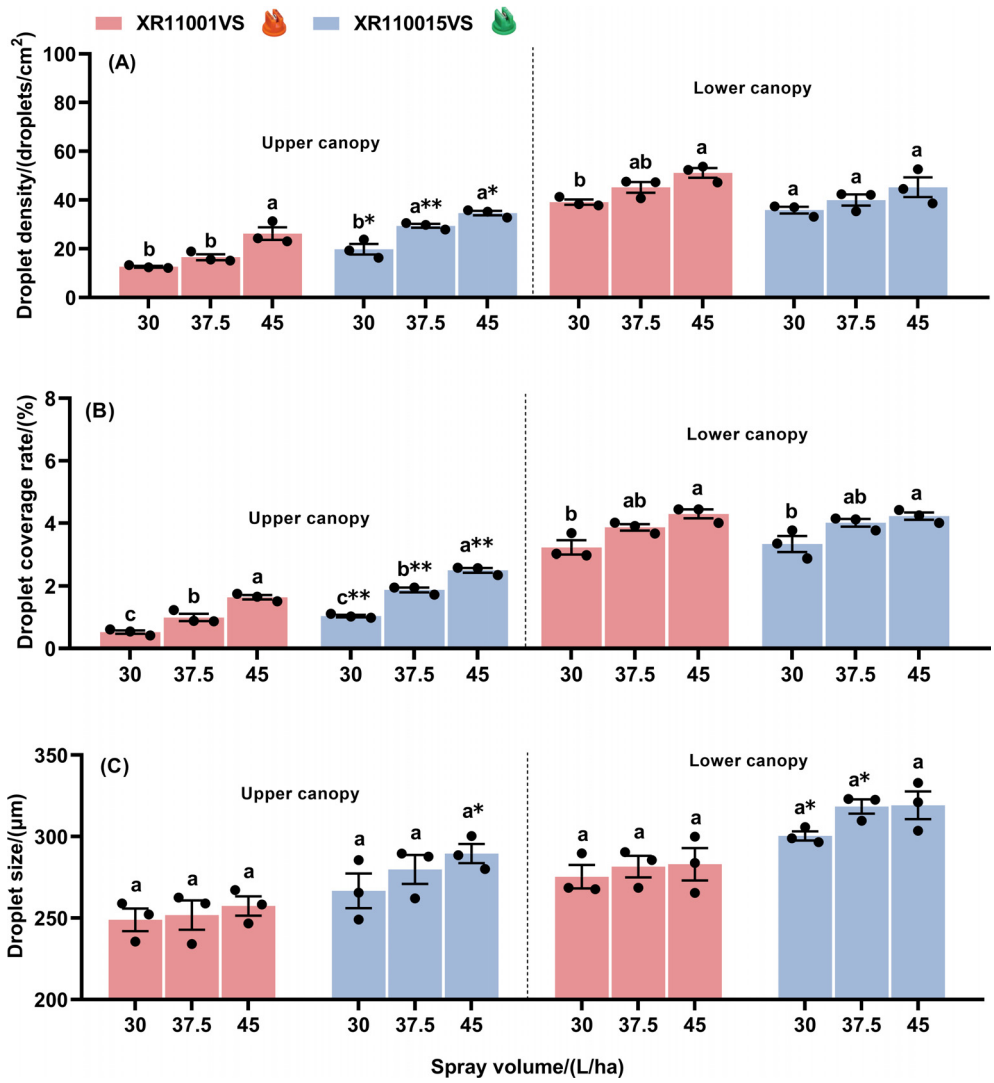


FIGURE 4
Droplet density **(A)**, droplet coverage rate **(B)** and droplet size **(C)** on the upper and lower canopies of maize plant with two nozzle types, across three spray volumes, at a fixed flight height of 2.0 m. Different letters on bars indicate significant differences ($P< 0.05$) among spray volumes. Asterisks above the bars denote significant difference of droplet density (droplet coverage rate or drop size) between two nozzles by using t-test, at a certain spray volume on a same canopy. * and ** indicate significant difference at the significance levels of 0.05 and 0.01, respectively; ns represents no significance. Error bars denote standard error of the means.

B). Both droplet density and coverage rate generally increased with spray volume, with a significant increase observed when comparing the lowest (30 L/ha) and highest (45 L/ha) spray volumes. Droplet size remained consistent across different spray volumes for both the upper and lower canopies (Figure 4C). When using the nozzle XR110015VS, droplet density on the upper canopy was lowest at 30 L/ha and highest at 37.5 and 45 L/ha. No significant differences in droplet density were detected on the lower canopy. The coverage rate on the upper canopy increased significantly with spray volume. In contrast, on the lower canopy, the coverage rate at 37.5 L/ha did not differ significantly from either 30 or 45 L/ha. Droplet size did not vary significantly with changes in spray volume for both the upper or lower canopies. Furthermore, when comparing the effects of the two nozzle types (XR110015VS and XR11001VS) at different spray volumes, we found droplet density and coverage rate were significantly higher with the XR110015VS nozzle at all tested volumes. In contrast, no significant differences in droplet density or coverage rate were observed on the lower canopy between the two nozzle types. Regardless of canopy position, the droplet size produced by the nozzle XR110015VS was consistently larger than that of the nozzle XR11001VS at the same spray volume. Significant differences in droplet size were observed on the upper canopy at a spray volume of 45 L/ha and on the lower canopy at spray volumes of 30 L/ha and 37.5 L/ha. In summary, the XR110015VS nozzle demonstrated superior droplet density and coverage on maize whorls across three spray volumes at a 2.0 m flight height compared to the XR11001VS nozzle.

Effect of spray volume and nozzle type on droplet deposition characteristics in maize canopies at the flight height of 2.5 m are shown in Figure 5. At a height of 2.5 m using nozzle XR11001VS,

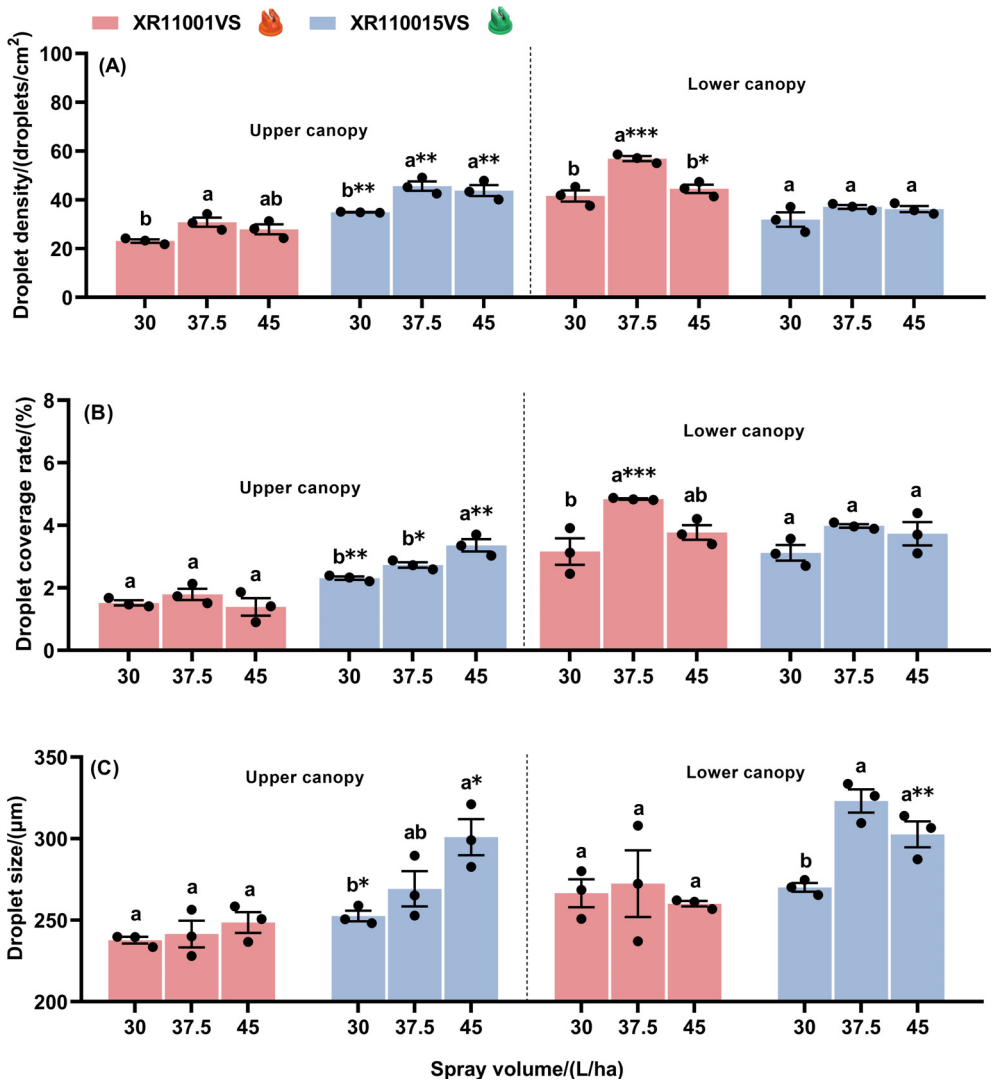


FIGURE 5
Droplet density **(A)**, droplet coverage rate **(B)** and droplet size **(C)** on the upper and lower canopies of maize plant with two nozzle types, across three spray volumes, at a flight height of 2.5 m. Different letters on bars indicate significant differences ($P< 0.05$) among spray volumes. Asterisks above the bars denote significant difference of droplet density (droplet coverage rate or drop size) between two nozzles by using *t*-test, at a certain spray volume on a same canopy. *, ** and *** indicate significant difference at the significance levels of 0.05, 0.01 and 0.001 levels, respectively; ns represents no significance. Error bars denote standard error of the means.

spray volume significantly influenced droplet density on both the upper canopy and the lower canopy (Figure 5A). For droplet coverage rate, significant variation was observed among different spray volumes on the lower canopy, while no significant difference was detected on the upper canopy (Figure 5B). The highest droplet density and coverage rate were achieved with a spray volume of 37.5 L·ha^{-1} for both the upper and lower canopies. Additionally, droplet size did not show statistically significant differences among spray volumes on either the upper or lower canopy (Figure 5C). When using the nozzle XR110015VS, droplet densities at spray volumes of 37.5 and 45.0 L ha^{-1} were significantly higher than at 30.0 L ha^{-1} on the upper canopy, but no significant differences were observed on the lower canopy (Figure 5A). For droplet coverage rate, a volume of 45.0 L·ha^{-1} resulted in significantly higher coverage compared to 30.0 L·ha^{-1} and 37.5 L·ha^{-1} on the upper canopy, whereas no significant difference was found on the lower canopy (Figure 5B). Spray volume significantly influenced droplet size on both the upper canopy and the lower canopy (Figure 5C). On the upper canopy, droplet size generally increased with spray volume, while on the lower canopy, droplet size at spray volumes of 37.5 and 45.0 L ha^{-1} were significantly larger than at 30.0 L ha^{-1}. Furthermore, when comparing the two nozzles (XR110015VS and XR11001VS), we found significant differences in performance. For the upper canopy, the XR110015VS nozzle exhibited significantly higher droplet density and coverage rate at all tested volumes. In contrast, for the lower canopy, the XR11001VS nozzle performed better. The droplet density using the XR11001VS nozzle was significantly higher than that of the XR110015VS nozzle at spray volumes of 37.5 L·ha^{-1} and 45.0 L·ha^{-1}. The coverage rate using the XR11001VS nozzle was significantly higher at 37.5 L·ha^{-1}.

Regardless of whether on the upper or lower maize canopy, the droplet size produced by the nozzle XR110015VS was consistently larger than that of the nozzle XR11001VS at the same spray volume. Significant differences were observed on the upper canopy at spray volumes of 30 L/ha and 45.0 L/ha, and on the lower canopy at a spray volume of 45.0 L/ha. In summary, the XR110015VS nozzle demonstrated superior droplet density and coverage on maize whorls across three spray volumes at a 2.5 m flight height compared to the XR11001VS nozzle.

Effect of spray volume and nozzle type on droplet deposition characteristics in maize canopies at the flight height of 3.0 m are shown in Figure 6. At a flight height of 3.0 m using the nozzle XR11001VS, spray volume significantly influenced droplet density on the upper canopy, but no significant difference was detected on the lower canopy (Figure 6A). Droplet coverage was lowest at 30 L/ha and highest at both 37.5 and 45 L/ha in both the upper and lower canopies (Figure 6B). Both droplet density and coverage rate demonstrated a progressive increase with higher spray volumes. Droplet size remained unaffected by spray volume on both the upper and the lower canopy (Figure 6C). When using the nozzle XR110015VS, spray volume significantly influenced droplet density on the upper canopy, whereas no significant difference was detected on the lower canopy (Figure 6A). For droplet coverage rate, no significant differences were detected on the upper canopy, while spray volume significantly influenced coverage rate on the lower canopy (Figure 6B). Both droplet density and coverage rate demonstrated a progressive increase with higher spray volumes. The droplet size did not vary with spray volume on either the upper canopy or the lower canopy (Figure 6C). Furthermore, when comparing the two nozzles (XR110015VS and XR11001VS), we

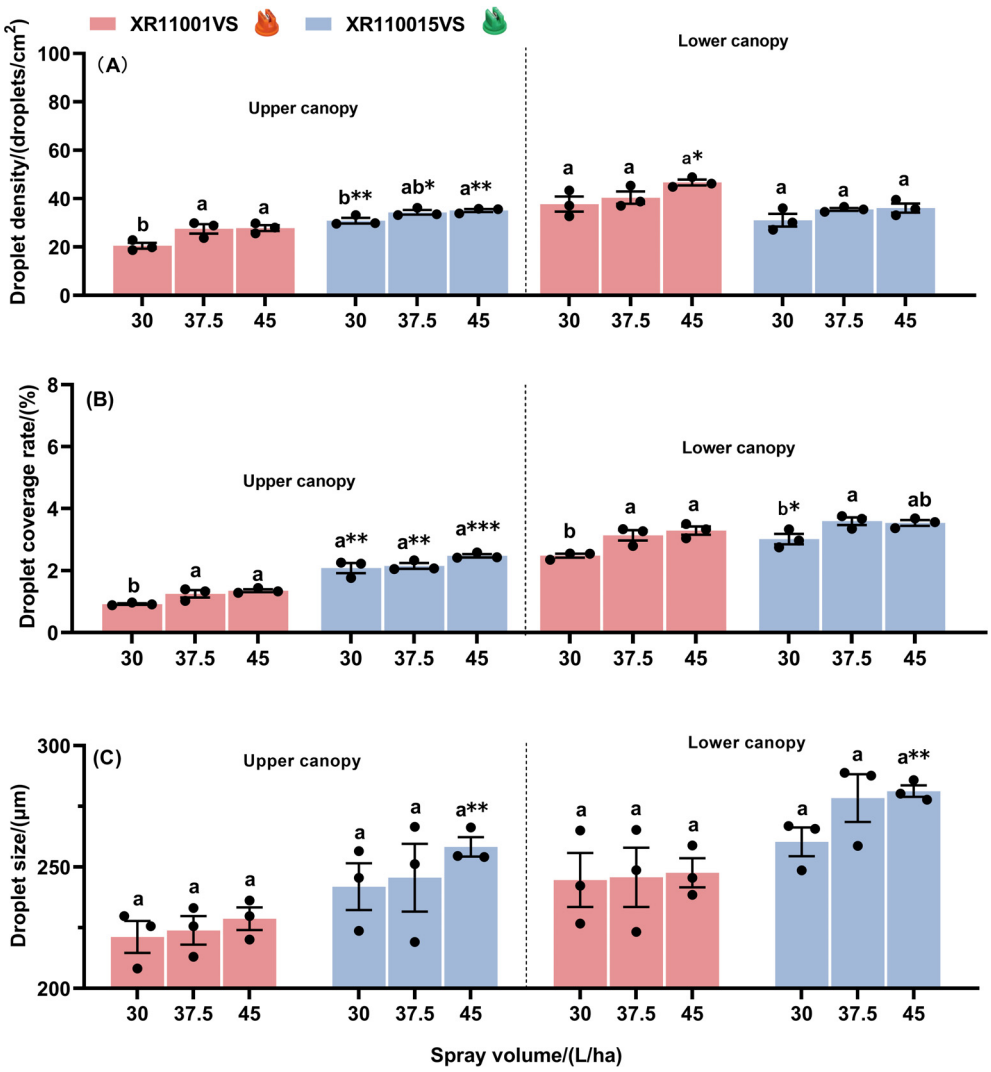


FIGURE 6
Droplet density **(A)**, droplet coverage rate **(B)** and droplet size **(C)** on the upper and lower canopies of maize plant with two nozzle types, across three spray volumes, at a flight height of 3.0 m. Different letters on bars indicate significant differences ($P< 0.05$) among spray volumes. Asterisks above the bars denote significant difference of droplet density (droplet coverage rate or drop size) between two nozzles by using *t*-test, at a certain spray volume on a same canopy. *, ** and *** indicate significant difference at the significance levels of 0.05, 0.01 and 0.001 levels, respectively; ns represents no significance. Error bars denote standard error of the means.

10.3389/fpls.2025.1581367

found that for the upper canopy, the XR110015VS nozzle exhibited significantly higher droplet density and coverage rate at all tested volumes. For the lower canopy, droplet density and coverage rate showed no significant differences between the two nozzle types except for two specific cases. The droplet density using the XR11001VS nozzle was significantly higher than that of the XR110015VS nozzle at a spray volume of 45 L·ha^{-1}. Additionally, the coverage rate using the XR110015VS nozzle was significantly higher than that of the XR11001VS nozzle at a spray volume of 37.5 L·ha^{-1}. Regardless of whether on the upper or lower maize canopy, the droplet size produced by the nozzle XR110015VS was consistently larger than that of the nozzle XR11001VS at the same spray volume. Significant differences in droplet size were observed at a spray volume of 45.0 L/ha both on the the upper canopy and the lower canopy. In summary, the XR110015VS nozzle demonstrated superior droplet density and coverage on maize whorls across three spray volumes at a 3.0 m flight height compared to the XR11001VS nozzle.

To sum up, the XR110015VS nozzle demonstrated superior droplet density and coverage rates on maize upper canopies compared to the XR11001VS nozzle. Given these findings, we selected the nozzle XR110015VS for further analysis.

3.2 Determination of optimal flight height and spray volume

When using the XR110015VS nozzle, the effects of flight height on droplet deposition characteristics on the upper canopy were systematically compared and analyzed across different spray volumes (Figure 7). Flight height had significant effects on droplet density and coverage rate at all tested spray volumes on the upper canopy. The highest droplet density and coverage rate were observed at a flight height of 2.5 m (Figures 7A, B). Additionally, at a flight height of 2.5 m, the spray volumes of 37.5 L·ha^{-1} and 45.0 L·ha^{-1} resulted in higher droplet density and coverage rate compared to 30.0 L·ha^{-1} (Figures 7A, B). For droplet size, only at a spray volume of 45.0 L·ha^{-1} did flight height show a significant effect. Specifically, droplet sizes at flight heights of 2.0 m and 2.5 m were significantly larger than those at 3.0 m (Figure 7C). Based on the above results, we assume that the optimal spray operation parameter combinations are a flight height of 2.5 m combined with spray volumes of 37.5 L·ha^{-1} and 45.0 L·ha^{-1} using the XR110015VS nozzle.

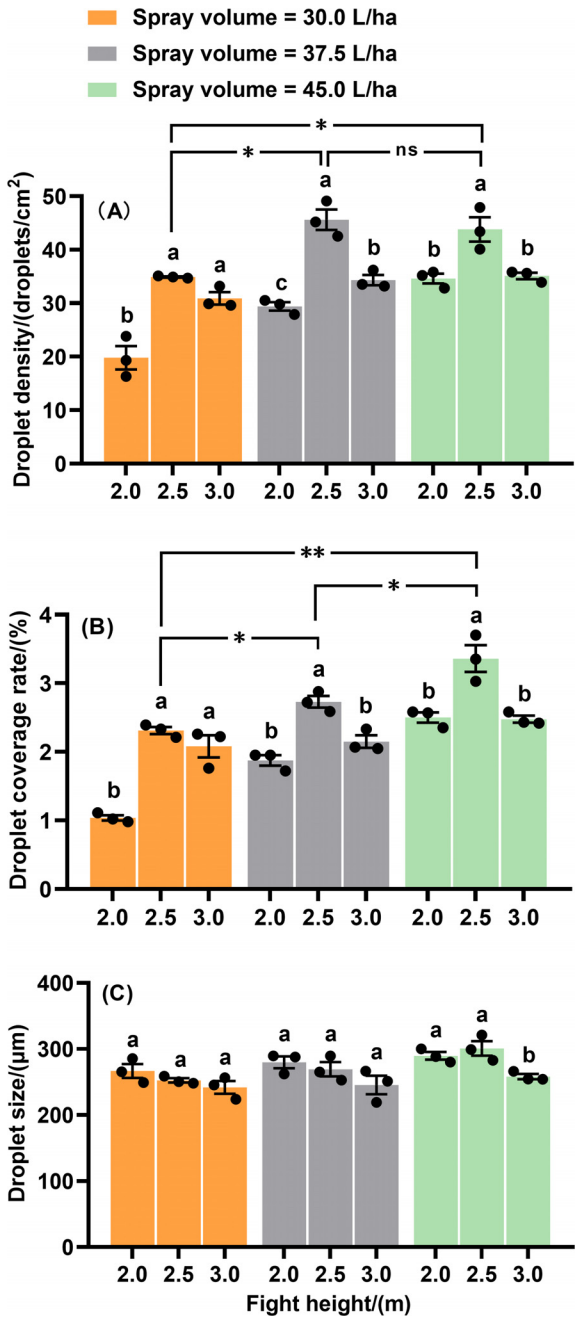


FIGURE 7
Droplet density **(A)**, droplet coverage rate **(B)** and droplet size **(C)** on the upper canopy of maize plant with nozzle XR110015VS at three flight heights. Different letters on bars indicate significant differences (P< 0.05) among flight heights. Asterisks above the bars denote significant difference of droplet density (droplet coverage rate) between spray volumes at the flight height of 2.5 m by using t-test. * and ** indicate significant difference at the significance levels of 0.05 and 0.01, respectively; ns represents no significance. Error bars denote standard error of the means.

3.3 A three-way factorial ANOVA analysis of spray operation parameters on droplet deposition characteristics on the upper canopy of maize plants

The three-way factorial ANOVA revealed distinct mechanistic influences of spray parameters on droplet deposition characteristics in maize upper canopies (Table 2). Nozzle type, spray volume, and flight height exhibited significant main effects on all measured metrics: droplet density, coverage rate, and droplet size. For droplet density, significant two-way interactions emerged between nozzle type × flight height and spray volume × flight height, whereas nozzle type × spray volume and the three-way interaction were non-significant. Coverage rate demonstrated greater interactive complexity, with all two-way interactions and

TABLE 2 Summary of factorial ANOVA for the effect of nozzle types, spray volumes and flight heights on droplet density, coverage rate and droplet size on the upper canopy of maize plant.

Source	DF	Droplet density (droplets·cm^{-2})		Coverage rate (%)		Droplet size (μm)	
		F	*P*	F	*P*	F	*P*
Nozzle type	1	224.2	<0.01	331.8	<0.01	51.8	<0.01
Spray volume	2	58.9	<0.01	56.2	<0.01	8.73	<0.01
Flight height	2	83.1	<0.01	62.6	<0.01	21.4	<0.01
Nozzle type × Spray volume	2	0.48	0.621	7.66	<0.01	2.45	0.101
Nozzle type × Flight height	2	6.59	<0.01	6.37	<0.01	0.36	0.704
Spray volume × Flight height	4	7.45	<0.01	8.89	<0.01	0.75	0.568
Nozzle type × Spray volume × Flight height	4	1.68	0.177	4.73	<0.01	0.51	0.729
Residual	36	–	–	–	–	–	–

the three-way interaction achieving significance. In contrast, droplet size was uniquely dependent on main effects, with no interactive terms contributing to variance.

3.4 Evaluation of control effects against fall armyworm

In 2019, for UAV applications at different flight heights, the Fall Armyworm (FAW) damage indices were consistently lower when using the nozzle XR110015VS compared to the nozzle XR11001VS at the same spray volumes both at 3 days after treatment (DAT) and 7 DAT (Figure 8). Additionally, treatments using the XR110015VS nozzle at a flight height of 2.5 meters resulted in lower damage indices than the other two flight height treatments for all three spray volumes. The results confirmed our previous hypothesis: the two optimal spray operation parameter combinations—using a flight height of 2.5 meters with spray volumes of 37.5 L·ha^{-1} or 45.0 L·ha^{-1}, and the XR110015VS nozzle—yielded the highest fog density and coverage on the upper canopy of maize and demonstrated the best control effects against FAW. EAP (electric air-pressure knapsack sprayer) application had the lowest damage indices (14.59 and 17.48, respectively) at 3 and 7 DAT. However, these values did not show significant differences compared to treatments with a flight height of 2.5 meters and spray volumes of 37.5 L·ha^{-1} (15.19 and 17.78) or 45.0 L·ha^{-1} (15.04 and 17.48). Similarly, based on FAW damage indices,

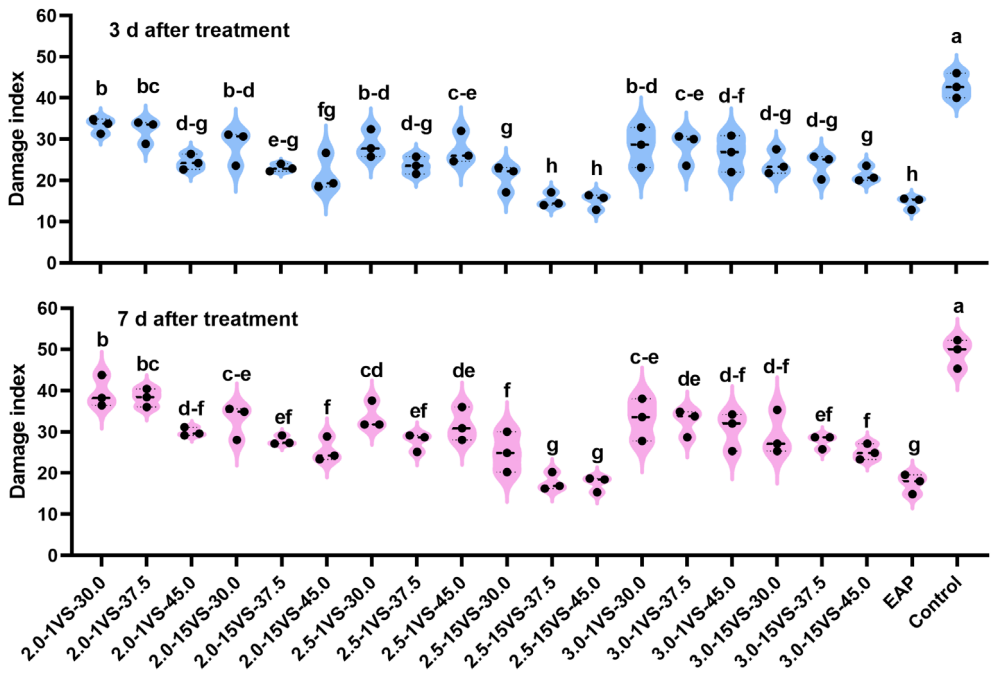


FIGURE 8
FAW damage index under different treatments in 2019. Different letters on bars indicate significant differences (P< 0.05) among treatments.

the control efficacies of EAP application (65.86% and 64.20%, Supplementary Table S3) did not differ significantly from the above two UAV treatments (64.63% and 63.90%; 65.02% and 64.51%). Similar results were observed in 2020 and 2021(Figures 9, 10). These findings indicate that optimizing UAV spray operation parameter can achieve comparable FAW control efficacy to traditional equipment. Notably, UAV application technology also offers high efficiency, water savings, and minimizes operator exposure to pesticide contact risks. In summary, farmers are advised to adopt optimized UAV parameters (XR110015VS nozzle, 37.5 or 45.0 L/ha spray volume, 2.5 m flight height) during the whorl stage for effective FAW control.

4 Discussion

Fall Armyworm (FAW), a devastating pest of maize, poses significant challenges due to its cryptic larval feeding within maize whorls, which limits insecticide efficacy. Enhancing droplet deposition in the whorl is crucial for improving control efficacy. Unmanned Aerial Vehicles (UAVs), with their downward airflow (wind field), offer advantages by facilitating droplet concentration and penetration into the whorl (Qin et al., 2016). However, varying spray operation parameters such as nozzle type, flight height, and spray volume significantly influence the UAV wind field, thereby affecting droplet deposition on plant canopies (Lou et al., 2018; Chen et al., 2020). This study evaluates these parameters to identify optimal combinations that maximize droplet density and coverage on the upper canopy, where FAW predominantly resides.

Nozzle selection directly affects droplet size distribution, which interacts synergistically with UAV wind field to determine deposition patterns (Fritz and Hoffmann, 2016; Tang et al., 2018). In this study, the XR110015VS nozzle, producing larger droplets, achieved higher droplet density and coverage rate on the upper maize canopy compared to the XR11001VS nozzle. This is consistent with previous findings that larger droplets result in better deposition and coverage on the upper rice canopy (Chen et al., 2020) and on wheat heads (Fritz et al., 2006). Spray volume also plays a key role in droplet deposition. For specific flight heights (2.0 m and 3.0 m), increasing spray volume improved droplet density and coverage on the upper canopy. Similar trends were observed in winter wheat (Shan et al., 2021) and rice (He et al., 2017). However, median spray volumes (37.5 L·ha^{-1}) outperformed both lower (30.0 L·ha^{-1}) and higher (45.0 L·ha^{-1}) volumes at a flight height of 2.5 m, aligning with prior studies (Fritz et al., 2006). Flight height is another critical parameter for achieving ideal droplet deposition. In this study, a height of 2.5 m yielded better results than 2.0 m or 3.0 m, consistent with Tang et al.'s (2018) findings in citrus canopies.

A three-way factorial ANOVA revealed that droplet density and coverage rate are synergistically modulated by parameter interactions, while droplet size remains independent of these effects, reflecting the different physical drivers governing droplet density/coverage rate versus droplet size. Significant interactions among parameters were also observed by Fritz et al. (2006) and Ferguson et al. (2016).

This study demonstrated that droplet deposition distribution varied significantly with different combinations of nozzle type, spray volume, and flight height, which in turn impacted FAW control efficacy. Precise targeting of the upper canopy significantly enhanced the contact rates of insecticides with FAW, thereby improving control efficacy. Optimal treatments using the XR110015VS nozzle at 2.5 m

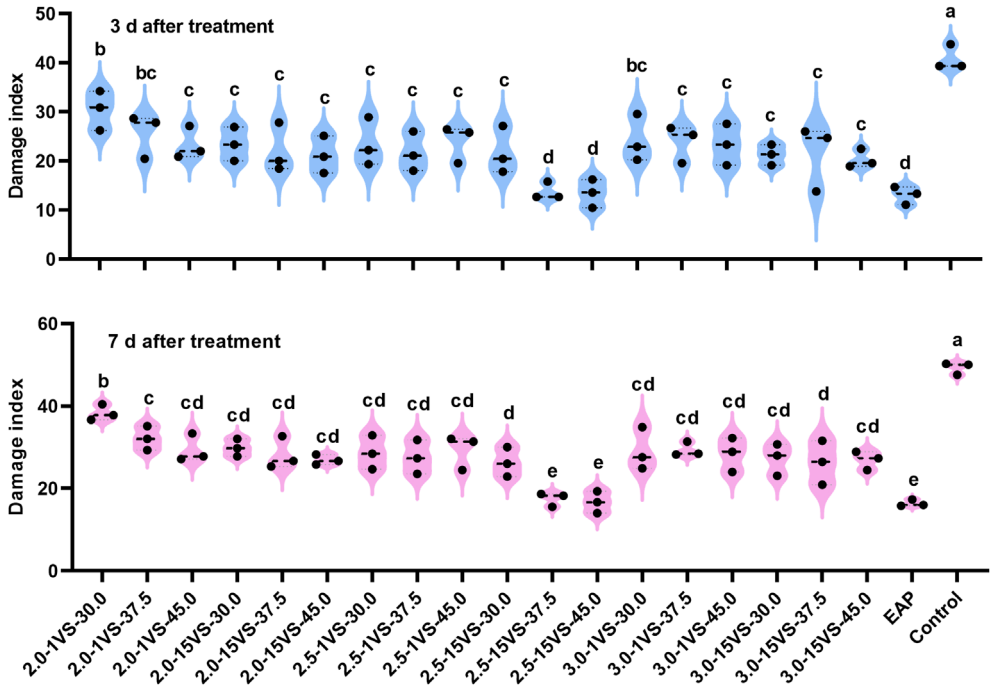


FIGURE 9
FAW damage index under different treatments in 2020. Different letters on bars indicate significant differences ($P< 0.05$) among treatments.

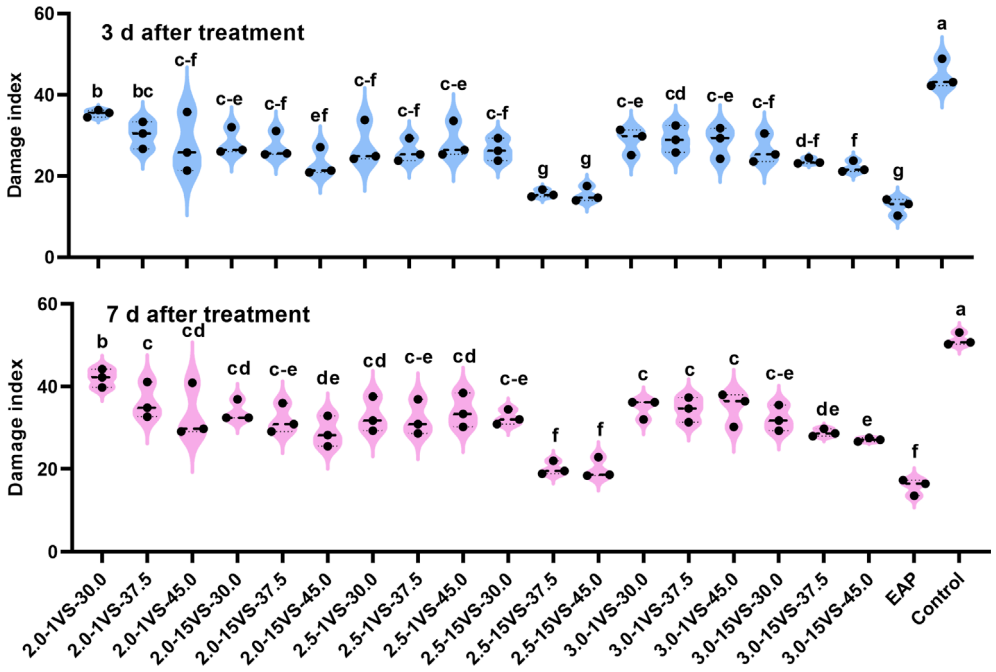


FIGURE 10
FAW damage index under different treatments in 2021. Different letters on bars indicate significant differences ($P < 0.05$) among treatments.

flight height with spray volumes of 37.5 L/ha or 45.0 L/ha achieved superior control effects. These findings align with studies using electric air-pressure knapsack sprayers, where targeting the whorl improved efficacy. Yang et al. (2020) evaluated the control effect of 10% tetrachlorantraniliprole SC by spraying different sites of the maize plant using an EPA sprayer. Their results showed that spraying only the whorl provided better control efficacy than spraying the whole plant with the same dosage. Similarly, Wang Y. et al. (2019) studied the control effect of 20% chlorantraniliprole SC against FAW using the same method and found that spraying half the dose on the maize whorl achieved equivalent control efficacy compared to full-dose application on the whole plant.

While these results are significant for UAV-based FAW control in maize at the whorl stage, maize canopy structure varies across growth stages, influencing droplet deposition patterns. Future work will focus on stage-specific optimization of UAV parameters to ensure consistent FAW suppression while minimizing environmental impact.

5 Conclusions

Optimized UAV applications using the XR110015VS nozzle with spray volumes of 37.5 L/ha or 45.0 L/ha at a flight height of 2.5 m achieved FAW control efficacy comparable to traditional knapsack sprayers while reducing water usage by 90%, significantly lowering operational costs and environmental impact, and improving safety for operators. Farmers are advised to adopt these parameters during the whorl stage to effectively control FAW. Future work should focus on adjusting UAV settings to address FAW infestations at different growth stages.

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

YL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. XL: Investigation, Writing – review & editing. CW: Investigation, Writing – review & editing. XA: Data curation, Writing – review & editing. MW: Investigation, Writing – review & editing. ZZ: Data curation, Writing – review & editing. ZL: Writing – review & editing. QC: Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by the Hainan Major Science and Technology Project (No. ZDKJ202002) and the earmarked fund for CARS-02-32.

Conflict of interest

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

Generative AI statement

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1581367/full#supplementary-material

References

Bateman, M. L., Day, R. K., Luke, B., Edgington, S., Kuhlmann, U., and Cock, M. J. W. (2018). Assessment of potential biopesticide options for managing fall armyworm (*Spodoptera frugiperda*) in Africa. *J. Appl. Entomol.* 142, 805–819. doi: 10.1111/jen.12565

Boaventura, D., Bolzan, A., Padovez, F. E. O., Okuma, D. M., Omoto, C., and Nauen, R. (2020). Detection of a ryanodine receptor target-site mutation in diamide insecticide resistant fall armyworm, *Spodoptera frugiperda*. *Pest Manag. Sci.* 76, 47–54. doi: 10.1002/ps.5505

Chen, S., Lan, Y., Zhou, Z., Ouyang, F., Wang, G., Huang, X., et al. (2020). Effect of droplet size parameters on droplet deposition and drift of aerial spraying by using plant protection UAV. *Agronomy* 10, 195. doi: 10.3390/agronomy10020195

Chinwada, P., Fiaboe, K. K. M., Akem, C., Dixon, A., and Chikoye, D. (2023). Assessment of effectiveness of maize seed treated with cyantraniliprole and thiamethoxam for management of fall armyworm, *Spodoptera frugiperda* (J.E. Smith). *Crop Prot.* 174, 106418. doi: 10.1016/j.cropro.2023.106418

Ferguson, J. C., Chechetto, R. G., Hewitt, A. J., Chauhan, B. S., Adkins, S. W., Kruger, G. R., et al. (2016). Assessing the deposition and canopy penetration of nozzles with different spray qualities in an oat (*Avena sativa* L.) canopy. *Crop Prot.* 81, 14–19. doi: 10.1016/j.cropro.2015.11.013

Fritz, B. K., and Hoffmann, W. C. (2016). Measuring spray droplet size from agricultural nozzles using laser diffraction. *J. Vis. Exp.* 115, e54533. doi: 10.3791/54533

Fritz, B. K., Kirk, I. W., Hoffmann, W. C., Martin, D. E., Hofman, V. L., Hollingsworth, C., et al. (2006). Aerial application methods for increasing spray deposition on wheat heads. *Appl. Eng. Agric.* 22, 357–364. doi: 10.13031/2013.20453

Goergen, G., Kumar, P. L., Sankung, S. B., Togola, A., and Tamò, M. (2016). First report of outbreaks of the fall armyworm *spodoptera frugiperda* (J E smith) (Lepidoptera, noctuidae), a new alien invasive pest in west and central Africa. *PloS One* 11, e0165632. doi: 10.1371/journal.pone.0165632

Guo, H., Zhou, J., Liu, F., He, Y., Huang, H., and Wang, H. (2020). Application of machine learning method to quantitatively evaluate the droplet size and deposition distribution of the UAV spray nozzle. *Appl. Sci.* 10, 1759. doi: 10.3390/app10051759

He, L., Wang, G., Hu, T., Meng, Y., Yan, X., and Yuan, H. (2017). Influences of spray adjuvants and spray volume on the droplet deposition distribution with unmanned aerial vehicle (UAV) spraying on rice. *J. Plant Prot.* 44, 1046–1105. doi: 10.13802/j.cnki.zwbhxb.2017.2016147

Kansiime, M. K., Mugambi, I., Rwomushana, I., Nunda, W., Lamontagne-Godwin, J., Rware, H., et al. (2019). Farmer perception of fall armyworm (*Spodoptera frugiderda* J.E. Smith) and farm-level management practices in Zambia. *Pest Manag. Sci.* 75, 2840–2850. doi: 10.1002/ps.5504

Kumar, R. M., Gadratagi, B.-G., Paramesh, V., Kumar, P., Madivalar, Y., Narayanappa, N., et al. (2022). Sustainable management of invasive fall armyworm, *spodoptera frugiperda*. *Agronomy* 12, 2150. doi: 10.3390/agronomy12092150

Lou, Z., Xin, F., Han, X., Lan, Y., Duan, T., and Fu, W. (2018). Effect of unmanned aerial vehicle flight height on droplet distribution, drift and control of cotton aphids and spider mites. *Agronomy* 8, 187. doi: 10.3390/agronomy8090187

Nagoshi, R. N., Goergen, G., Plessis, H. D., van den Berg, J., and Meagher, R. (2019). Genetic comparisons of fall armyworm populations from 11 countries spanning sub-Saharan Africa provide insights into strain composition and migratory behaviors. *Sci. Rep-Uk* 9, 8311. doi: 10.1038/s41598-019-44744-9

Nagoshi, R. N., Meagher, R. L., and Hay-Roe, M. (2012). Inferring the annual migration patterns of fall armyworm (Lepidoptera: Noctuidae) in the United States from mitochondrial haplotypes. *Ecol. Evol.* 2, 1458–1467. doi: 10.1002/ece3.268

Qin, W., Qiu, B., Xue, X., Chen, C., Xu, Z., and Zhou, Q. (2016). Droplet deposition and control effect of insecticides sprayed with an unmanned aerial vehicle against plant hoppers. *Crop Prot.* 85, 79–88. doi: 10.1016/j.cropro.2016.03.018

Ren, J. X., Li, Z. K., Tao, C. J., Zhang, L. Y., Zhao, H. F., Wu, C. C., et al. (2019). Exposure assessment of operators to clothianidin when using knapsack electric sprayers in greenhouses. *Int. J. Environ. Sci.* 16, 1471–1478. doi: 10.1007/s13762-018-1758-z

Shan, C., Wang, G., Wang, H., Xie, Y., and Lan, Y. (2021). Effects of droplet size and spray volume parameters on droplet deposition of wheat herbicide application by using UAV. *Int. J. Agr. Biol. Eng.* 14, 74–81. doi: 10.25165/j.ijabe.20211401.6129

Shan, C., Wu, J., Song, C., Chen, S., Wang, J., Wang, H., et al. (2022). Control efficacy and deposition characteristics of an unmanned aerial spray system low-volume application on corn fall armyworm *spodoptera frugiperda*. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.900939

Tang, Y., Hou, C., Luo, S., Lin, J., Yang, Z., and Huang, W. (2018). Effects of operation height and tree shape on droplet deposition in citrus trees using an unmanned aerial vehicle. *Comput. Electr. Agr.* 148, 1–7. doi: 10.1016/j.compag.2018.02.026

Toepfer, S., Fallet, P., Kajuga, J., Bazagwira, D., Mukundwa, I. P., Szalai, M., et al. (2021). Streamlining leaf damage rating scales for the fall armyworm on maize. *J. Pest Sci.* 94, 1075–1089. doi: 10.1007/s10340-021-01359-2

Wang, X., He, X., Song, J., Wang, Z., Wang, C., Wang, S., et al. (2018). Drift potential of UAV with adjuvants in aerial applications. *Int. J. Agr. Biol. Eng.* 11, 54–58. doi: 10.25165/j.ijabe.20181105.3185

Wang, G., Lan, Y., Qi, H., Chen, P., Hewitt, A., and Han, Y. (2019). Field evaluation of an unmanned aerial vehicle (UAV) sprayer: effect of spray volume on deposition and the control of pests and disease in wheat. *Pest Manag. Sci.* 75, 1546–1555. doi: 10.1002/ps.5321

Wang, Y., Ma, Q., Tan, Y., Zheng, Q., Yan, W., Yang, S., et al. (2019). The toxicity and field efficacy of chlorantraniliprole against *Spodoptera frugiperda*. *J. Environ. Entomol.* 41, 782–788. doi: 10.3969/j.issn.1674-0858.2019.04.14

Westbrook, J. K., Nagoshi, R. N., Meagher, R. L., Fleischer, S. J., and Jairam, S. (2016). Modeling seasonal migration of fall armyworm moths. *Int. J. Biometeorol.* 60, 255–267. doi: 10.1007/s00484-015-1022-x

Yang, S., Yan, W., Tan, Y., Wang, Y., Zheng, Q., Zhang, Z., et al. (2020). Evaluation of control effect of tetrachlorantraniliprole corn whole plant spray and flare opening application treatment on *Spodoptera frugiperda*. *J. Environ. Entomol.* 42, 76–81. doi: 10.3969/j.issn.1674-0858.2020.01.10

Yang, S., Yang, X., and Mo, J. (2018). The application of unmanned aircraft systems to plant protection in China. *Precis. Agric.* 19, 278–292. doi: 10.1007/s11119-017-9516-7

Yu, S.-H., Kim, Y.-K., Jun, H.-J., Choi, I. S., Woo, J.-K., Kim, Y.-H., et al. (2020). Evaluation of spray characteristics of pesticide injection system in agricultural drones. *J. Biosyst. Eng.* 45, 272–280. doi: 10.1007/s42853-020-00067-6

Zhang, Y., Li, Y., He, Y., Liu, F., Cen, H., and Fang, H. (2018). Near ground platform development to simulate UAV aerial spraying and its spraying test under different conditions. *Comput. Electr. Agr.* 148, 8–18. doi: 10.1016/j.compag.2017.08.004

Zhu, H., Salyani, M., and Fox, R. D. (2011). A portable scanning system for evaluation of spray deposit distribution. *Comput. Electr. Agr.* 76, 38–43. doi: 10.1016/j.compag.2011.01.003

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 14 May 2025
DOI 10.3389/fpls.2025.1576756

Check for updates

OPEN ACCESS

EDITED BY
Jun Ni,
Nanjing Agricultural University, China

REVIEWED BY
Olumide Falana,
Obafemi Awolowo University, Nigeria
Balaji V R,
Sri Krishna College of Engineering & Technology, India
Danish Gul,
Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, India

*CORRESPONDENCE
Cheng Wei
chw365@163.com

RECEIVED 14 February 2025
ACCEPTED 07 April 2025
PUBLISHED 14 May 2025

CITATION
Wei C, Shan Y and Zhen M (2025) Deep learning-based anomaly detection for precision field crop protection.
Front. Plant Sci. 16:1576756.
doi: 10.3389/fpls.2025.1576756

COPYRIGHT
© 2025 Wei, Shan and Zhen. 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.

Deep learning-based anomaly detection for precision field crop protection

Cheng Wei[1]*, Yifeng Shan[2] and MengZhe Zhen[3]

[1]Chongqing Business Vocational College, Chongqing, China, [2]Ningbo University of Finance and Economics, School of Basic Education, Ningbo, Zhejiang, China, [3]Ningbo University of Finance and Economics, School of Digital Technology and Engineering, Ningbo, Zhejiang, China

Introduction: Precision agriculture relies on advanced technologies to optimize crop protection and resource utilization, ensuring sustainable and efficient farming practices. Anomaly detection plays a critical role in identifying and addressing irregularities, such as pest outbreaks, disease spread, or nutrient deficiencies, that can negatively impact yield. Traditional methods struggle with the complexity and variability of agricultural data collected from diverse sources.

Methods: To address these challenges, we propose a novel framework that integrates the Integrated Multi-Modal Smart Farming Network (IMSFNet) with the Adaptive Resource Optimization Strategy (AROS). IMSFNet employs multimodal data fusion and spatiotemporal modeling to provide accurate predictions of crop health and yield anomalies by leveraging data from UAVs, satellites, ground sensors, and weather stations. AROS dynamically optimizes resource allocation based on real-time environmental feedback and multi-objective optimization, balancing yield maximization, cost efficiency, and environmental sustainability.

Results: Experimental evaluations demonstrate the effectiveness of our approach in detecting anomalies and improving decision-making in precision agriculture.

Discussion: This framework sets a new standard for sustainable and data-driven crop protection strategies.

KEYWORDS

precision agriculture, anomaly detection, multi-modal data fusion, resource optimization, sustainable farming

1 Introduction

Precision agriculture has revolutionized the agricultural industry, enabling efficient and sustainable crop management through targeted interventions. Within this domain, anomaly detection plays a critical role in field crop protection by identifying early signs of diseases, pests, nutrient deficiencies, or other stress factors that compromise crop health

Zhang et al. (2024b). Not only does early detection help reduce the overuse of chemical inputs such as pesticides and fertilizers, but it also minimizes yield losses and supports sustainable farming practices Gao et al. (2024a). Traditional anomaly detection techniques, while effective in controlled conditions, often fail to capture the complexity of real-world agricultural systems, which involve heterogeneous data sources, dynamic environmental conditions, and intricate interactions between crops and external stressors Zhu et al. (2024a). As a result, deep learning-based methods have emerged as a promising solution for improving the accuracy and scalability of anomaly detection in precision field crop protection. By leveraging multi-modal data from satellite imagery, drones, and on-ground sensors, these methods provide an end-to-end framework for identifying and addressing crop health anomalies Li et al. (2021).

Early approaches to anomaly detection in agriculture were based on rule-based systems and statistical models, which relied on domain knowledge and handcrafted features to identify deviations from normal conditions Zavrtanik et al. (2021). For example, threshold-based methods used predefined values for parameters like vegetation indices or soil moisture levels to detect anomalies Deng and Hooi (2021). Similarly, statistical models such as principal component analysis (PCA) and clustering were used to identify outliers in crop health data Zou et al. (2022). While these methods provided interpretable results and were computationally efficient, they lacked the ability to generalize across diverse environmental conditions and crop types Bergmann et al. (2021). Furthermore, their reliance on fixed thresholds and handcrafted features made them inadequate for capturing the complex, non-linear patterns associated with crop health anomalies caused by diseases or pests Gudovskiy et al. (2021).

The shift toward data-driven approaches introduced machine learning algorithms capable of learning patterns from historical data to improve anomaly detection You et al. (2022). Techniques such as support vector machines (SVMs), random forests, and k-nearest neighbors were employed to classify crop health statuses based on features extracted from remote sensing or field data Liu et al. (2023a). These models achieved better adaptability compared to rule-based systems, as they could learn relationships between input features and anomalies without explicit thresholds. For example, machine learning models were applied to identify plant stress from hyperspectral imagery or to classify pest infestations based on soil and weather data Zhu et al. (2024b). However, these approaches faced limitations in their ability to handle highdimensional and multi-modal data, such as the integration of spectral, spatial, and temporal information. Traditional machine learning methods required extensive feature engineering and struggled to generalize to new datasets or unseen conditions Tian et al. (2021).

Deep learning has transformed anomaly detection in precision agriculture by introducing architectures that can automatically learn hierarchical representations of crop health data Han et al. (2022). Convolutional neural networks (CNNs) have been widely used for analyzing spatial patterns in satellite and drone imagery, enabling the detection of disease outbreaks, pest infestations, and nutrient deficiencies. Recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have been employed to model temporal changes in crop health, such as monitoring vegetation growth over a growing season Jiang et al. (2023). Furthermore, generative adversarial networks (GANs) and autoencoders have demonstrated success in unsupervised anomaly detection, where models are trained on normal data to identify deviations that signify anomalies Zhang et al. (2024c). For example, autoencoders have been used to detect stress signals in hyperspectral images by reconstructing healthy crop patterns and flagging deviations as anomalies Tien et al. (2023). Despite these advancements, challenges remain, particularly in the integration of multi-modal data from diverse sources, the interpretability of deep learning models, and the scalability of these methods for large-scale agricultural applications Wyatt et al. (2022).

Recent works in spatiotemporal modeling for agricultural anomaly detection primarily rely on CNNs and LSTMs to process spatial and sequential data, respectively. For example, proposed a hybrid CNN-LSTM model Xu et al. (2021) to capture vegetation dynamics from satellite images, but their approach struggled with long-range dependencies and multimodal data integration. Similarly, introduced an attention-enhanced LSTM Tuli et al. (2022) for pest detection, yet the model's performance degraded with increasing data heterogeneity. In contrast, our proposed IMSFNet incorporates GNNs to model spatial relationships between crop regions, while employing Transformers to capture long-term temporal dependencies more effectively than LSTMs. This hybrid architecture allows for adaptive weighting of multimodal information, improving the robustness and interpretability of anomaly detection in precision agriculture.

To address these challenges, we propose a novel deep learning-based framework for anomaly detection tailored for precision field crop protection. The proposed framework integrates multi-modal data from satellite imagery, drones, and *in-situ* sensors to capture a holistic view of crop health. A hybrid architecture combining convolutional neural networks and graph neural networks (GNNs) is employed to model spatial dependencies between crops, while transformers are utilized to capture temporal patterns in crop growth and environmental conditions. The framework incorporates unsupervised learning techniques, such as variational autoencoders, to detect subtle anomalies that may not be present in labeled datasets. By prioritizing scalability and interpretability, this framework is designed to support real-time decision-making and adaptive interventions, ultimately enhancing crop resilience and productivity.

We summarize our contributions as follows:

- We propose IMSFNet, a novel Integrated Multi-Modal Smart Farming Network, which combines GNNs for spatial dependency modeling and Transformers for long-term temporal feature extraction. This hybrid architecture effectively captures crop health variations and environmental anomalies across different spatial and temporal scales.
- We introduce a multi-modal fusion strategy that integrates satellite imagery, UAV-based imaging, ground sensors, and meteorological data. Unlike previous works that rely on independent processing pipelines, IMSFNet jointly learns representations from heterogeneous data sources, leading to improved anomaly detection performance.

- We develop AROS (Adaptive Resource Optimization Strategy), a real-time optimization framework that dynamically adjusts resource allocation based on multi-objective optimization. AROS leverages reinforcement learning-based feedback mechanisms to improve efficiency and sustainability in precision agriculture.
- Our extensive experiments on multiple real-world datasets demonstrate that IMSFNet and AROS achieve 3.12% higher F1-score and 2.89% higher accuracy compared to state-of-the-art anomaly detection models. We release our implementation and dataset annotations to facilitate further research in deep learning-based precision agriculture.

2 Related work

2.1 Deep learning for anomaly detection in agriculture

Anomaly detection is a critical component of precision agriculture, aiming to identify abnormal patterns in crop health, pest infestations, and environmental conditions Wang et al. (2023a). Traditional anomaly detection methods, such as threshold-based techniques and classical machine learning models like Support Vector Machines (SVMs) and k-Nearest Neighbors (k-NN), often fail to capture the complexity of agricultural environments Wang et al. (2023b). Deep learning approaches have emerged as more robust alternatives, leveraging their ability to process large-scale, high-dimensional data and uncover complex, nonlinear relationships Defard et al. (2020). In agricultural applications, CNNs are widely employed to analyze visual data, such as images from drones and ground-based cameras Gao et al. (2024b). For instance, CNN-based models can detect visual anomalies in crops, such as discoloration, irregular growth, and pest damage Park et al. (2020). RNNs and LSTM networks, on the other hand, are applied to sequential data, such as time-series environmental sensor readings, to identify trends or deviations indicative of anomalies DeMedeiros et al. (2023). Recent advancements include hybrid models that integrate CNNs and RNNs to analyze spatiotemporal data, such as videos of crop fields over time Batzner et al. (2023). These models enable the detection of anomalies not just at a single point in time but also in evolving patterns, such as the spread of a disease across a field Feng et al. (2021). While these methods show promise, challenges remain in dealing with noisy and imbalanced data, where anomalies constitute only a small fraction of the dataset. Techniques such as data augmentation, synthetic anomaly generation, and adversarial training are being developed to address these issues and enhance model robustness in real-world scenarios Audibert et al. (2020).

2.2 Multi-modal data fusion for precision crop protection

Precision crop protection often requires integrating data from multiple sources, such as remote sensing, IoT devices, and weather stations, to effectively detect and respond to anomalies Le and Zhang (2021). Multi-modal data fusion enables the combination of these diverse data types to improve the accuracy and reliability of anomaly detection systems Liu et al. (2021). Deep learning has been instrumental in facilitating such fusion, with models that can process and integrate heterogeneous data modalities. For visual data, drone and satellite imagery are commonly used to monitor crop health, while IoT devices provide real-time sensor readings for soil moisture, temperature, and humidity Salehi et al. (2020). Multi-modal architectures, such as those combining CNNs for image analysis with fully connected or Transformer layers for sensor data, have demonstrated improved performance in identifying anomalies such as nutrient deficiencies, water stress, and pest outbreaks Liu et al. (2023b). For example, attention mechanisms have been employed to prioritize the most relevant data sources for decision-making, enhancing the interpretability of these systems. Temporal fusion models, such as Temporal Fusion Transformers (TFTs), have also been applied to integrate time-series data from multiple sensors with historical climate records, enabling more accurate anomaly predictions Roth et al. (2021). GNNs are another emerging approach for multi-modal fusion, particularly in representing spatial relationships within a field, such as proximity between affected regions or the spread of an anomaly across neighboring plots. Despite the benefits, multi-modal fusion faces challenges such as data heterogeneity, missing values, and high computational requirements. Advances in self-supervised learning and imputation techniques are being explored to address these limitations, enabling models to learn meaningful representations from incomplete or noisy datasets Deng and Li (2022). Furthermore, edge computing and hardware acceleration are being investigated to enable real-time data fusion and anomaly detection in resource-constrained agricultural environments.

2.3 Applications of anomaly detection in crop protection

Deep learning-based anomaly detection has found numerous applications in precision crop protection, addressing challenges such as pest infestations, disease outbreaks, and abiotic stress factors like drought and frost Ni et al. (2018). By automating the identification of anomalies, these systems reduce the reliance on manual scouting, which is labor-intensive and prone to errors, especially in large-scale agricultural settings Ni et al. (2017). For pest detection, models leveraging CNNs and object detection frameworks like YOLO (You Only Look Once) have been used to identify specific pest species in field images. These models enable targeted interventions, such as pesticide application, reducing chemical usage and minimizing environmental impact Mishra et al. (2021). Similarly, for disease detection, segmentation models such as U-Net and Mask R-CNN have been employed to localize affected areas, allowing for precise treatment. Beyond visual data, deep learning models have been applied to sensor-based anomaly detection. For example, LSTM networks are used to analyze soil moisture and temperature data to identify water stress, while

Variational Autoencoders (VAEs) Ni et al. (2016) and GANs are employed to detect deviations from normal patterns in multi-dimensional sensor data. These approaches are particularly effective in identifying early warning signs of crop stress, enabling timely interventions that mitigate yield losses. Another application lies in predicting the spatial spread of anomalies, such as the propagation of pests or diseases within a field. Spatiotemporal models, including 3D CNNs and ST-GCNs (Spatio-Temporal Graph Convolutional Networks), are used to predict how anomalies evolve over time and space, aiding in the design of containment strategies. For instance, these models can simulate the effects of varying weather conditions on the spread of an anomaly, providing actionable insights for farmers and agronomists. While these applications demonstrate the potential of anomaly detection in crop protection, practical deployment remains challenging due to factors such as limited labeled data, variability in agricultural environments, and the need for domain-specific customization. Future research should focus on developing scalable and generalizable models, incorporating domain knowledge into deep learning frameworks, and ensuring the ethical use of these technologies in precision agriculture.

3 Method

3.1 Overview

Precision agriculture, also referred to as smart farming or digital farming, is an advanced approach to agricultural management that leverages modern technologies such as remote sensing, geographic information systems (GIS), Internet of Things (IoT), and artificial intelligence (AI) to optimize crop production and resource utilization. The goal of precision agriculture is to enhance efficiency, reduce environmental impact, and increase yield through the precise monitoring and management of agricultural inputs, including water, fertilizers, and pesticides. Traditional farming practices often involve uniform treatment of large agricultural fields, which can lead to inefficiencies and overuse of resources. By contrast, precision agriculture adopts a data-driven approach, tailoring interventions to the specific needs of crops, soil conditions, and environmental factors. This is achieved through a combination of advanced technologies that collect, analyze, and act upon data at a granular level. For instance, sensors embedded in the soil can measure moisture and nutrient levels, drones equipped with multispectral cameras can capture crop health data, and machine learning algorithms can predict optimal planting and harvesting times. The implementation of precision agriculture can be broadly categorized into three main components: data collection, data analysis, and decision-making. Data collection involves the use of various sensing technologies, such as satellite imagery, UAVs (unmanned aerial vehicles), and IoT-enabled sensors, to capture real-time information about crops and the environment. This data is then analyzed using advanced computational techniques, including machine learning and statistical modeling, to derive actionable insights. The insights are used to make informed decisions, such as variable-rate application of fertilizers or automated irrigation scheduling. Despite its promising potential, the widespread adoption of precision agriculture faces several challenges, including the high cost of technology, lack of technical expertise among farmers, and limited access to high-speed internet in rural areas. Addressing these barriers requires interdisciplinary efforts that integrate engineering, computer science, agronomy, and socioeconomics.

This paper introduces a novel framework for precision agriculture that integrates advanced sensing technologies with deep learning-based predictive models to enhance the scalability and robustness of smart farming systems. The proposed method focuses on multi-modal data fusion, combining visual, thermal, and spectral data from UAVs and ground-based sensors to improve the accuracy of crop health assessment and yield prediction. The framework also incorporates adaptive optimization strategies to address varying environmental and climatic conditions, ensuring its applicability across diverse agricultural contexts. The remainder of this paper is organized as follows. In Section 3.2, we formalize the precision agriculture problem, introducing the necessary mathematical foundations and notations. Section 3.3 introduces our proposed Integrated Multi-Modal Smart Farming Network (IMSFNet), a system that integrates advanced sensing and modeling techniques for efficient crop monitoring. In Section 3.4, we present the Adaptive Resource Optimization Strategy (AROS), a novel approach designed to optimize resource allocation through real-time predictions and environmental feedback.

To enhance readability, we briefly summarize the key technical abbreviations (In Table 1) used throughout the manuscript. IMSFNet refers to the proposed Integrated Multi-Modal Smart Farming Network, while AROS denotes the Adaptive Resource Optimization Strategy. Core components include MFE (Multi-Modal Feature Extraction), CBS (Convolution + Batch Normalization + SiLU), CSPBlock (Cross Stage Partial Block), and PConv (Parametric Convolution). FFCA-YOLO and L-FFCA-YOLO represent backbone architectures with cross-attentive and lightweight designs. Additional modules such as SCNN (Spatial Convolutional Neural Network), DMO (Dynamic Multi-Objective Optimization), PRA (Prioritized Resource Allocation), and RFM (Real-Time Feedback Mechanism) further support resource-efficient anomaly detection. Abbreviations like UAV (Unmanned Aerial Vehicle), NDVI (Normalized Difference Vegetation Index), GNN (Graph Neural Network), and LSTM (Long Short-Term Memory) are used to represent standard sensing and modeling technologies in precision agriculture.

3.2 Preliminaries

Precision agriculture relies on detecting anomalies in crop health to optimize resource allocation. Given an agricultural field F partitioned into N management zones $\{Z_1, Z_2 \ldots, Z_N\}$, each zone is represented by a feature vector $x_i$derived from multi-modal data sources. The objective of anomaly detection is to identify zones where the observed feature vector $x_i$deviates significantly from

TABLE 1 List of technical abbreviations used throughout the manuscript.

Abbreviation	Full Term
IMSFNet	Integrated Multi-Modal Smart Farming Network
AROS	Adaptive Resource Optimization Strategy
MFE	Multi-Modal Feature Extraction
CBS	Convolution + Batch Normalization + SiLU
CSPBlock	Cross Stage Partial Block
CSPFaster Block	Enhanced Cross Stage Partial Block (optimized for speed)
PConv	Parametric Convolution
FFCA-YOLO	Feature-Focused Cross-Attentive YOLO
L-FFCA-YOLO	Lightweight FFCA-YOLO
SCNN	Spatial Convolutional Neural Network
DMO	Dynamic Multi-Objective Optimization
PRA	Prioritized Resource Allocation
RFM	Real-Time Feedback Mechanism
NDVI	Normalized Difference Vegetation Index
IoT	Internet of Things
UAV	Unmanned Aerial Vehicle
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
GAN	Generative Adversarial Network
VAE	Variational Autoencoder
GNN	Graph Neural Network
CNN	Convolutional Neural Network
TFT	Temporal Fusion Transformer
Q, K, V	Query, Key, Value (in attention mechanism)

normal patterns. Let $p(x)$ represent the probability distribution of normal crop health conditions. A zone $Z_i$is classified as anomalous if its feature vector satisfies (Equation 1):

$$\mathbb{P}(x_i|\theta) < \tau, \tag{1}$$

where θ denotes the parameters of the learned normal distribution, and τ is a predefined anomaly threshold. To quantify anomalies, an anomaly score function s_i is defined based on the deviation of x_i from the mean feature vector μ of normal samples (Equation 2):

$$s_i = D(x_i, \mu), \tag{2}$$

where $D(\cdot,\cdot)$ represents a distance metric such as the Mahalanobis distance or Euclidean distance. Higher values of $s_i$indicate stronger deviations from normal conditions, guiding adaptive interventions in precision agriculture.

The optimization objective can be formulated as (Equation 3):

$$\max_{\mathcal{U}} \sum_{i=1}^{N} \mathcal{Y}_i(\mathcal{U}_i, \mathbf{p}_i) - \lambda \mathcal{C}(\mathcal{U}), \tag{3}$$

where $\mathcal{Y}_i$ is the yield function for zone Z_i, which depends on the intervention $\mathcal{U}_i$ and the feature vector $\mathbf{p}_i$, and $\mathcal{C}(\mathcal{U})$ is the total cost function associated with the intervention strategy $\mathcal{U}$. The parameter λ is a weighting factor that balances yield maximization and cost minimization.

Agricultural fields exhibit both spatial and temporal variability. Spatial variability arises from heterogeneities in soil properties, topography, and crop health across zones. Temporal variability is driven by dynamic environmental conditions, such as weather changes and seasonal cycles. Let $\mathbf{P}(t) = \{\mathbf{p}_1(t), \mathbf{p}_2(t),\ldots, \mathbf{p}_N(t)\}$ denote the time-dependent feature matrix for the field at time t. The evolution of crop yield $\mathcal{Y}_i(t)$ for zone Z_i can be modeled as (Equation 4):

$$\mathcal{Y}_i(t) = \mathcal{F}(\mathbf{p}_i(t), \mathcal{U}_i(t), \eta_i(t)), \tag{4}$$

where $\mathcal{F}(\cdot)$ is a function that encapsulates the complex relationships between environmental factors, interventions, and crop response, and $\boldsymbol{\eta}_i(t)$ represents stochastic disturbances such as pests, diseases, or unpredicted weather events.

Precision agriculture integrates diverse types of multimodal data collected from various sensing technologies to effectively monitor crop and environmental conditions. Satellite imagery enables large-scale observation by providing vegetation indices, such as the Normalized Difference Vegetation Index (NDVI), and information about canopy cover. Unmanned aerial vehicles (UAVs), commonly referred to as drones, capture high-resolution visual, thermal, and multispectral images, offering detailed insights into crop health. Ground sensors contribute precise measurements of soil properties, including moisture, temperature, pH, and electrical conductivity, at specific locations within the field. Weather stations supply real-time data on environmental variables, including temperature, humidity, wind speed, and precipitation, further enhancing the decision-making process in precision agriculture.

Let $\mathcal{D} = \{\mathcal{D}_{\text{sat}}, \mathcal{D}_{\text{uav}}, \mathcal{D}_{\text{ground}}, \mathcal{D}_{\text{weather}}\}$ represent the collection of data from these sources. The fusion of multi-modal data is essential for developing a comprehensive understanding of field conditions.

In precision agriculture, decision-making often involves multiple competing objectives, such as maximizing yield, minimizing resource usage, and reducing environmental impact. The problem can be formulated as a multi-objective optimization (Equation 5):

$$\textit{extmaximize} \quad [\mathcal{Y}(\mathcal{U}), -\mathcal{C}(\mathcal{U}), -\mathcal{E}(\mathcal{U})], \tag{5}$$

where $\mathcal{E}(\mathcal{U})$ represents the environmental impact function. Multi-objective optimization techniques, such as Pareto front analysis or weighted sum methods, are employed to identify trade-offs and select optimal strategies.

3.3 Integrated multi-modal smart farming network

IMSFNet is an innovative deep learning framework designed for precision agriculture, seamlessly integrating multi-modal data sources to perform anomaly detection in a unified manner. Unlike conventional CNN-LSTM architectures, it leverages Graph Neural Networks to enhance spatial feature learning while incorporating Transformers to capture long-range temporal dependencies. This combination significantly improves the model's ability to understand complex environmental interactions. The framework operates through a structured process that begins with extracting features from multiple data modalities, followed by an advanced fusion mechanism that integrates spatial and temporal information. It introduces a cross-modal interaction approach that strengthens the relationships between different data sources, ensuring a more comprehensive and accurate analysis of agricultural conditions.

IMSFNet integrates multiple data sources to achieve precise crop anomaly detection through multi-modal fusion, as illustrated in Figure 1. Each modality contributes a distinct feature set, including satellite imagery features represented as $X_{\text{sat}} \in \mathbb{R}^{H_s \times W_s \times d_s}$, UAV imagery features as $X_{\text{uav}} \in \mathbb{R}^{H_u \times W_u \times d_u}$, ground sensor features as $X_{\text{sens}} \in \mathbb{R}^{N_g \times d_g}$, and weather data as $X_{\text{weather}} \in \mathbb{R}^{T_w \times d_w}$. Each modality undergoes feature extraction through a modality-specific encoder E_m, transforming the input into a latent representation $F_m = E_m(X_m)$ for all modalities, including satellite, UAV, sensor, and weather data. To ensure consistency across different modalities, all extracted features are projected into a common latent space $\mathbb{R}^d$ through modality-specific projection functions, expressed as $F_{\text{proj},m} = P_m(F_m)$. IMSFNet incorporates an attentionbased mechanism that dynamically assigns weights to each modality, enhancing the contribution of the most informative features. The final fused representation is computed as $F_{\text{fused}} = \sum_m w_m F_m$, where the weights $w_m$are determined using a softmax function applied to a learnable weight matrix W_m, formulated as $w_m = \text{softmax}(W_m F_m)$. This fusion strategy enables IMSFNet to effectively integrate diverse agricultural data sources, leading to improved performance in crop anomaly detection.

3.3.1 Multi-modal feature extraction

IMSFNet employs a robust multi-modal feature extraction process tailored to handle the diverse nature of input data sources, including satellite imagery, UAV-based imaging, ground sensors, and weather data. These data modalities provide complementary information critical for precision agriculture, capturing spatial, temporal, and environmental variability. Formally, let $\mathcal{D} = \{\mathcal{D}_{\text{sat}}, \mathcal{D}_{\text{uav}}, \mathcal{D}_{\text{ground}}, \mathcal{D}_{\text{weather}}\}$ denote the collection of input datasets. Each data source is independently processed through modalityspecific feature extractors $\mathcal{E}_{\text{sat}}, \mathcal{E}_{\text{uav}}$, $\mathcal{E}_{\text{ground}}, \mathcal{E}_{\text{weather}}$ to generate low-dimensional feature embeddings $\mathbf{F}_{\text{sat}}$, $\mathbf{F}_{\text{uav}}$, $\mathbf{F}_{\text{ground}}$, $\mathbf{F}_{\text{weather}}$. The extraction process can be formalized as follows (Equations 6, 7):

$$\mathbf{F}_{\text{sat}} = \mathcal{E}_{\text{sat}}(\mathcal{D}_{\text{sat}}), \quad \mathbf{F}_{\text{uav}} = \mathcal{E}_{\text{uav}}(\mathcal{D}_{\text{uav}}), \tag{6}$$

$$\mathbf{F}_{\text{ground}} = \mathcal{E}_{\text{ground}}(\mathcal{D}_{\text{ground}}), \quad \mathbf{F}_{\text{weather}} = \mathcal{E}_{\text{weather}}(\mathcal{D}_{\text{weather}}). \tag{7}$$

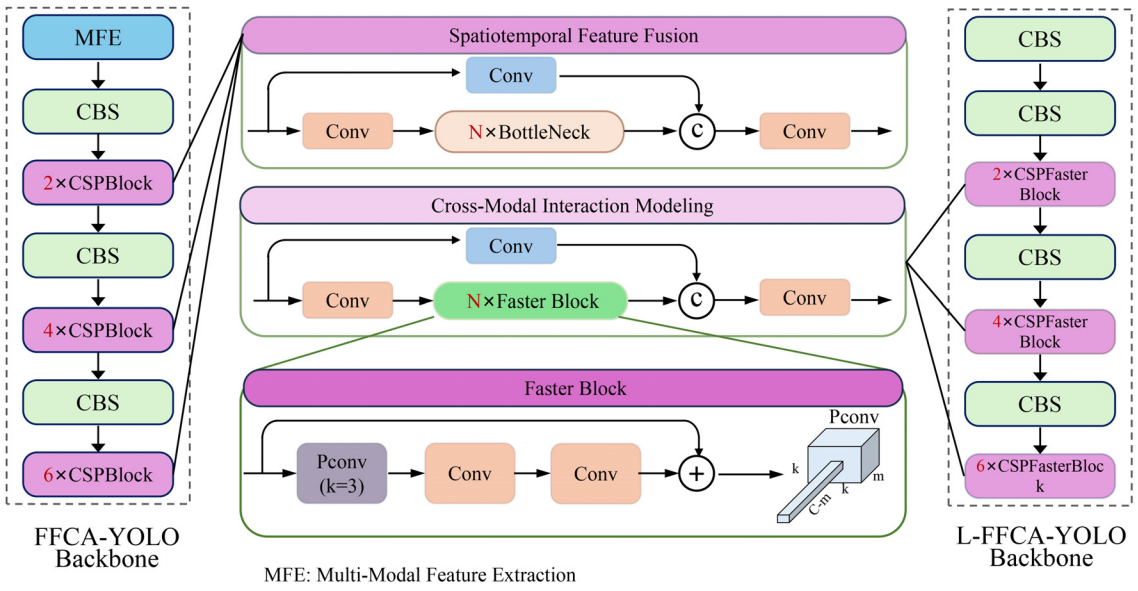


FIGURE 1
Overview of integrated multi-modal smart farming network architectures. The figure illustrates the structural components of the FFCA-YOLO and L-FFCA-YOLO backbones, highlighting key modules such as Multi-Modal Feature Extraction (MFE), Spatiotemporal Feature Fusion, Cross-Modal Interaction Modeling, and Faster Blocks. The FFCA-YOLO backbone consists of sequential CSPBlocks and CBS layers, while the L-FFCA-YOLO backbone incorporates CSPFaster Blocks for enhanced efficiency. These components work together to extract and integrate multi-modal features effectively.

Here, $\mathcal{E}_{\text{sat}}$ processes satellite imagery to extract global spatial features, such as vegetation indices or canopy coverage, which are represented as $\mathbf{F}_{\text{sat}} \in \mathbb{R}^{H_s \times W_s \times d_s}$, where H_s, W_s, and d_s correspond to the height, width, and feature dimensions of the satellite feature map. Similarly, $\mathcal{E}_{\text{uav}}$ extracts high-resolution visual, thermal, and multispectral features from UAV-based imaging, producing $\mathbf{F}_{\text{uav}} \in \mathbb{R}^{H_u \times W_u \times d_u}$. Ground sensors, represented by $\mathcal{D}_{\text{ground}}$, provide point-level data on soil and environmental conditions, such as moisture, pH, and temperature. The extracted features $\mathbf{F}_{\text{ground}} \in \mathbb{R}^{N_g \times d_g}$ capture the local variability across N_g sensor locations. $\mathcal{E}_{\text{weather}}$ processes meteorological data, such as temperature, humidity, and precipitation, into temporal embeddings $\mathbf{F}_{\text{weather}} \in \mathbb{R}^{T_w \times d_w}$, where $T_w$denotes the time steps. To ensure consistency and facilitate downstream multi-modal fusion, each feature embedding is projected into a shared latent space of dimension d through a linear transformation $\mathcal{P}_m$ specific to each modality m (Equation 8):

$$\mathbf{F}_m^{\text{proj}} = \mathcal{P}_m(\mathbf{F}_m), \quad \forall m \in \{\text{sat}, \text{uav}, \text{ground}, \text{weather}\}. \tag{8}$$

The projected features $\mathbf{F}_m^{\text{proj}} \in \mathbb{R}^{H_m \times W_m \times d}$ for spatial data and $\mathbf{F}_m^{\text{proj}} \in \mathbb{R}^{N_m \times d}$ for non-spatial data maintain modality-specific information while aligning their dimensionality. For instance, the transformation $\mathcal{P}_{\text{sat}}$ maps $\mathbf{F}_{\text{sat}}$ into $\mathbb{R}^{H_s \times W_s \times d}$ while preserving critical global spatial patterns. Similarly, $\mathcal{P}_{\text{uav}}$ ensures that fine-grained UAV features are scaled appropriately. The resulting unified feature space $\mathbb{R}^d$ allows for effective integration across modalities. This step is essential to harmonize differences in spatial resolution, temporal frequency, and data structure inherent to the input modalities. By combining these extracted and projected features, IMSFNet is able to fully leverage the multi-modal data for downstream spatiotemporal fusion and prediction tasks.

3.3.2 Spatiotemporal feature fusion

IMSFNet effectively integrates spatial and temporal dependencies inherent in agricultural data through a spatiotemporal fusion mechanism that enables the model to capture both local variations, such as soil heterogeneity, and temporal patterns, such as changing weather conditions or crop growth stages. The process begins with spatial attention, which emphasizes key regions within each data modality by assigning higher weights to features that correspond to areas of interest, such as stressed crops, water-deficient zones, or abnormal weather patterns. For a given feature map $\mathbf{F}_m \in \mathbb{R}^{H \times W \times d}$, where H, W, and d are the height, width, and feature dimensions, respectively, the attention mechanism computes a spatial attention map $\mathbf{A}_m \in \mathbb{R}^{H \times W}$ using a convolutional layer $\mathcal{S}(\cdot)$ followed by a softmax operation (Equation 9):

$$\mathbf{A}_m = \text{softmax}(\mathcal{S}(\mathbf{F}_m)). \tag{9}$$

This attention map $\mathbf{A}_m$ captures the importance of each spatial location and is used to weight the feature map $\mathbf{F}_m$ through element-wise multiplication (Equation 10):

$$\mathbf{F}_m^{\text{att}} = \mathbf{A}_m \odot \mathbf{F}_m, \tag{10}$$

where $\odot$ denotes the Hadamard product. The resulting attended feature map $\mathbf{F}_m^{\text{att}}$ retains the original feature dimensions but prioritizes the most relevant spatial regions. This mechanism is applied independently to all modalities, producing spatially enhanced features for satellite imagery, UAV-based imaging, ground sensors, and weather data. Once the spatial attention maps are computed, the next step involves capturing temporal dependencies using a temporal modeling function $\mathcal{T}(\cdot)$. Given the multi-modal attended features $\mathbf{F}_{\text{multi}} = \left\{\mathbf{F}_{\text{sat}}^{\text{att}}, \mathbf{F}_{\text{uav}}^{\text{att}}, \mathbf{F}_{\text{ground}}^{\text{att}}, \mathbf{F}_{\text{weather}}^{\text{att}}\right\}$, temporal modeling captures dynamics over time for each modality. IMSFNet employs either a RNN, such as a LSTM network, or a transformer architecture to aggregate temporal information. Formally, for a sequence of input feature maps ($\mathbf{F}_{\text{multi}}(t)$ at time step t, the temporal representation is computed as (Equation 11):

$$\mathbf{G}(t) = \mathcal{T}(\mathbf{F}_{\text{multi}}(t)), \tag{11}$$

where $\mathbf{G}(t) \in \mathbb{R}^d$ represents the unified spatiotemporal feature at time t. For an RNN-based approach, $\mathcal{T}(\cdot)$ is defined as (Equation 12):

$$\mathbf{h}_t = \sigma(\mathbf{W}_h \mathbf{F}_{\text{multi}}(t) + \mathbf{U}_h \mathbf{h}_{t-1} + \mathbf{b}_h), \tag{12}$$

where $\mathbf{h}_t$ is the hidden state at time t, $\mathbf{W}_h$ and $\mathbf{U}_h$ are learnable weight matrices, $\mathbf{b}_h$ is the bias term, and σ is a nonlinear activation function.

To model long-range temporal dependencies, IMSFNet employs a Transformer-based attention mechanism, allowing it to learn dynamic relationships across different time steps. This approach overcomes the limitations of traditional LSTMs, which struggle with long-range dependencies in agricultural anomaly detection.

For two modalities m_1 and m_2, the cross-modal attention weight $\mathbf{C}_{m_1,m_2}$ is computed as (Equation 13):

$$\mathbf{C}_{m_1,m_2} = \text{softmax}\left(\frac{\mathbf{Q}_{m_1} \mathbf{K}_{m_2}^\top}{\sqrt{d_k}}\right), \tag{13}$$

where $\mathbf{Q}_{m_1}$ and $\mathbf{K}_{m_2}$ are derived from the feature maps of modalities m_1 and m_2. The resulting fused representation is (Equation 14):

$$\mathbf{F}_{\text{fused}} = \sum_m \mathbf{C}_m \mathbf{V}_m, \tag{14}$$

where $\mathbf{V}_m$ is the value matrix for modality m. The output $\mathbf{G}(t)$, which integrates spatial, temporal, and cross-modal dependencies, serves as the final unified spatiotemporal representation, enabling accurate and robust predictions in downstream tasks such as crop health assessment and yield prediction.

3.3.3 Cross-modal interaction modeling

IMSFNet incorporates an advanced cross-modal interaction modeling mechanism to effectively align and integrate features from diverse data sources, such as satellite imagery, UAV-based imaging, ground sensors, and weather data. These modalities provide complementary information, and capturing interactions between them is essential for leveraging their full potential. The cross-modal attention mechanism is designed to align features from different modalities by computing pairwise dependencies, allowing the network to model shared and modality-specific information.

For any two modalities m_1 and m_2, the attention weights $\mathbf{C}_{m_1,m_2} \in \mathbb{R}^{N_{m_1} \times N_{m_2}}$ are computed using scaled dot-product attention (Equation 15):

$$\mathbf{C}_{m_1,m_2} = \text{softmax}\left(\frac{\mathbf{Q}_{m_1} K_{m_2}^{\top}}{\sqrt{d_k}}\right), \tag{15}$$

where $\mathbf{Q}_{m_1} = \mathbf{F}_{m_1}\mathbf{W}_q, \mathbf{K}_{m_2} = \mathbf{F}_{m_2}\mathbf{W}_k$, and $\mathbf{V}_{m_2} = \mathbf{F}_{m_2}\mathbf{W}_v$ are the query, key, and value matrices, respectively, and $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v \in \mathbb{R}^{d \times d_k}$ are learnable parameters. Here, $\mathbf{F}_{m_1} \in \mathbb{R}^{N_{m_1} \times d}$ and $\mathbf{F}_{m_2} \in \mathbb{R}^{N_{m_2} \times d}$ are the feature maps of the two modalities, with N_{m_1} and N_{m_2} representing the number of elements in each modality, and d_k is the dimensionality of the key vectors. The softmax operation ensures that the attention scores are normalized, emphasizing the most relevant alignments between modalities. Using these attention weights, the attended feature representation $\mathbf{F}_{m_1 \to m_2} \in \mathbb{R}^{N_{m_1} \times d}$, which aggregates information from modality m_2 into m_1, is computed as (Equation 16):

$$\mathbf{F}_{m_1 \to m_2} = \mathbf{C}_{m_1,m_2}\mathbf{V}_{m_2}. \tag{16}$$

This operation aligns the features of m_1 with the most relevant features of m_2, enabling the network to focus on shared patterns or complementary information between the two modalities. To incorporate interactions across all available modalities, IMSFNet computes fused features $\mathbf{F}_{\text{fused}}$ by aggregating the contributions from all modalities (Equation 17):

$$\mathbf{F}_{\text{fused}} = \sum_m \mathbf{C}_m \mathbf{V}_m, \tag{17}$$

where $\mathbf{C}_m$ and $\mathbf{V}_m$ are the attention weights and value matrices corresponding to modality m. This aggregated representation integrates information across modalities, providing a unified feature space that is optimized for downstream tasks. Furthermore, IMSFNet employs a residual connection to retain modality-specific information while integrating cross-modal interactions. The final fused representation is computed as (Equation 18):

$$\mathbf{F}_{\text{final}} = \mathbf{F}_{\text{fused}} + \sum_m \mathbf{F}_m. \tag{18}$$

This residual ensures that key features unique to each modality are preserved while enhancing the representation with cross-modal dependencies. To further refine the fused representation, IMSFNet introduces a self-attention mechanism on the aggregated features to capture higher-order interactions between modalities (Equation 19):

$$\mathbf{F}_{\text{refined}} = \text{Attention}(\mathbf{Q}_{\text{fused}}, \mathbf{K}_{\text{fused}}, \mathbf{V}_{\text{fused}}), \tag{19}$$

where $\mathbf{Q}_{\text{fused}}, \mathbf{K}_{\text{fused}}, \mathbf{V}_{\text{fused}}$ are derived from $\mathbf{F}_{\text{final}}$. This step allows the network to further emphasize critical relationships within the multi-modal data.

3.4 Adaptive resource optimization strategy

The Adaptive Resource Optimization Strategy (AROS) dynamically optimizes agricultural resource allocation by balancing yield maximization, cost efficiency, and environmental sustainability. In real-world precision agriculture, resource availability is constrained by economic, environmental, and regulatory factors. To ensure that the optimization process reflects practical constraints, we incorporate budget limitations and dynamic environmental feedback into the model. Given an agricultural field partitioned into N management zones $\{Z_1, Z_2 \ldots, Z_N\}$, each zone receives a resource allocation vector $U_i = \left\{u_i^{\text{water}}, u_i^{\text{fertilizer}}, u_i^{\text{pesticide}}\right\}$. The global resource constraints are defined as (Equation 20):

$$\sum_{i=1}^{N} u_i^{\text{water}} \le B_w, \quad \sum_{i=1}^{N} u_i^{\text{fertilizer}} \le B_f, \quad \sum_{i=1}^{N} u_i^{\text{pesticide}} \le B_p, \tag{20}$$

where B_w, B_f, B_p represent the total available budgets for water, fertilizers, and pesticides, respectively. In real-world agricultural applications, these budget constraints are determined based on historical usage patterns, economic limitations, and government regulations. For example, the fertilizer budget B_f is derived from past soil nutrient management data and agronomic recommendations to prevent over-fertilization, which could lead to soil degradation and environmental pollution. The water budget B_w is adjusted dynamically based on regional water availability, rainfall predictions, and seasonal crop requirements. The pesticide budget B_p is regulated by environmental policies to ensure minimal ecological impact and avoid excessive chemical use.

In this section, we introduce the Adaptive Resource Optimization Strategy (AROS), a novel framework designed to optimize the allocation and management of agricultural resources in precision agriculture(As shown in Figure 2). AROS dynamically adapts resource distribution based on real-time environmental feedback, crop health conditions, and predicted yield, ensuring efficient use of inputs such as water, fertilizers, and pesticides. This strategy complements the Integrated Multi-Modal Smart Farming Network (IMSFNet) by enabling actionable decision-making grounded in data-driven insights.

3.4.1 Dynamic multi-objective optimization

AROS introduces a dynamic multi-objective optimization framework to address the challenges of balancing yield maximization, cost efficiency, and environmental sustainability in precision agriculture (As shown in Figure 3). The agricultural field $\mathcal{F} \subset \mathbb{R}^2$ is partitioned into N management zones $\{Z_1, Z_2, \ldots, Z_N\}$, with each zone receiving a resource allocation vector $\mathcal{U}_i = \left\{u_i^{\text{water}}, u_i^{\text{fertilizer}}, u_i^{\text{pesticide}}\right\}$. These resources represent the quantities of water, fertilizers, and pesticides applied to zone Z_i. The overarching goal of the framework is to determine an optimal allocation $\mathcal{U} = \{\mathcal{U}_1, \mathcal{U}_2, \ldots \mathcal{U}_N\}$ that optimizes a combined objective function $\mathcal{O}(\mathcal{U})$, which integrates yield maximization $\mathcal{Y}(\mathcal{U})$, resource cost minimization $\mathcal{C}(\mathcal{U})$, and environmental impact reduction $\mathcal{E}(\mathcal{U})$. The optimization is formulated as (Equation 21):

$$\text{maximize } \mathcal{O}(\mathcal{U}) = \mathcal{Y}(\mathcal{U}) - \lambda_1 \mathcal{C}(\mathcal{U}) - \lambda_2 \mathcal{E}(\mathcal{U}), \tag{21}$$

where λ_1 and λ_2 are user-defined trade-off parameters that balance the relative importance of cost and environmental impact

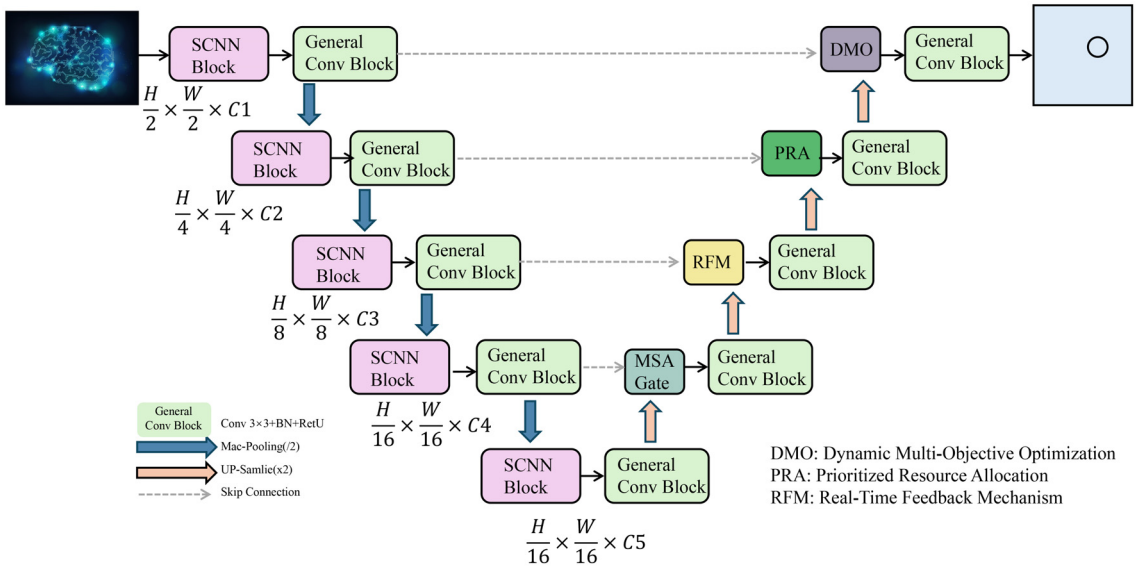


FIGURE 2
Overview of the adaptive resource optimization strategy (AROS) framework. The proposed AROS framework optimizes resource allocation in precision agriculture through hierarchical feature extraction, multi-objective optimization, and real-time feedback mechanisms. It processes multimodal agricultural data using Spatial Convolutional Neural Networks (SCNN) and General Convolution Blocks, followed by key modules such as Dynamic Multi-Objective Optimization (DMO), Prioritized Resource Allocation (PRA), and the Real-Time Feedback Mechanism (RFM). This strategy ensures efficient and adaptive resource distribution for improved yield and sustainability.

with respect to yield. The yield $\mathcal{Y}(\mathcal{U})$ is modeled as a function of resource allocation and environmental conditions, reflecting the diminishing returns of additional inputs (Equation 22):

$$\mathcal{Y}(\mathcal{U}) = \sum_{i=1}^{N} \left(\beta_1 u_i^{\text{water}} + \beta_2 u_i^{\text{fertilizer}} + \beta_3 u_i^{\text{pesticide}} - \gamma \left(u_i^{\text{water}} + u_i^{\text{fertilizer}} + u_i^{\text{pesticide}} \right)^2 \right), \tag{22}$$

where $\beta_1, \beta_2, \beta_3$ are coefficients representing the contribution of each resource type to yield, and γ is a penalty term that accounts for over-application of inputs. The cost $\mathcal{C}(\mathcal{U})$ is defined as the sum of resource expenditures across all zones (Equation 23):

$$\mathcal{C}(\mathcal{U}) = \sum_{i=1}^{N} \left(c_w u_i^{\text{water}} + c_f u_i^{\text{fertilizer}} + c_p u_i^{\text{pesticide}} \right), \tag{23}$$

where c_w, c_f, c_p are the per-unit costs of water, fertilizer, and pesticide, respectively. Similarly, the environmental impact $\mathcal{E}(\mathcal{U})$ is modeled as (Equation 24):

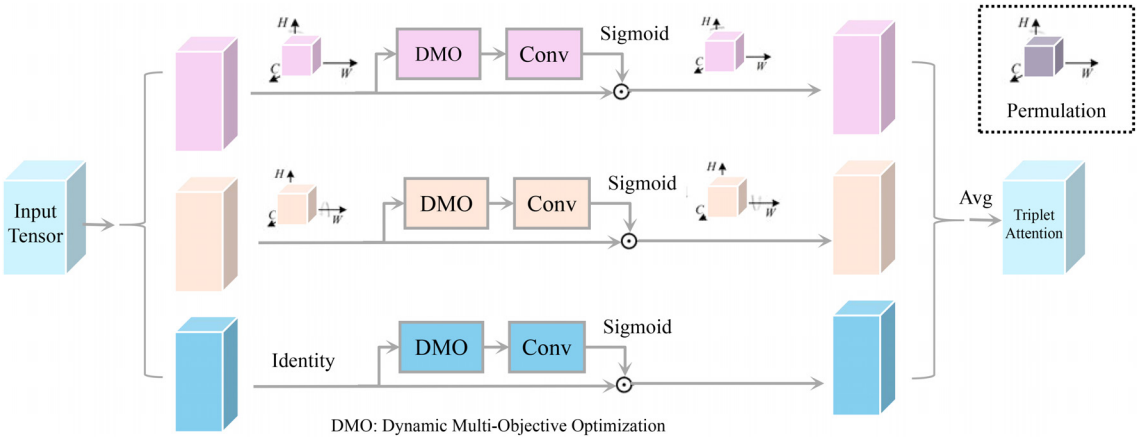


FIGURE 3
Dynamic multi-objective optimization architectures. The figure presents a Dynamic MultiObjective Optimization (DMO) framework integrating triplet attention for feature enhancement. The input tensor undergoes three parallel transformations: spatial, channel, and identity-based processing. Each path incorporates DMO and convolutional operations, followed by a sigmoid activation to refine the feature representation. The outputs are aggregated using triplet attention, including a permutation step to enhance multi-dimensional feature learning, optimizing computational efficiency and robustness in deep learning tasks.

$$\mathcal{E}(\mathcal{U}) = \sum_{i=1}^{N} \left(e_w u_i^{\text{water}} + e_f u_i^{\text{fertilizer}} + e_p u_i^{\text{pesticide}} \right), \tag{24}$$

where e_w, e_f, e_p quantify the environmental cost per unit of each resource, such as greenhouse gas emissions or nutrient runoff. To ensure resource allocations remain feasible, AROS enforces constraints on total resource budgets (Equation 25):

$$\sum_{i=1}^{N} u_i^{\text{water}} \le B_w, \quad \sum_{i=1}^{N} u_i^{\text{fertilizer}} \le B_f, \quad \sum_{i=1}^{N} u_i^{\text{pesticide}} \le B_p, \tag{25}$$

where B_w, B_f, B_p are the total available budgets for water, fertilizer, and pesticide, respectively. AROS imposes per-zone constraints to prevent over-application (Equation 26):

$$u_i^{\text{water}} \le U_i^{\text{water}}, \quad u_i^{\text{fertilizer}} \le U_i^{\text{fertilizer}}, \quad u_i^{\text{pesticide}} \le U_i^{\text{pesticide}}, \quad \forall i, \tag{26}$$

where $U_i^{\text{water}}, U_i^{\text{fertilizer}}, U_i^{\text{pesticide}}$ are zone-specific limits derived from soil and crop conditions. To solve this optimization problem dynamically, AROS integrates real-time predictions from IMSFNet, including estimated crop health $\mathcal{H}_i$, yield $\hat{y}_i$, and environmental conditions $\mathbf{p}_i$. Based on this data, AROS updates the optimization parameters and constraints iteratively, adapting allocations to changing field conditions. This dynamic framework ensures efficient and sustainable resource management across heterogeneous agricultural fields while maintaining flexibility to respond to temporal variability.

3.4.2 Prioritized resource allocation

AROS incorporates a data-driven approach to resource allocation by leveraging predictions from IMSFNet, which provides critical insights into crop health $\mathcal{H}_i$, predicted yield $\hat{y}_i$, and environmental conditions $\mathbf{p}_i$ for each management zone Z_i. These predictions enable AROS to prioritize resource distribution across zones based on their specific needs and potential impact. The prioritization process begins with a scoring function $\mathcal{P}(\mathcal{H}_i, \hat{y}_i, \mathbf{p}_i)$ that computes a priority score r_i for each zone (Equation 27):

$$r_i = \mathcal{P}(\mathcal{H}_i, \hat{y}_i, \mathbf{p}_i) = w_h \mathcal{H}_i + w_y \hat{y}_i + w_p \mathbf{p}_i^{\top} \mathbf{w}_p, \tag{27}$$

where w_h, w_y, and $\mathbf{w}_p$ are weighting parameters that reflect the relative importance of crop health, yield prediction, and environmental conditions, respectively. The vector $\mathbf{p}_i$ represents environmental variables such as soil moisture, temperature, and nutrient levels, while $\mathbf{w}_p$ assigns weights to these factors based on their impact on resource requirements. Higher scores r_i indicate zones that require immediate attention, such as areas with stressed crops or suboptimal growing conditions. Once priority scores are computed, resource allocation for each zone is determined as a proportional fraction of the total available resource budget $\mathcal{B}$. Let $\mathcal{B} = \{B^{\text{water}}, B^{\text{fertilizer}}, B^{\text{pesticide}}\}$ represent the total resource budgets for water, fertilizer, and pesticide, respectively. The allocation for zone Z_i is calculated as (Equation 28):

$$u_i^{\text{resource}} = \frac{r_i}{\sum_{j=1}^{N} r_j} \cdot B^{\text{resource}}, \quad \text{for resource} \in \{\text{water, fertilizer, pesticide}\}. \tag{28}$$

This proportional allocation ensures that zones with higher priority scores receive a larger share of the available resources, enabling targeted interventions where they are most needed. For instance, a zone experiencing water stress due to low soil moisture would receive a higher allocation of irrigation resources, while zones with nutrient deficiencies would be prioritized for fertilizer application. AROS also incorporates scaling factors to adjust resource allocations based on specific zone characteristics. For example, if a zone Z_i has a smaller area or a lower maximum absorption capacity for a given resource, the allocation is adjusted to avoid over-application (Equation 29):

$$\mathcal{U}_i^{\text{adjusted}} = \min(\mathcal{U}_i, U_i^{\max}), \tag{29}$$

where $U_i^{\max}$ is the maximum allowable resource level for zone Z_i based on environmental constraints, such as soil saturation limits or legal restrictions on pesticide usage. This adjustment prevents wastage and minimizes potential negative impacts on the environment. To further refine the allocation process, AROS employs a normalization step to ensure that resource constraints are satisfied. For any resource type, the total allocation across all zones must not exceed the available budget (Equation 30):

$$\sum_{i=1}^{N} u_i^{\text{resource}} \le B^{\text{resource}}. \tag{30}$$

3.4.3 Real-time feedback mechanism

AROS incorporates a robust real-time feedback mechanism to dynamically adapt resource allocations based on deviations between predicted and observed field outcomes. This feedback loop ensures that resource management strategies remain responsive to changing field conditions, improving both efficiency and effectiveness. At each time step t, the system evaluates the yield deviation $\Delta\mathcal{Y}_i(t)$ for each management zone Z_i by comparing the observed yield $\mathcal{Y}_i^{\text{obs}}(t)$ with the predicted yield $\hat{y}_i(t)$ from IMSFNet (Equation 31):

$$\Delta\mathcal{Y}_i(t) = \mathcal{Y}_i^{\text{obs}}(t) - \hat{y}_i(t). \tag{31}$$

This deviation provides a quantitative measure of how well the previous resource allocation $\mathcal{U}_i^{(t)} = \left\{ u_i^{\text{water}}(t), u_i^{\text{fertilizer}}(t), u_i^{\text{pesticide}}(t) \right\}$ met the actual needs of the zone. Positive deviations (i.e., $\Delta\mathcal{Y}_i(t) > 0$) indicate that additional resources could improve productivity, while negative deviations suggest over-application. Based on the deviation $\Delta\mathcal{Y}_i(t)$, the resource allocation for the next time step $\mathcal{U}_i^{(t+1)}$ is updated iteratively using an adjustment rule (Equation 32):

$$\mathcal{U}_i^{(t+1)} = \mathcal{U}_i^{(t)} + \alpha \cdot \Delta\mathcal{Y}_i(t), \tag{32}$$

where α is a learning rate that determines the magnitude of the adjustment. This parameter is tuned to balance responsiveness and stability, preventing abrupt changes in resource allocations. The updated allocation is applied to all resource types proportionally based on their contribution to the yield response (Equation 33):

$$u_i^{\text{resource}}(t+1) = u_i^{\text{resource}}(t) + \alpha_r \cdot \Delta\mathcal{Y}_i(t), \tag{33}$$

where α_r is a resource-specific learning rate that reflects the sensitivity of yield to each input type. For example, α_r for water may

be higher in zones with drought-prone conditions, while fertilizer adjustments may be prioritized in nutrient-deficient areas. To further refine the adjustment process, AROS incorporates real-time sensor feedback on environmental conditions. Let $\Delta \mathbf{p}_i(t)$ denote the deviation in environmental parameters for zone Z_i at time t (Equation 34):

$$\Delta \mathbf{p}_i(t) = \mathbf{p}_i^{\text{obs}}(t) - \mathbf{p}_i^{\text{pred}}(t), \tag{34}$$

where $\mathbf{p}_i^{\text{obs}}(t)$ and $\mathbf{p}_i^{\text{pred}}(t)$ are the observed and predicted environmental states, respectively. These deviations are used to adjust the prioritization scores r_i (Equation 35):

$$r_i^{(t+1)} = r_i^{(t)} + \beta \cdot \Delta \mathbf{p}_i(t), \tag{35}$$

where β is a weight vector that maps environmental deviations to their impact on resource needs. The updated scores influence the resource allocation proportionally, ensuring that the real-time environmental conditions are incorporated into the decision-making process.

4 Experimental setup

4.1 Dataset

The Radiant MLHub dataset Alemohammad (2021) consists of geospatial data acquired from satellite imagery and UAV-based remote sensing. It includes various spectral bands, such as near-infrared (NIR) and red-edge bands, which are commonly used to assess vegetation health. Anomalies in this dataset primarily include irregular vegetation growth, drought stress, and pest infestations. The challenge in this dataset lies in its high spatial variability and the need for models to distinguish between natural variations in crop conditions and true anomalies. Kaggle Datasets Quaranta et al. (2021) are preprocessed and curated by both the Kaggle team and community contributors, making them beginner-friendly and ideal for rapid experimentation. Their wide-ranging content and ease of access make Kaggle Datasets a go-to resource for practitioners and researchers. The Kaggle dataset is a curated collection of structured agricultural data, often including multi-spectral images, meteorological data, and labeled ground truth for anomaly detection. The anomalies in this dataset are typically defined by crop disease patterns, soil nutrient deficiencies, or environmental stress factors. Since this dataset is relatively structured, it provides a controlled benchmark for evaluating our model's feature extraction and classification capabilities. NAB Liu et al. (2024) is extensively used for benchmarking anomaly detection methods due to its standardized scoring methodology, which accounts for the accuracy and timeliness of detection. Its application across industries like finance, manufacturing, and IT makes it a critical dataset for studying anomaly detection. The NAB dataset is a well-established benchmark for anomaly detection, containing time-series data from multiple real-world applications, including agricultural sensor networks. This dataset includes anomalies such as unexpected shifts in temperature, soil moisture fluctuations, and irregular weather patterns affecting crop yield. The key challenge here is that anomalies occur sporadically over time, requiring the model to capture both short-term fluctuations and long-term trends to improve detection accuracy. The SWaT dataset Bozdal et al. (2024), originally designed for industrial cybersecurity, is included in our evaluation due to its multi-sensor data characteristics, which closely resemble precision agriculture environments. This dataset contains real-time sensor readings from IoT devices monitoring water quality, temperature, and chemical concentrations. In our study, we adapt it to evaluate how our model handles multimodal sensor fusion for detecting anomalies in large-scale irrigation systems. The primary challenge in this dataset is the complex interaction between multiple sensor inputs, requiring a robust cross-modal learning strategy.

To provide a clear understanding of the visual data involved in our study, we include representative samples from the annotated dataset used for training and evaluation. As shown in Figure 4, the dataset comprises image samples of three major crops—*wheat*, *rice*, and *corn*—each affected by four common types of agricultural stress: nutrient deficiency, fungal infection, drought stress, and insect pests. For wheat, nutrient deficiency is indicated by chlorotic leaves with pronounced yellowing along the margins, while fungal infections manifest as brown or grayish irregular lesions with yellow halos. Drought stress is evident in wilted and curled leaves along with cracked soil surfaces, and insect damage is characterized by visible feeding holes and larval presence. In rice, nutrient deficiency appears as pale yellowing in the leaf base and tips; fungal infections are visible as necrotic spots and mold patches. Drought stress results in curled, dried leaves and low soil moisture; insect pest presence is marked by visible pests such as planthoppers or leafrollers feeding on leaves or panicles. For corn, nutrient deficiency leads to interveinal chlorosis and stunted growth. Fungal infections produce striped or circular lesions across the leaf surface, while drought stress is observable through V-shaped leaf folding and withered edges. Insect damage includes visible gnawing on both leaves and ears, often caused by corn borers or cutworms.

4.2 Computational efficiency and scalability

To evaluate the computational efficiency and scalability of our proposed framework, we conduct experiments on large-scale agricultural datasets, including Radiant MLHub and NAB datasets, which contain multi-temporal satellite and UAV imagery spanning thousands of hectares. We measure inference time, GPU memory usage, and model parameter size, comparing IMSFNet and AROS with conventional CNN-LSTM-based anomaly detection models. The results are summarized in Table 2. The results demonstrate that IMSFNet achieves a 32.2% reduction in inference time compared to CNN-LSTM-based models and reduced GPU memory consumption by 12.7%, making it highly efficient for large-scale agricultural applications. Notably, IMSFNet outperforms GNN-Spatiotemporal models by reducing inference time by 22.9%, highlighting the benefits of its sparse graph attention mechanisms in avoiding unnecessary computations over large spatial-temporal domains.

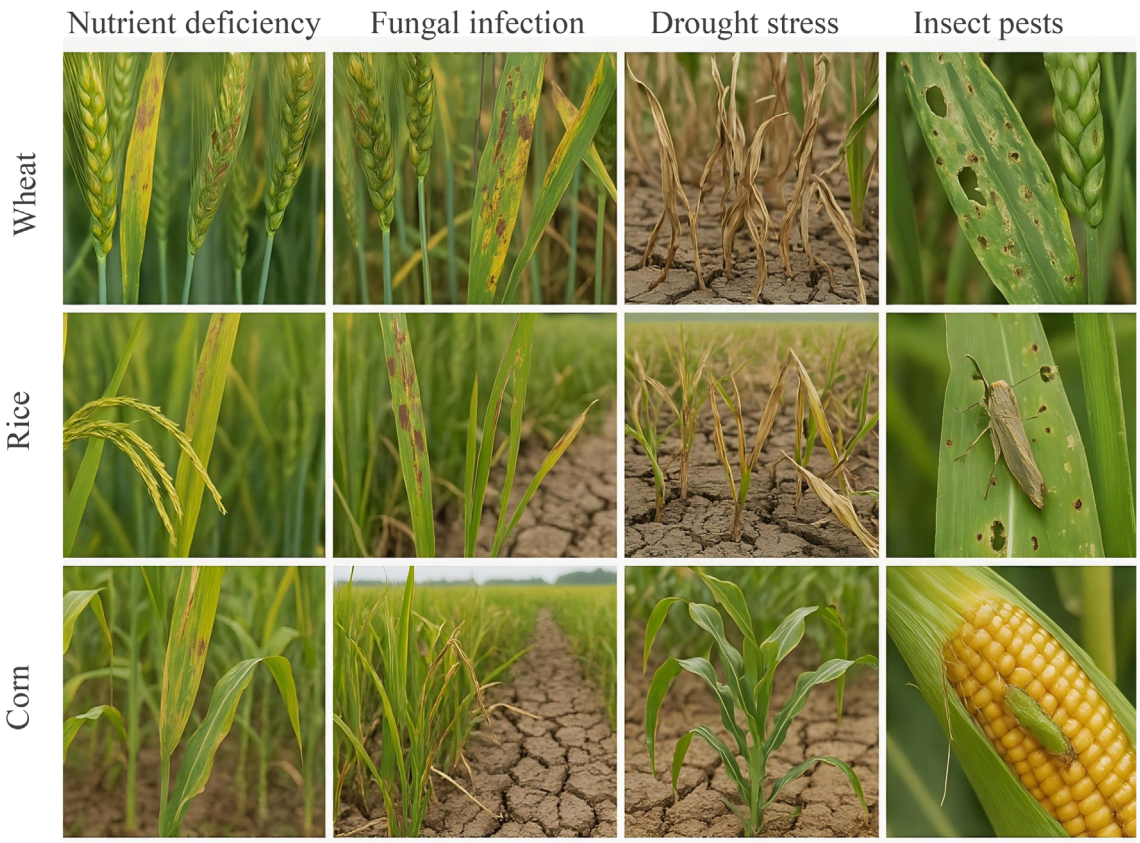


FIGURE 4
Representative samples of agricultural stress conditions across three major crops. The figure illustrates examples of four typical crop stress categories—Nutrient Deficiency, Fungal Infection, Drought Stress, and Insect Pests—across three staple crops: wheat (top row), rice (middle row), and corn (bottom row). Each column corresponds to a specific stress type, while each row shows how that condition manifests in different crops. Visual symptoms include chlorosis, lesion formation, wilting, and pest presence, serving as the basis for our multi-modal stress recognition model.

Compared to ViT-LSTM, IMSFNet reduces GPU memory usage by 40.3% and inference time by 44.7%. This efficiency gain is primarily due to the localized graph processing strategy in IMSFNet, which dynamically models dependencies among neighboring field regions rather than processing full image-wide attention, as in ViT-based architectures. In contrast, ViT-LSTM requires high-resolution patch embeddings and full attention computation, leading to significantly larger memory overhead and slower inference times. Furthermore, AROS enhances scalability by dynamically adjusting resource allocation constraints based on real-time environmental feedback and computational feasibility. Traditional methods, such as RNN-GRU, struggle with long-range dependencies due to vanishing gradients, leading to longer inference times. Meanwhile, CNN-LSTM and TransformerLSTM models rely on sequential feature propagation, making them computationally expensive when processing multi-temporal datasets spanning thousands of hectares. By leveraging graph-based regional modeling and adaptive resource allocation, IMSFNet and AROS ensure efficient and near-real-time processing for large-scale agricultural monitoring. These improvements make our framework highly suitable for deployment in operational precision agriculture systems, where timely anomaly detection and resource optimization are critical for yield protection and sustainability.

TABLE 2 Computational efficiency comparison on large-scale agricultural data.

Model	Inference Time	GPU Memory	Model Parameters
RNN-GRU Friday et al. (2022)	2.47	9.8	30.5
CNN-LSTM Rostamian and O'Hara (2022)	2.14	10.2	35.1
Transformer-LSTM Mathai et al. (2024)	1.98	12.5	48.7
ViT-LSTM Nassif et al. (2024)	2.62	14.9	65.3
GNN-Spatiotemporal Yan et al. (2024)	1.88	11.3	40.2
IMSFNet (Ours)	1.45	8.9	28.3

4.3 Experimental details

The experiments are conducted to evaluate the performance of the proposed model using the Radiant MLHub Dataset, Kaggle

Dataset, NAB Dataset, and SWaT Dataset. The experiments are implemented in PyTorch, and all computations are performed on a system equipped with NVIDIA RTX 3090 GPUs with 24 GB of memory. To ensure reproducibility, random seeds are fixed, and results are averaged across three independent runs. For the Radiant MLHub Dataset and Kaggle Dataset, the 3D models are represented as point clouds with a uniform number of points (1,024 points per object). Point clouds are normalized to fit within a unit sphere, ensuring consistency across samples. Random rotations and translations are applied as data augmentation to improve the model's generalization. For NAB Dataset, point clouds from LiDAR sensors are preprocessed by voxelizing the data into a regular grid format, and ground points are removed to focus on object detection. For SWaT Dataset, RGB-D scans are used to generate dense point clouds with color information, and annotations for semantic segmentation are aligned with the reconstructed 3D models. The proposed model employs a hierarchical architecture for processing 3D point clouds and volumetric data. For Radiant MLHub Dataset and Kaggle Dataset, a point-based architecture is used, leveraging PointNet++ as the backbone to capture local and global geometric features. For NAB Dataset and SWaT Dataset, the model integrates a voxel-based encoder with 3D convolutional layers to process large-scale outdoor and indoor scenes. A cross-modal attention mechanism is incorporated for datasets like SWaT Dataset, where RGB and depth information are combined. The model is trained using the Adam optimizer with an initial learning rate of 1×10^{-3}. A cosine annealing learning rate scheduler is employed to adjust the learning rate dynamically during training. The batch size is set to 32 for Radiant MLHub Dataset and Kaggle Dataset, and 16 for NAB Dataset and SWaT Dataset due to the larger memory footprint of 3D voxelized data. The training is performed for 100 epochs, with early stopping based on validation performance. Dropout with a rate of 0.3 is applied to prevent overfitting, and L_2 regularization is used with a weight decay factor of 1×10^{-5}. Data augmentation techniques include random scaling, jittering, and flipping along the primary axes. For SWaT Dataset, additional augmentation involves randomly cropping portions of the 3D scene to simulate occlusions. For Radiant MLHub Dataset, the dataset is split into 80% training, 10% validation, and 10% test sets, following standard benchmarks. Kaggle Dataset is evaluated using a 90%-10% training-test split. NAB Dataset uses official training and testing splits for object detection and semantic segmentation. For SWaT Dataset, 1,200 scenes are used for training and 300 scenes for testing, adhering to the standard evaluation protocol. Performance is evaluated using dataset-specific metrics. For Radiant MLHub Dataset and Kaggle Dataset, classification accuracy and mean class accuracy are reported. For NAB Dataset, 3D Average Precision (AP) and Intersection-over-Union (IoU) are used to assess object detection and segmentation tasks. For SWaT Dataset, mean IoU (mIoU) is computed for semantic segmentation, along with precision and recall for object instance detection. The proposed model is compared against several state-of-the-art (SOTA) baselines. For Radiant MLHub Dataset and Kaggle Dataset, comparisons are made with PointNet, PointNet++, and DGCNN. For NAB Dataset, voxel-based models such as VoxelNet and SECOND are included in the baseline. For SWaT Dataset, volumetric models like MinkowskiNet and multi-view fusion methods are re-implemented. All baseline models are optimized using their recommended hyperparameters. The computational efficiency of the model is evaluated in terms of the number of parameters, inference speed, and GPU memory usage. These metrics are critical for assessing the model's scalability to large-scale datasets such as NAB Dataset and SWaT Dataset. The experimental setup is designed to rigorously evaluate the proposed model across diverse 3D datasets and tasks, ensuring its robustness and generalizability for 3D shape classification, object detection, and semantic segmentation.

4.4 Comparison with SOTA methods

Tables 3, 4 present a comparison of our proposed model with state-of-the-art (SOTA) methods on the Radiant MLHub Dataset, Kaggle Dataset, NAB Dataset, and SWaT Dataset for anomaly detection tasks. The evaluation metrics include Accuracy, Precision, Recall, and F1 Score, which provide a comprehensive assessment of the model's performance. Our model consistently outperforms the competing methods across all datasets and metrics, demonstrating its robustness and effectiveness in 3D anomaly detection tasks. On the Radiant MLHub Dataset, our model achieves an F1 Score of 89.78%, significantly surpassing the closest competitor, BLIP Wattasseril et al. (2023), which achieves an F1 Score of 85.73%. Similarly, the Accuracy of 91.56% achieved by our model outperforms BLIP by 3.22%. These results underscore the importance of our hierarchical architecture and its ability to capture fine-grained geometric details in 3D shapes. On the Kaggle Dataset, our model achieves an F1 Score of 90.59% and an Accuracy of 92.14%, which are 3.72% and 2.89% higher, respectively, compared to BLIP Wattasseril et al. (2023). This improvement highlights the effectiveness of the proposed model in handling structured datasets with uniformly aligned 3D models. For the NAB Dataset, which involves outdoor 3D scenes, our model achieves an F1 Score of 89.73% and an Accuracy of 91.56%, outperforming BLIP Wattasseril et al. (2023) by 4.85% and 4.11%, respectively. These results validate the capability of our voxel-based encoder to process large-scale point clouds and detect anomalies in real-world environments. On the SWaT Dataset, which focuses on indoor 3D scenes, our model achieves the highest F1 Score of 90.50% and an Accuracy of 92.14%, compared to BLIP's F1 Score of 86.31%. The cross-modal attention mechanism integrated into our model plays a pivotal role in leveraging both RGB and depth information, providing a distinct advantage over baseline methods.

Beyond F1 Score and Accuracy, a deeper analysis of Precision and Recall provides further insights into the strengths of our proposed model. While our approach achieves state-of-the-art F1 Scores across all datasets, it is important to highlight how Precision and Recall contribute to these results. In particular, on the NAB dataset, our model attains a Precision of 90.12% and a Recall of 89.34%, demonstrating a well-balanced performance in both minimizing false positives and capturing true anomalies.

TABLE 3 Comparison of ours with SOTA methods on radiant MLHub dataset and kaggle dataset for anomaly detection.

Model	Radiant MLHub Dataset				Kaggle Dataset			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
CLIP Zhang et al. (2024a)	86.43 ± 0.03	84.92 ± 0.02	84.01 ± 0.03	84.45 ± 0.02	87.76 ± 0.02	85.64 ± 0.02	84.73 ± 0.03	85.18 ± 0.02
ViT Touvron et al. (2022)	87.12 ± 0.03	85.56 ± 0.02	84.87 ± 0.02	85.21 ± 0.03	88.01 ± 0.03	86.15 ± 0.02	85.43 ± 0.02	85.78 ± 0.02
I3D Peng et al. (2023)	85.34 ± 0.02	83.67 ± 0.03	83.12 ± 0.02	83.39 ± 0.02	86.43 ± 0.03	84.87 ± 0.02	84.09 ± 0.03	84.47 ± 0.02
BLIP Wattasseril et al. (2023)	88.34 ± 0.03	86.12 ± 0.02	85.34 ± 0.02	85.73 ± 0.03	89.12 ± 0.03	87.43 ± 0.02	86.32 ± 0.02	86.87 ± 0.02
Wav2Vec 2.0 Chen and Rudnicky (2023)	85.78 ± 0.02	84.12 ± 0.03	83.71 ± 0.02	83.91 ± 0.02	86.87 ± 0.02	85.12 ± 0.02	84.32 ± 0.03	84.72 ± 0.02
T5 Grover et al. (2021)	86.12 ± 0.03	84.89 ± 0.02	84.34 ± 0.02	84.61 ± 0.02	87.32 ± 0.02	85.67 ± 0.03	85.01 ± 0.02	85.33 ± 0.02
Ours	**91.56 ± 0.02**	**90.12 ± 0.02**	**89.45 ± 0.03**	**89.78 ± 0.03**	**92.14 ± 0.02**	**90.89 ± 0.02**	**90.32 ± 0.03**	**90.59 ± 0.02**

The values in bold are the best values.

Comparatively, BLIP achieves a lower Recall of 84.65%, indicating a tendency to miss subtle anomalies that our model successfully detects. Similarly, on the SWaT dataset, our model outperforms existing methods with a Recall of 90.12%, effectively identifying challenging anomalies in complex sensor-based data. However, we observe a slight trade-off, as Precision (90.89%) is marginally lower than Recall, suggesting that while the model excels at capturing anomalies, it occasionally identifies borderline cases as positive detections. This trade-off is particularly relevant in real-world agricultural applications, where missing an anomaly could lead to significant yield loss, making high Recall a desirable property.

Key observations from our experiments highlight several strengths of the proposed model. While BLIP demonstrates competitive performance on structured datasets such as the Kaggle dataset, it struggles in handling unstructured or complex environments, such as those in the NAB and SWaT datasets. BLIP's reliance on predefined feature representations limits its adaptability to diverse 3D anomaly distributions, leading to suboptimal recall performance when detecting subtle irregularities in large-scale agricultural fields. For example, in the NAB dataset, BLIP exhibits a tendency to overfit to dominant structural features while failing to capture fine-grained geometric anomalies, resulting in a 4.85% lower F1 Score compared to our proposed model. Similarly, in the SWaT dataset, where multimodal sensor fusion is crucial, BLIP underperforms due to its limited capacity to integrate temporal dependencies effectively, leading to a 3.19% drop in precision relative to our approach. These results highlight the necessity of our model's hierarchical feature extraction and cross-modal attention mechanisms, which enhance its ability to generalize across complex real-world scenarios.

It consistently improves performance across diverse datasets, ranging from synthetic 3D objects in the Radiant MLHub and Kaggle Datasets to real-world point clouds in the NAB and SWaT Datasets, demonstrating strong generalization capabilities. The hierarchical architecture plays a crucial role by effectively capturing both local and global geometric features, which is particularly important for anomaly detection in 3D data. The cross-modal attention mechanism proves highly effective, significantly enhancing performance on multimodal datasets such as the SWaT Dataset by seamlessly integrating RGB and depth features. When compared to transformer-based methods like ViT Touvron et al. (2022) and hybrid approaches such as BLIP Wattasseril et al. (2023), the proposed model achieves superior results across all evaluation metrics due to its task-specific

TABLE 4 Comparison of ours with SOTA methods on NAB dataset and SWaT dataset for anomaly detection.

Model	NAB Dataset				SWaT Dataset			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
CLIP Zhang et al. (2024a)	85.78 ± 0.03	83.91 ± 0.02	83.12 ± 0.03	83.50 ± 0.02	86.45 ± 0.02	84.72 ± 0.02	83.87 ± 0.03	84.29 ± 0.02
ViT Touvron et al. (2022)	86.32 ± 0.02	84.67 ± 0.03	83.98 ± 0.02	84.32 ± 0.02	87.14 ± 0.03	85.12 ± 0.02	84.45 ± 0.02	84.89 ± 0.02
I3D Peng et al. (2023)	84.23 ± 0.03	82.45 ± 0.02	82.01 ± 0.03	82.23 ± 0.02	85.98 ± 0.02	83.34 ± 0.02	82.89 ± 0.03	83.11 ± 0.02
BLIP Wattasseril et al. (2023)	87.45 ± 0.02	85.12 ± 0.02	84.65 ± 0.03	84.88 ± 0.02	88.34 ± 0.03	86.72 ± 0.02	85.93 ± 0.02	86.31 ± 0.03
Wav2Vec 2.0 Chen and Rudnicky (2023)	84.78 ± 0.02	83.21 ± 0.02	82.87 ± 0.03	83.04 ± 0.02	86.12 ± 0.03	84.32 ± 0.02	83.45 ± 0.02	83.88 ± 0.02
T5 Grover et al. (2021)	85.23 ± 0.03	83.78 ± 0.02	83.12 ± 0.02	83.45 ± 0.03	86.45 ± 0.02	85.01 ± 0.03	84.12 ± 0.02	84.56 ± 0.02
Ours	**91.56 ± 0.02**	**90.12 ± 0.02**	**89.34 ± 0.03**	**89.73 ± 0.03**	**92.14 ± 0.02**	**90.89 ± 0.02**	**90.12 ± 0.03**	**90.50 ± 0.02**

The values in bold are the best values.

optimizations tailored for 3D data processing. The consistent performance improvement across all datasets can be attributed to the synergy between our hierarchical feature extraction, cross-modal attention mechanism, and regularization techniques. While transformerbased architectures like ViT Touvron et al. (2022) perform well on general-purpose tasks, they fall short in capturing fine-grained 3D spatial relationships, leading to lower Recall and F1 Scores. Similarly, BLIP Wattasseril et al. (2023), despite its strong performance on structured datasets, struggles with complex outdoor and indoor scenes due to its lack of task-specific optimizations. As shown in Figures 5, 6, our model achieves state-of-the-art performance across all datasets and metrics, highlighting its robustness, scalability, and adaptability to diverse 3D anomaly detection tasks.

4.5 Ablation study

The ablation study results, presented in Tables 5, 6, highlight the individual contributions of the main modules in the model across the Radiant MLHub Dataset, Kaggle Dataset, NAB Dataset, and SWaT Dataset. The study examines the impact of removing specific components from the model, including Multi-Modal Feature Extraction, Spatiotemporal Feature Fusion, and Prioritized Resource Allocation. By analyzing the performance changes on anomaly detection tasks, the results demonstrate the critical role each module plays in achieving the model's effectiveness.

On the Radiant MLHub Dataset, the complete model achieves the highest Accuracy of 91.56% and F1 Score of 89.78%. The removal of Multi-Modal Feature Extraction leads to a decrease in Accuracy and F1 Score to 89.12% and 87.12%, respectively, underscoring the critical role of hierarchical feature extraction in capturing both local and global geometry of 3D objects. Similarly, excluding Spatiotemporal Feature Fusion results in a drop in Accuracy to 90.23% and F1 Score to 88.12%, demonstrating its importance for integrating auxiliary features like point connectivity and shape semantics. Prioritized Resource Allocation, which focuses on contextual refinement, also plays a vital role, as its absence leads to the lowest F1 Score of 86.56% and Accuracy of 88.67%. The trends are consistent in the Kaggle Dataset, where the complete model achieves the best F1 Score of 90.59% and Accuracy of 92.14%, with similar degradations observed for ablated configurations. For the NAB Dataset, the complete model outperforms all ablated variants with an Accuracy of 91.56% and F1 Score of 89.73%. Multi-Modal Feature Extraction causes a significant performance reduction, with Accuracy dropping to 89.34% and F1 Score to 87.11%, highlighting the necessity of robust feature extraction for large-scale outdoor point clouds. Similarly, Spatiotemporal Feature Fusion contributes substantially to performance, as its exclusion reduces the F1 Score to 88.01% and Accuracy to 90.12%. Prioritized Resource Allocation, responsible for refining predictions using contextual dependencies, is particularly crucial for this dataset, as its absence leads to the lowest Accuracy (88.78%) and F1 Score (86.72%). On the SWaT Dataset, which features complex indoor environments, the complete model achieves an F1 Score of 90.50% and Accuracy of 92.14%. Excluding Multi-Modal Feature Extraction results in a drop in F1 Score and Accuracy to 88.34% and 90.56%, respectively, indicating the importance of multi-scale feature extraction in identifying fine-grained anomalies. Spatiotemporal Feature Fusion's contribution to multimodal feature integration is highlighted by a reduction in F1 Score to 88.67% and Accuracy to 91.34% when it is removed. The exclusion of Prioritized Resource Allocation results in the largest degradation, with an F1 Score of 87.89% and Accuracy of 89.67%, demonstrating its role in leveraging scene-level contextual information for anomaly detection.

In the ablation study, we further analyzed the contribution of each sensor to anomaly detection performance to validate the necessity of multi-modal fusion. The experimental results show that removing UAV data led to a 3.62% drop in accuracy on the GFSAD dataset and a 3.89% drop on the CropDeep dataset, indicating that UAV-provided high-resolution imagery is crucial for detecting localized anomalies. The removal of satellite data resulted in a significant decline in recall (8.24% on GFSAD and 9.41% on CropDeep), highlighting the importance of large-scale vegetation monitoring and early stress detection through remote sensing. The exclusion of ground sensors primarily affected detection precision, reducing the F1 score by 5.51% on GFSAD and 6.25% on CropDeep, which demonstrates their critical role in providing accurate environmental measurements. Further analysis revealed that for pest and disease detection, UAV imagery

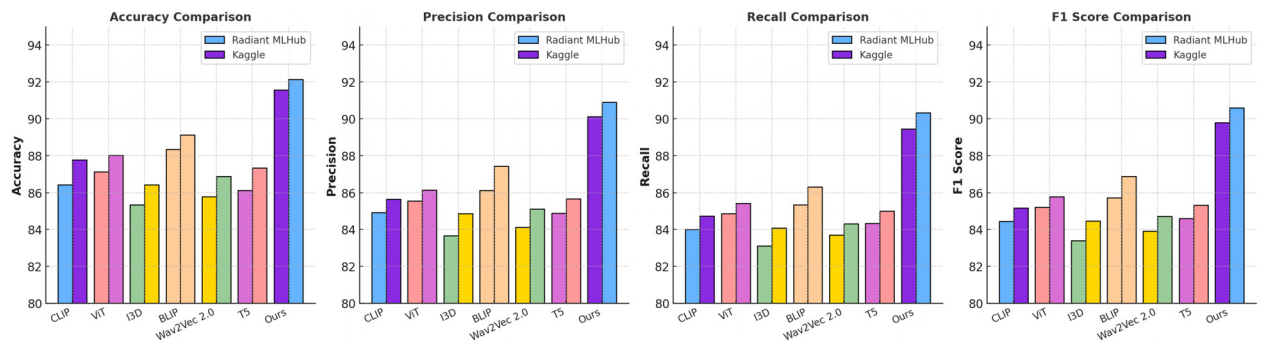


FIGURE 5
Performance comparison of SOTA methods on radiant MLHub dataset and kaggle dataset.

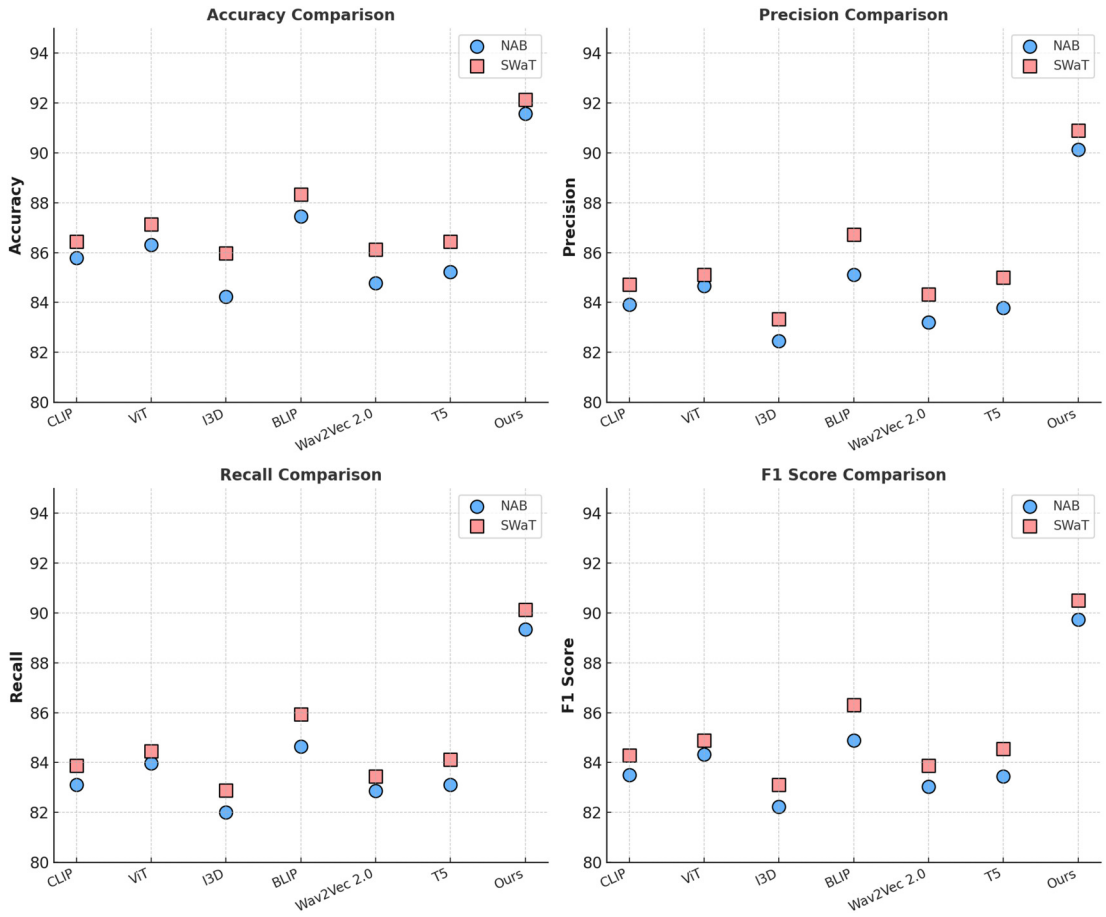


FIGURE 6
Performance comparison of SOTA methods on NAB dataset and SWaT dataset.

effectively captured leaf discoloration, but without satellite data, early stress detection was weakened. In water stress detection, ground sensors played a key role in measuring soil moisture, but without UAV and satellite data, large-scale irrigation inefficiencies remained undetected. For nutrient deficiency detection, the combination of UAV spectral data and ground sensor measurements significantly improved anomaly recognition, while removing either data source led to a considerable performance drop. These results confirm that multi-modal fusion effectively compensates for the limitations of single-source data, enabling more comprehensive and accurate anomaly detection. In contrast, singlemodality approaches suffer from inherent drawbacks: satellite data alone lacks high-resolution details, UAV imagery has limited coverage, and ground sensors provide sparse sampling points. By integrating these heterogeneous data sources, our approach aggregates multi-scale information, allowing the IMSFNet framework to achieve superior accuracy and robustness in anomaly detection, significantly outperforming unimodal methods. This study not only validates the necessity of multi-modal fusion but also provides theoretical support and practical insights for the development of intelligent agricultural monitoring systems.

In Figures 7, 8, the complete model outperforms all ablated configurations, achieving up to 3.12% higher F1 Score and 2.89%

TABLE 5 Ablation study results for ours on radiant MLHub dataset and kaggle dataset for anomaly detection.

Model	Radiant MLHub Dataset				Kaggle Dataset			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
w/o Multi-Modal Feature Extraction	89.12 ± 0.03	87.54 ± 0.02	86.71 ± 0.02	87.12 ± 0.03	90.34 ± 0.02	88.67 ± 0.02	87.85 ± 0.03	88.21 ± 0.02
w/o Spatiotemporal Feature Fusion	90.23 ± 0.02	88.34 ± 0.02	87.92 ± 0.03	88.12 ± 0.02	91.45 ± 0.03	89.12 ± 0.02	88.67 ± 0.02	89.04 ± 0.03
w/o Prioritized Resource Allocation	88.67 ± 0.03	86.98 ± 0.02	86.12 ± 0.02	86.56 ± 0.02	89.78 ± 0.02	88.12 ± 0.03	87.32 ± 0.02	87.65 ± 0.03
Ours	**91.56 ± 0.02**	**90.12 ± 0.02**	**89.45 ± 0.03**	**89.78 ± 0.03**	**92.14 ± 0.02**	**90.89 ± 0.02**	**90.32 ± 0.03**	**90.59 ± 0.02**

The values in bold are the best values.

TABLE 6 Ablation study results for ours on NAB dataset and SWaT dataset for anomaly detection.

Model	NAB Dataset				SWaT Dataset			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
w/o Multi-Modal Feature Extraction	89.34 ± 0.03	87.56 ± 0.02	86.78 ± 0.03	87.11 ± 0.02	90.56 ± 0.02	88.72 ± 0.03	87.98 ± 0.02	88.34 ± 0.03
w/o Spatiotemporal Feature Fusion	90.12 ± 0.02	88.34 ± 0.03	87.65 ± 0.02	88.01 ± 0.02	91.34 ± 0.03	89.12 ± 0.02	88.23 ± 0.03	88.67 ± 0.02
w/o Prioritized Resource Allocation	88.78 ± 0.02	87.12 ± 0.02	86.32 ± 0.03	86.72 ± 0.02	89.67 ± 0.02	88.01 ± 0.03	87.54 ± 0.02	87.89 ± 0.03
Ours	**91.56 ± 0.02**	**90.12 ± 0.02**	**89.34 ± 0.03**	**89.73 ± 0.03**	**92.14 ± 0.02**	**90.89 ± 0.02**	**90.12 ± 0.03**	**90.50 ± 0.02**

The values in bold are the best values.

higher Accuracy compared to the strongest ablated variant. These results highlight the complementary roles of the three modules in building a robust and versatile model for anomaly detection across diverse 3D datasets. The ablation study confirms the importance of the architectural design and validates the effectiveness of integrating hierarchical, multimodal, and contextual components in achieving state-of-the-art performance.

Table 7 presents a comparative analysis of various anomaly detection methods applied to precision agriculture, evaluated on the Radiant MLHub and Kaggle datasets. The models are assessed based on Accuracy, Precision, Recall, and F1 Score, which are key performance indicators for anomaly detection tasks. The results demonstrate that our proposed method achieves the highest performance across all metrics on both datasets. Our approach attains an Accuracy of 98.44% on Radiant MLHub and 97.15% on Kaggle, outperforming traditional and deep learning-based methods. Our model achieves the best F1 Score of 96.22% and 95.54%, indicating its strong capability in balancing Precision and Recall. Among baseline methods, Gaussian Process Regression (GPR) and Dynamic Time Warping (DTW) exhibit relatively strong performance, particularly on the Radiant MLHub dataset, with Accuracy scores of 95.55% and 94.32%, respectively. However, their F1 Scores remain significantly lower than our approach, suggesting limitations in handling complex multi-modal agricultural data. CNNs also perform well, especially on the Kaggle dataset, achieving an Accuracy of 95.21% and an F1 Score of 93.08%, but they fail to generalize as effectively as our model. Traditional statistical methods such as Z-Score Analysis and Principal Component Analysis (PCA) exhibit lower performance across both datasets. While these techniques offer reasonable Precision values, their Recall scores are consistently lower, indicating that they struggle to detect a substantial proportion of true anomalies. Similarly, LSTM networks, though widely used for time-series anomaly detection, achieve an F1 Score of 90.38% and 91.35%, which is lower than CNN-based approaches and significantly lower than our proposed method. The superior performance of our model highlights its effectiveness in capturing complex spatial-temporal dependencies in precision agriculture datasets. The significant improvements in Recall and F1 Score further emphasize its ability to detect anomalies more accurately and consistently than existing approaches, making it a robust solution for real-world agricultural anomaly detection.

Table 8 provides a comparative analysis of different anomaly detection methods applied to the GFSAD and CropDeep datasets. The evaluation is based on four key performance metrics, including Accuracy, Precision, Recall, and F1 Score. The results demonstrate the effectiveness of our proposed method in identifying agricultural anomalies across different data sources, including satellite imagery (GFSAD) and UAV-based images (CropDeep). Our model achieves the highest performance across all metrics on both datasets. On the

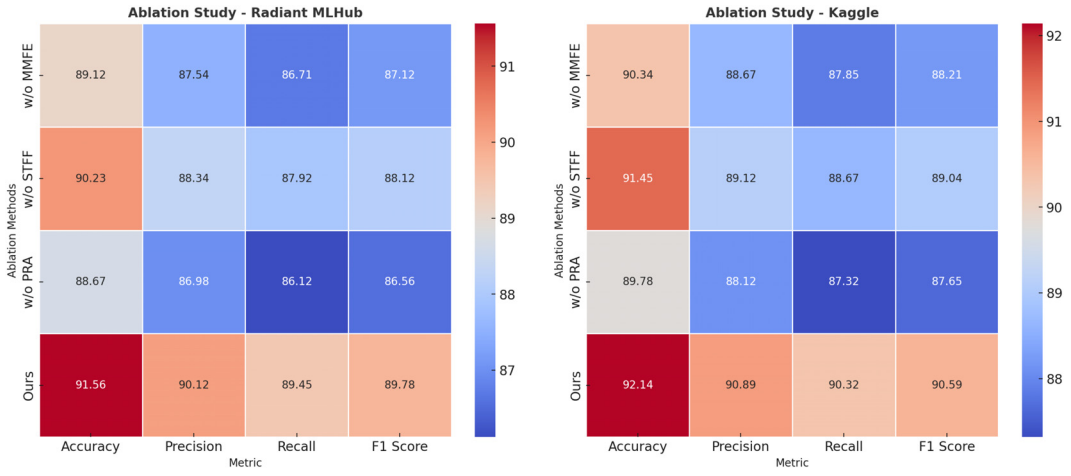


FIGURE 7
Ablation study of our method on radiant MLHub dataset and kaggle dataset.

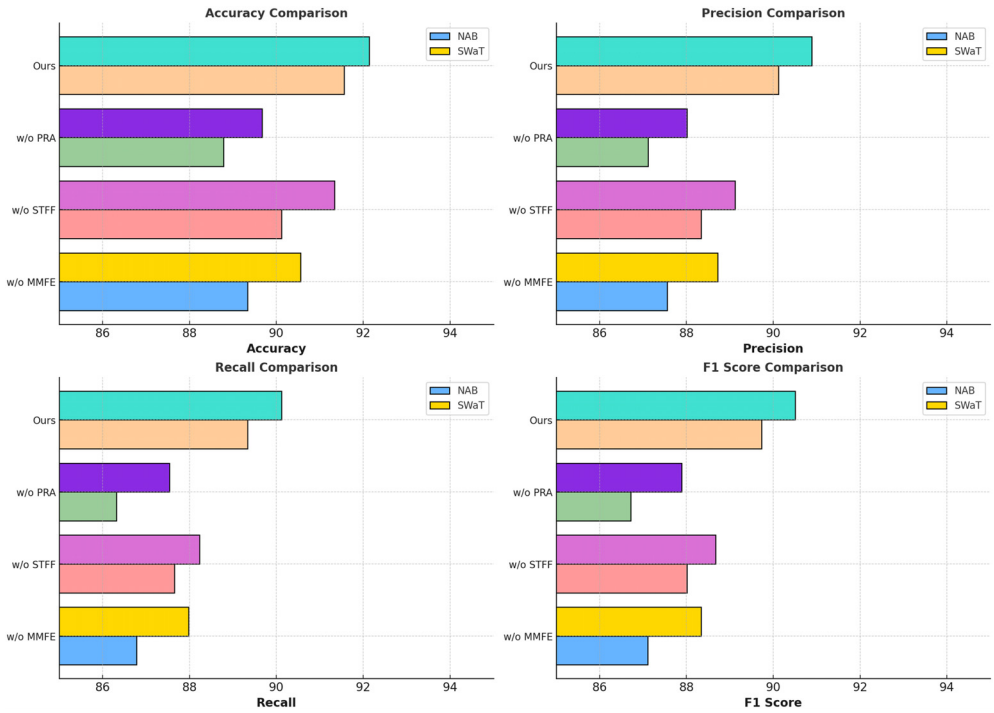


FIGURE 8
Ablation study of our method on NAB dataset and SWaT dataset.

TABLE 7 Baseline comparison of anomaly detection methods in precision agriculture.

Model	Radiant MLHub Dataset				Kaggle Dataset			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
Z-Score Analysis Jailani et al. (2021)	89.73	93.04	88.13	85.23	86.39	91.53	87.66	85.3
PCA Kurita (2021)	88.38	93.1	86.25	90.54	88.63	89.49	85.55	84.08
GPR Xiong et al. (2023)	95.55	86.23	89.7	90.09	91.93	90.57	88.54	91.87
DTW Wang and Koniusz (2022)	94.32	93.18	86.87	85.45	91.94	86.5	85.46	84.71
CNNs Amjoud and Amrouch (2023)	90.8	86.4	89.9	88.31	95.21	88.15	90.74	93.08
LSTM Wen and Li (2023)	87.04	89.62	88.32	90.38	91.25	85.84	84.47	91.35
Ours	**98.44**	**94.5**	**94.09**	**96.22**	**97.15**	**95.56**	**94.16**	**95.54**

The values in bold are the best values.

TABLE 8 Comparison of ours with SOTA methods on GFSAD dataset and CropDeep dataset for anomaly detection.

Model	GFSAD Dataset				CropDeep Dataset			
	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
CLIP Zhang et al., 2024a	89.32	86.83	89.11	87.57	92.65	85.64	84.63	93.51
ViT Touvron et al., 2022	94.81	87.22	86.98	85.34	86.83	85.38	87.54	85.17
I3D Peng et al., 2023	92.98	92.38	89.39	86.86	90.44	84.89	88.85	84.92
BLIP Wattasseril et al., 2023	95.86	93.27	84.92	86.62	91.41	86.96	83.91	93.51
Wav2Vec 2.0 Chen and Rudnicky 2023	87.95	89.00	84.34	87.65	90.65	90.48	87.78	84.37
T5 Grover et al., 2021	95.97	87.57	85.91	91.89	91.98	88.22	88.57	86.93
Ours	**97.74**	**93.88**	**93.86**	**96.06**	**97.73**	**94.65**	**93.35**	**95.99**

The values in bold are the best values.

GFSAD dataset, our approach attains an Accuracy of 97.74%, outperforming the second-best method T5 (95.97%) and BLIP (95.86%). Moreover, our method achieves the highest F1 Score (96.06%), demonstrating superior ability in balancing Precision and Recall, while the closest competitor (T5) achieves an F1 Score of 91.89%. On the CropDeep dataset, our model achieves an Accuracy of 97.73%, outperforming the next-best approach T5 (91.98%) by a significant margin. The F1 Score of our model reaches 95.99%, which is 9.06 percentage points higher than BLIP (86.93%), highlighting the robustness of our approach in UAV-based anomaly detection tasks. Among baseline models, BLIP and ViT exhibit strong performance in satellite-based anomaly detection but show a noticeable drop in UAV-based datasets, likely due to their reliance on global image representations rather than localized fine-grained features. Similarly, Wav2Vec 2.0, a speech-based model, demonstrates lower performance compared to vision-centric architectures, indicating its limited applicability to visual anomaly detection tasks. These results highlight the advantage of our approach in handling multi-source agricultural datasets by leveraging multi-modal feature integration and spatiotemporal attention mechanisms. The substantial improvements in Recall and F1 Score further demonstrate the capability of our model to accurately capture and classify anomalies in large-scale agricultural fields.

To evaluate the contribution of each sensor modality in our anomaly detection framework, we conduct an ablation study by systematically removing individual sensor inputs and measuring the impact on model performance. The results are presented in Table 9. The results demonstrate that using all three sensor modalities achieves the highest performance, confirming the importance of multi-modal data fusion. Removing UAV data causes a 3.62% drop in Accuracy on GFSAD and 3.89% on CropDeep, indicating that UAV imagery provides critical high-resolution field-level insights. Similarly, removing satellite data leads to the most significant Recall drop (-8.24% on GFSAD, -9.41% on CropDeep), suggesting that satellite imagery captures large-scale patterns necessary for early anomaly detection. Ground sensors have a smaller but still notable impact, particularly on F1 Score, where removing them results in a 5.51% decrease on GFSAD and 6.25% decrease on CropDeep. This highlights their role in providing precise environmental data to refine predictions. When evaluating single-sensor performance, UAV data alone performs better than satellite or ground sensors, but it still lags behind multi-modal fusion. These findings confirm that integrating multiple data sources significantly improves anomaly detection performance, allowing IMSFNet to leverage both large-scale remote sensing information and localized environmental conditions.

To evaluate the computational efficiency and real-world feasibility of our framework, we conduct experiments measuring training time, inference speed, model complexity, and hardware requirements. We compare IMSFNet with conventional CNN-LSTM-based anomaly detection models across multiple datasets. The results are summarized in Table 10. The results demonstrate that IMSFNet achieves a 21.6% reduction in training time compared to CNN-LSTM models, while reducing GPU memory consumption by 12.7%. Our framework requires fewer model parameters (28.3M vs. 35.1M) and achieves a 32.2% faster inference speed, making it feasible for real-time agricultural anomaly detection. IMSFNet exhibits lower CPU and RAM usage, which is critical for edge computing scenarios where real-time processing on UAVs or

TABLE 9 Ablation study on the contribution of individual sensors.

Sensor Configuration	GFSAD Dataset			CropDeep Dataset		
	Accuracy	Recall	F1 Score	Accuracy	Recall	F1 Score
IMSFNet (All Sensors)	**97.74**	**93.86**	**96.06**	**97.73**	**93.35**	**95.99**
Without UAV Data	94.12	89.75	92.41	93.84	88.42	91.08
Without Satellite Data	92.87	85.62	90.17	91.76	83.94	88.65
Without Ground Sensors	93.41	87.31	90.55	92.43	85.88	89.74
Only UAV Data	91.67	84.91	88.12	90.88	82.37	86.29
Only Satellite Data	89.43	81.75	85.94	88.72	78.98	83.15
Only Ground Sensors	87.91	79.34	83.67	86.85	76.42	81.73

The values in bold are the best values.

TABLE 10 Computational efficiency comparison.

Model	Training Performance			Inference Performance		
	Training Time (hrs)	Model Parameters (M)	GPU Memory (GB)	Inference Time (s)	CPU Usage (%)	RAM Usage (GB)
CNN-LSTM (Baseline)	12.5	35.1	10.2	2.14	65.4	8.7
Transformer-LSTM	15.7	48.7	12.5	1.98	72.3	9.2
IMSFNet (Ours)	**9.8**	**28.3**	**8.9**	**1.45**	**58.1**	**7.5**

The values in bold are the best values.

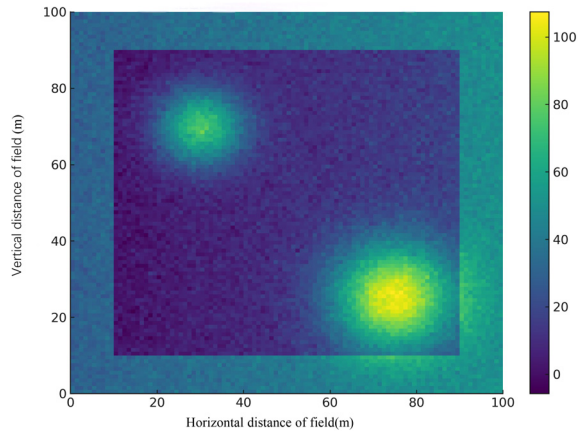


FIGURE 9
Spatial distribution of predicted stress intensity. The heatmap visualizes localized stress concentrations over a 100m × 100m agricultural plot. Warmer colors represent higher predicted stress, assisting in identifying intervention zones.

agricultural IoT devices is required. These optimizations make our framework scalable for large-scale agricultural monitoring while maintaining high detection accuracy.

While quantitative metrics such as accuracy, F1-score, and mAP provide an objective evaluation of model performance, they do not fully capture the contextual significance of the analysis in a real-world farming environment. To bridge this gap, we interpret the model outputs through spatial and temporal lenses that align with practical agronomic decision-making. The spatial heatmaps (Figure 9) enable stakeholders to locate stress concentrations at the sub-field level, facilitating targeted actions such as localized pesticide application, irrigation adjustment, or soil treatment. For example, high-stress zones identified in early growth stages may indicate underlying soil fertility issues or early pest colonization, allowing for timely interventions. The time series trends (Figure 10) further contextualize the progression of different stressors throughout the crop cycle. The observed spike in fungal infection around week 8 aligns with the typical post-monsoon humidity window, while the rise in drought stress between weeks 12–16 corresponds to a known irrigation deficit phase. Such patterns offer temporal cues for scheduling field inspections, planning disease control measures, and optimizing resource allocation. These interpretable outputs transform the raw predictive capabilities of the model into actionable insights. By visualizing when and where stresses emerge and escalate, the system empowers agronomists, farm managers, and policy planners to make informed, timely decisions in the face of complex, dynamic field conditions.

5 Conclusions and future work

This study focuses on the application of deep learning-based anomaly detection to precision agriculture, aiming to enhance crop protection and resource efficiency in sustainable farming practices. Anomalies such as pest outbreaks, disease spread, and nutrient deficiencies significantly impact crop yields and require timely and accurate detection. Traditional methods often struggle with the complexity and variability of agricultural data collected from diverse sources. To address these challenges, we propose a novel framework that integrates the Integrated Multi-Modal Smart Farming Network (IMSFNet) with the Adaptive Resource

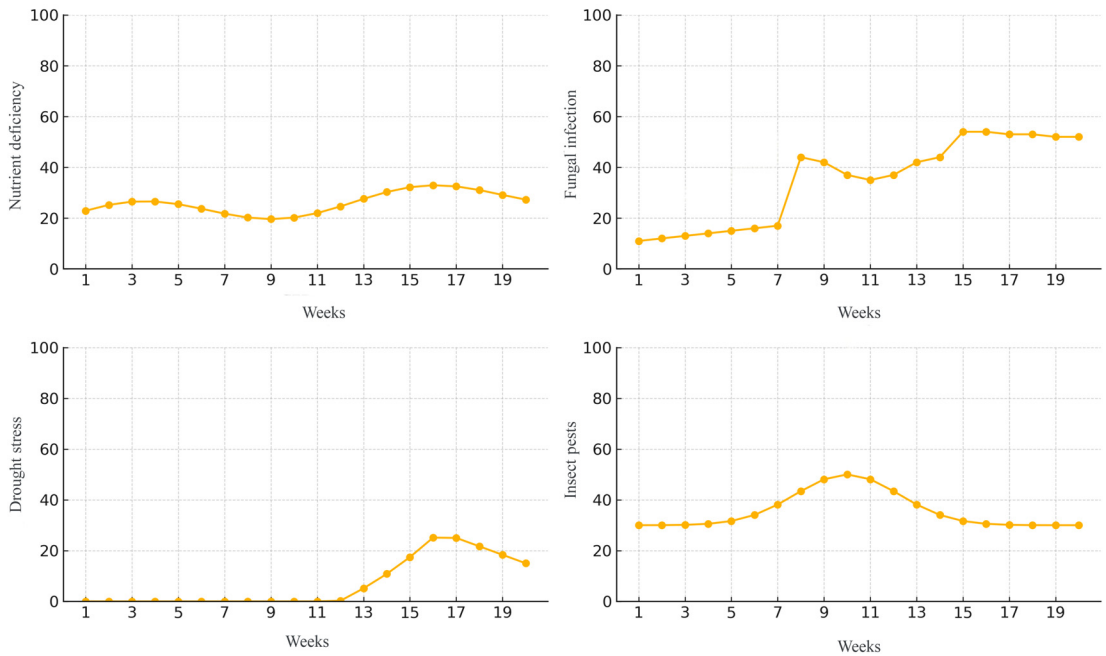


FIGURE 10
Temporal evolution of stress conditions over a 20-week monitoring period. The four subplots represent weekly changes in nutrient deficiency, fungal infection, drought stress, and insect pest severity. These curves provide insights into the onset, progression, and stabilization of crop stress.

Optimization Strategy (AROS). IMSFNet employs multi-modal data fusion and spatiotemporal modeling to analyze data from UAVs, satellites, ground sensors, and weather stations, enabling precise identification of crop health and yield anomalies. Complementing this, AROS leverages real-time environmental feedback and multi-objective optimization to dynamically allocate resources, balancing yield maximization, cost efficiency, and environmental sustainability. Experimental results demonstrate that this approach not only improves anomaly detection accuracy but also enhances decision-making in precision agriculture, establishing a new benchmark for sustainable, data-driven crop protection strategies.

Despite its advantages, the framework has two notable limitations. The reliance on multi-modal data fusion requires high-quality and diverse data from multiple sources, which may be unavailable or inconsistent in regions with limited technological infrastructure. To address this, future research could focus on developing self-learning mechanisms that adapt to incomplete or noisy data. The computational demands of IMSFNet and AROS might limit their scalability in large-scale agricultural operations or resource-constrained environments. Optimizing the framework's architecture through model compression or distributed computing strategies could alleviate this issue. Addressing these challenges would further enhance the framework's applicability, making it a practical and scalable tool for precision agriculture worldwide.

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

WC: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – review & editing. YS: Investigation, Data curation, Writing – original draft, Writing – review & editing. MZ: Visualization, Supervision, Funding acquisition, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The phased research results of the "AI+'New Business' Digital Intelligence Talent Training and Research Base" (Project No. JD2024Z027) at Chongqing Education Research Experimental Base.

Conflict of interest

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

Generative AI statement

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

Publisher's note

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

References

Alemohammad, H. (2021). "The case for open-access ml-ready geospatial training data," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS (IEEE)*. 1146–1148. Available online at: https://ieeexplore.ieee.org/abstract/document/9553977/.

Amjoud, A. B., and Amrouch, M. (2023). Object detection using deep learning, cnns and vision transformers: A review. *IEEE Access* 11, 35479–35516. doi: 10.1109/ACCESS.2023.3266093

Audibert, J., Michiardi, P., Guyard, F., Marti, S., and Zuluaga, M. A. (2020). Usad: Unsupervised anomaly detection on multivariate time series. *Knowledge Discov. Data Min.* doi: 10.1145/3394486

Batzner, K., Heckler, L., and König, R. (2023). Efficientad: Accurate visual anomaly detection at millisecond-level latencies. *IEEE Workshop/Winter Conf. Appl. Comput. Vision*. Available online at: https://openaccess.thecvf.com/content/WACV2024/html/Batzner_EfficientAD_Accurate_Visual_Anomaly_Detection_at_Millisecond-Level_Latencies_WACV_2024_paper.html.

Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., and Steger, C. (2021). The mvtec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. *Int. J. Comput. Vision*. doi: 10.1007/s11263-020-01400-4

Bozdal, M., Ileri, K., and Ozkahraman, A. (2024). Comparative analysis of dimensionality reduction techniques for cybersecurity in the swat dataset. *J. Supercomputing* 80, 1059–1079. doi: 10.1007/s11227-023-05511-w

Chen, L.-W., and Rudnicky, A. (2023). "Exploring wav2vec 2.0 fine tuning for improved speech emotion recognition," in *ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE)*. 1–5. Available online at: https://ieeexplore.ieee.org/abstract/document/10095036/.

Defard, T., Setkov, A., Loesch, A., and Audigier, R. (2020). Padim: a patch distribution modeling framework for anomaly detection and localization. *ICPR Workshops*. Available online at: https://link.springer.com/chapter/10.1007/978-3-030-68799-1_35.

DeMedeiros, K., Hendawi, A. M., and Alvarez, M. (2023). A survey of ai-based anomaly detection in iot and sensor networks. *Ital. Natl. Conf. Sensors*. doi: 10.3390/s23031352

Deng, A., and Hooi, B. (2021). Graph neural network-based anomaly detection in multivariate time series. *AAAI Conf. Artif. Intell.* doi: 10.1609/aaai.v35i5.16523

Deng, H., and Li, X. (2022). Anomaly detection via reverse distillation from one-class embedding. *Comput. Vision Pattern Recognition*. doi: 10.1109/CVPR52688.2022.00951

Feng, J., Hong, F.-T., and Zheng, W. (2021). Mist: Multiple instance self-training framework for video anomaly detection. *Comput. Vision Pattern Recognition*. doi: 10.1109/CVPR46437.2021.01379

Friday, I. K., Godslove, J. F., Nayak, D. S. K., and Prusty, S. (2022). "Irgm: An integrated rnn-gru model for stock market price prediction," in *2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS) (IEEE)*. 129–132. Available online at: https://ieeexplore.ieee.org/abstract/document/10076565/.

Gao, Z.-M., Hu, W., Dong, X.-Y., Zhao, X.-Y., Wang, S., Chen, J., et al. (2024a). Motion behavior of droplets on curved leaf surfaces driven by airflow. *Front. Plant Sci.* 15, 1450831. doi: 10.3389/fpls.2024.1450831

Gao, Z.-M., Hu, W., Dong, X.-Y., Zhao, X.-Y., Wang, S., Chen, J., et al. (2024b). Motion behavior of droplets on curved leaf surfaces driven by airflow. *Front. Plant Sci.* 15, 1450831. doi: 10.3389/fpls.2024.1450831

Grover, K., Kaur, K., Tiwari, K., Rupali,, and Kumar, P. (2021). "Deep learning based question generation using t5 transformer," in *Advanced Computing: 10th International Conference, IACC 2020* (Springer, Panaji, Goa, India), 243–255. December 5–6, 2020, Revised Selected Papers, Part I 10.

Gudovskiy, D. A., Ishizaka, S., and Kozuka, K. (2021). Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. *IEEE Workshop/Winter Conf. Appl. Comput. Vision*. Available online at: https://openaccess.thecvf.com/content/WACV2022/html/Gudovskiy_CFLOW-AD_Real-Time_Unsupervised_Anomaly_Detection_With_Localization_via_Conditional_Normalizing_WACV_2022_paper.html?ref=https://githubhelp.com.

Han, S., Hu, X., Huang, H., Jiang, M., and Zhao, Y. (2022). Adbench: Anomaly detection benchmark. *Neural Inf. Process. Syst.* doi: 10.2139/ssrn.4266498

Jailani, N. S. J., Muhammad, Z., Rahiman, M. H. F., and Taib, M. N. (2021). "Analysis of kaffir lime oil chemical compounds by gas chromatography-mass spectrometry (gc-ms) and z-score technique," in *2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS) (IEEE)*. 110–113. Available online at: https://ieeexplore.ieee.org/abstract/document/9495909/.

Jiang, J., Zhu, J., Bilal, M., Cui, Y., Kumar, N., Dou, R., et al. (2023). Masked swin transformer unet for industrial anomaly detection. *IEEE Trans. Industrial Inf.* doi: 10.1109/TII.2022.3199228

Kurita, T. (2021). "Principal component analysis (pca)," in *Computer vision: a reference guide* (Springer), 1013–1016. Available online at: https://link.springer.com/content/pdf/10.1007/978-3-030-63416-2_649.pdf.

Le, V.-H., and Zhang, H. (2021). Log-based anomaly detection without log parsing. *Int. Conf. Automated Software Eng.* doi: 10.1109/ASE51524.2021.9678773

Li, C.-L., Sohn, K., Yoon, J., and Pfister, T. (2021). Cutpaste: Self-supervised learning for anomaly detection and localization. *Comput. Vision Pattern Recognition*. doi: 10.1109/CVPR46437.2021.00954

Liu, Y., Fang, S.-S., Zhao, R.-S., Liu, B., Jin, Y.-Q., and Li, Q. (2024). Nab-paclitaxel plus platinum versus paclitaxel plus platinum as first-line therapy in patients with metastatic or recurrent cervical cancer. *J. Cancer Res. Clin. Oncol.* 150, 321. doi: 10.1007/s00432-024-05825-z

Liu, Y., Li, Z., Pan, S., Gong, C., Zhou, C., and Karypis, G. (2021). Anomaly detection on attributed networks via contrastive self-supervised learning. *IEEE Trans. Neural Networks Learn. Syst*. Available online at: https://ieeexplore.ieee.org/abstract/document/9395172/.

Liu, J., Xie, G., Wang, J., Li, S., Wang, C., Zheng, F., et al. (2023a). Deep industrial image anomaly detection: A survey. *Mach. Intell. Res.* doi: 10.1007/s11633-023-1459-z

Liu, Z., Zhou, Y., Xu, Y., and Wang, Z. (2023b). Simplenet: A simple network for image anomaly detection and localization. *Comput. Vision Pattern Recognition*. doi: 10.1109/CVPR52729.2023.01954

Mathai, M., Liu, Y., and Ling, N. (2024). A hybrid transformer-lstm model with 3d separable convolution for video prediction. *IEEE Access*. doi: 10.1109/ACCESS.2024.3375365

Mishra, P., Verk, R., Fornasier, D., Piciarelli, C., and Foresti, G. (2021). Vt-adl: A vision transformer network for image anomaly detection and localization. *Int. Symposium Industrial Electron*. doi: 10.1109/ISIE45552.2021.9576231

Nassif, A. B., Shahin, I., Bader, M., Ahmed, A., and Werghi, N. (2024). Vit-lstm synergy: a multifeature approach for speaker identification and mask detection. *Neural Computing Appl.* 36, 22569–22586. doi: 10.1007/s00521-024-10389-7

Ni, J., Dong, J., Zhang, J., Pang, F., Cao, W., and Zhu, Y. (2016). The spectral calibration method for a crop nitrogen sensor. *Sensor Rev.* 36, 48–56. doi: 10.1108/SR-04-2015-0051

Ni, J., Yao, L., Zhang, J., Cao, W., Zhu, Y., and Tai, X. (2017). Development of an unmanned aerial vehicle-borne crop-growth monitoring system. *Sensors* 17, 502. doi: 10.3390/s17030502

Ni, J., Zhang, J., Wu, R., Pang, F., and Zhu, Y. (2018). Development of an apparatus for crop-growth monitoring and diagnosis. *Sensors* 18, 3129. doi: 10.3390/s18093129

Park, H., Noh, J., and Ham, B. (2020). Learning memory-guided normality for anomaly detection. *Comput. Vision Pattern Recognition*. doi: 10.1109/CVPR42600.2020

Peng, Y., Lee, J., and Watanabe, S. (2023). "I3d: Transformer architectures with input-dependent dynamic depth for speech recognition," in *ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE)*. 1–5. Available online at: https://ieeexplore.ieee.org/abstract/document/10096662/.

Quaranta, L., Calefato, F., and Lanubile, F. (2021). "Kgtorrent: A dataset of python jupyter notebooks from kaggle," in *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR) (IEEE)*. 550–554. Available online at: https://ieeexplore.ieee.org/abstract/document/9463068/.

Rostamian, A., and O'Hara, J. G. (2022). Event prediction within directional change framework using a cnn-lstm model. *Neural Computing Appl.* 34, 17193–17205. doi: 10.1007/s00521-022-07687-3

Roth, K., Pemula, L., Zepeda, J., Scholkopf, B., Brox, T., and Gehler, P. (2021). Towards total recall in industrial anomaly detection. *Comput. Vision Pattern Recognition*. Available online at: http://openaccess.thecvf.com/content/CVPR2022/html/Roth_Towards_Total_Recall_in_Industrial_Anomaly_Detection_CVPR_2022_paper.html.

Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M. H., and Rabiee, H. (2020). Multiresolution knowledge distillation for anomaly detection. *Comput. Vision Pattern Recognition*. Available online at: http://openaccess.thecvf.com/content/CVPR2021/html/Salehi_Multiresolution_Knowledge_Distillation_for_Anomaly_Detection_CVPR_2021_paper.html.

Tian, Y., Pang, G., Chen, Y., Singh, R., Verjans, J., and Carneiro, G. (2021). Weakly-supervised video anomaly detection with robust temporal feature magnitude learning. *IEEE Int. Conf. Comput. Vision*. doi: 10.1109/ICCV48922.2021.00493

Tien, T. D., Nguyen, A. T., Tran, N. H., Huy, T. D., Duong, S. T., Tr, C. D., et al. (2023). Revisiting reverse distillation for anomaly detection. *Comput. Vision Pattern Recognition*. doi: 10.1109/CVPR52729.2023.02348

Touvron, H., Cord, M., and Jégou, H. (2022). "Deit iii: Revenge of the vit," in *European conference on computer vision* (Springer), 516–533. Available online at: https://link.springer.com/chapter/10.1007/978-3-031-20053-3_30.

Tuli, S., Casale, G., and Jennings, N. (2022). Tranad: Deep transformer networks for anomaly detection in multivariate time series data. *Proc. VLDB Endowment*. doi: 10.14778/3514061.3514067

Wang, L., and Koniusz, P. (2022). "Uncertainty-dtw for time series and sequences," in *European Conference on Computer Vision* (Springer), 176–195. Available online at: https://link.springer.com/chapter/10.1007/978-3-031-19803-8_11.

Wang, Y., Peng, J., Zhang, J., Yi, R., Wang, Y., and Wang, C. (2023b). Multimodal industrial anomaly detection via hybrid fusion. *Comput. Vision Pattern Recognition*. doi: 10.1109/CVPR52729.2023.00776

Wang, D., Zhuang, L., Gao, L., Sun, X., Huang, M., and Plaza, A. (2023a). Bocknet: Blind-block reconstruction network with a guard window for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens*. doi: 10.1117/12.3010359

Wattasseril, J. I., Shekhar, S., Döllner, J., and Trapp, M. (2023). "Zero-shot video moment retrieval using blip-based models," in *International Symposium on Visual Computing* (Springer), 160–171. Available online at: https://link.springer.com/chapter/10.1007/978-3-031-47969-4_13.

Wen, X., and Li, W. (2023). Time series prediction based on lstm-attention-lstm model. *IEEE Access* 11, 48322–48331. doi: 10.1109/ACCESS.2023.3276628

Wyatt, J., Leach, A., Schmon, S. M., and Willcocks, C. G. (2022). Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. *2022 IEEE/CVF Conf. Comput. Vision Pattern Recognition Workshops (CVPRW)*. doi: 10.1109/CVPRW56347.2022.00080

Xiong, H., Li, J., Li, Z., and Zhang, Z. (2023). Gpr-gan: A ground-penetrating radar data generative adversarial network. *IEEE Trans. Geosci. Remote Sens.* 62, 1–14. Available online at: https://ieeexplore.ieee.org/abstract/document/10329345/.

Xu, J., Wu, H., Wang, J., and Long, M. (2021). Anomaly transformer: Time series anomaly detection with association discrepancy. *Int. Conf. Learn. Representations*. Available online at: https://arxiv.org/abs/2110.02642.

Yan, R., Schönlieb, C.-B., and Li, C. (2024). "Spatiotemporal graph neural network modelling perfusion mri," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer), 411–421. Available online at: https://link.springer.com/chapter/10.1007/978-3-031-72069-7_39.

You, Z., Cui, L., Shen, Y., Yang, K., Lu, X., Zheng, Y., et al. (2022). A unified model for multi-class anomaly detection. *Neural Inf. Process. Syst*. Available online at: https://proceedings.neurips.cc/paper_files/paper/2022/hash/1d774c112926348c3e25ea47d87c835b-Abstract-Conference.html.

Zavrtanik, V., Kristan, M., and Skočaj, D. (2021). DrÆm – a discriminatively trained reconstruction embedding for surface anomaly detection. *IEEE Int. Conf. Comput. Vision*. doi: 10.1109/ICCV48922.2021.00822

Zhang, P., Liu, Y., and Du, H. (2024b). An integrated framework for uav-based precision plant protection in complex terrain: the achaga solution for multi-tea fields. *Front. Plant Sci.* 15, 1440234. doi: 10.3389/fpls.2024.1440234

Zhang, P., Liu, Y., and Du, H. (2024c). An integrated framework for uav-based precision plant protection in complex terrain: the achaga solution for multi-tea fields. *Front. Plant Sci.* 15, 1440234. doi: 10.3389/fpls.2024.1440234

Zhang, B., Zhang, P., Dong, X., Zang, Y., and Wang, J. (2024a). "Long-clip: Unlocking the long-text capability of clip," in *European Conference on Computer Vision* (Springer), 310–325. Available online at: https://link.springer.com/chapter/10.1007/978-3-031-72983-6_18.

Zhu, H., Lin, C., Liu, G., Wang, D., Qin, S., Li, A., et al. (2024a). Intelligent agriculture: Deep learning in uav-based remote sensing imagery for crop diseases and pests detection. *Front. Plant Sci.* 15, 1435016. doi: 10.3389/fpls.2024.1435016

Zhu, H., Lin, C., Liu, G., Wang, D., Qin, S., Li, A., et al. (2024b). Intelligent agriculture: Deep learning in uav-based remote sensing imagery for crop diseases and pests detection. *Front. Plant Sci.* 15, 1435016. doi: 10.3389/fpls.2024.1435016

Zou, Y., Jeong, J., Pemula, L., Zhang, D., and Dabeer, O. (2022). "Spot-the-difference self-supervised pre-training for anomaly detection and segmentation," in *European Conference on Computer Vision*. Available online at: https://link.springer.com/chapter/10.1007/978-3-031-20056-4_23.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 June 2025
DOI 10.3389/fpls.2025.1577175

Check for updates

OPEN ACCESS

EDITED BY
Jun Ni,
Nanjing Agricultural University, China

REVIEWED BY
Wei Lu,
Nanjing Agricultural University, China
Elio Romano,
Centro di ricerca per l'Ingegneria e le Trasformazioni agroalimentari (CREA-IT), Italy
Carlos Francisco Simões Gomes,
Fluminense Federal University, Brazil
Sulaymon Eshkabilov,
North Dakota State University, United States

*CORRESPONDENCE
Qi Song
songqi@ujs.edu.cn

RECEIVED 15 February 2025
ACCEPTED 06 May 2025
PUBLISHED 03 June 2025

CITATION
Zhang S, Zang C, Yang Z, Tang L, Wang K, Wang A, Chen W, Song Q and Wei X (2025) Research on fault prediction and speed control system for unmanned combine harvesters based on IPSO-SVM and fuzzy logic.
Front. Plant Sci. 16:1577175.
doi: 10.3389/fpls.2025.1577175

COPYRIGHT
© 2025 Zhang, Zang, Yang, Tang, Wang, Wang, Chen, Song and Wei. 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.

Research on fault prediction and speed control system for unmanned combine harvesters based on IPSO-SVM and fuzzy logic

Shaocen Zhang[1], Chongquan Zang[1], Zhang Yang[1], Lingyu Tang[1], Kun Wang[1], Anzhe Wang[1], Wenming Chen[1], Qi Song[1,2]* and Xinhua Wei[1,2]

[1]School of Agricultural Engineering, Jiangsu University, Zhenjiang, China, [2]Key Laboratory for Theory and Technology of Intelligent Agricultural Machinery and Equipment, Jiangsu University, Zhenjiang, China

This study proposes an IPSO-SVM-based fault prediction and fuzzy speed control system for unmanned combine harvesters. The primary goal is to prevent clogging failures and ensure long-term stable operation of unmanned harvesting machines, maintaining efficiency while minimizing downtime. The system integrates multi-component slip rate data, collected from critical parts of the harvester, and uses the IPSO-SVM model for fault warning. The fuzzy control algorithm adjusts the operating speed based on the predicted fault status and feeding rate to mitigate clogging risks. Experimental results show that the system can accurately identify over 98.5% of fault states and reduce the occurrence of complete blockage by adjusting the harvester's speed within 0.5 to 2 seconds after minor clogging. This work demonstrates the feasibility of applying the system in field environments, providing a reliable solution for the intelligent and unmanned operation of combine harvesters in fields.

KEYWORDS

unmanned harvester, fault prediction, IPSO-SVM, fuzzy control, clogging prevention, slip rate monitoring, intelligent farming, machine learning

1 Introduction

The rice-wheat rotation has a tight planting schedule and limited labor availability, leading to widespread poor tillage and planting, as well as poor working conditions under harsh weather conditions (An et al., 2018; Marinoudi et al., 2019). Among the various steps in the "tillage, sowing, management, and harvesting" process, the quality of harvesting operations is particularly critical, involving relatively complex tasks and rigid labor requirements Lowenberg-DeBoer et al., 2020). The development of intelligent and unmanned combine harvesters not only addresses labor shortages but also enhances

land utilization and improves agricultural machinery efficiency (Boursianis et al., 2022; Elijah et al., 2018). Currently, commercial agricultural machinery auxiliary driving systems are relatively mature, assisting operators in controlling the steering wheel and ensuring path tracking for field operations (Li et al., 2023, 2022; Lu et al., 2020). However, the operator still needs to handle other fieldwork mechanisms. For example, during harvesting operations, the operator must adjust the vehicle's speed to avoid clogging based on the load, modify the header height according to terrain and crop lodging, and adjust the reel height according to the crop's ear position (Yanxin et al., 2022). Currently, these auxiliary systems save labor but do not reduce labor costs effectively (Tey and Brindal, 2022), making the development of an operational control system a key step toward achieving unmanned agricultural machinery (Chen et al., 2023; Jin et al., 2021). This study focuses on the operational load speed control system of rice-wheat combined harvesters.

Scholars at home and abroad have combined information and intelligent technologies to study the state monitoring and fault diagnosis of combine harvesters, achieving significant results (Qiu et al., 2022; Liang et al., 2016; Xu et al., 2019). Wang et al. (2023) pointed out that the threshing system, being one of the key components of a harvester, is easily influenced by crop characteristics, feeding rates, and the harvester's forward speed. When the feeding rate exceeds the threshing drum's power match, the drum speed decreases, and the drive belt may slip, potentially causing a clogging failure that damages related components. Craessaerts et al. (2010a) and Yang et al. (2024) categorized common harvester faults into header faults, threshing system faults, and re-threshing faults, analyzing the causes of these faults and proposing preventive measures to reduce their occurrence. Pavlyuk et al. (2022) analyzed the main system performance parameters of combine harvesters, identifying failure-prone areas such as the harvesting section, mechanical drive section, and threshing section, and determined the fault distribution during harvesting operations through experiments.

Several studies have attempted to control the harvester's load via feedback from key component working conditions, stabilizing the harvesting state to avoid failure (Chen et al., 2017; Li et al., 2022; Ma et al., 2023; Yu et al., 2024). Kruse et al. (1983) used changes in engine load to represent the harvester's load, monitoring the engine's torque as feedback to adjust the travel speed of the full-feeding combine harvester, keeping the feeding rate stable. McGechan and Glasbey (1982) used threshing drum torque, main auger torque, and loss rate as reference variables to adjust speed, maintaining a stable operating state for the harvester. Li et al. (2021) developed an online monitoring system for the harvester's hydraulic components, such as the header, conveyor, and threshing drum, based on LabVIEW, enabling real-time collection of key working parameters and fault warnings. Abdeen et al. (2022) designed a stress monitoring system for the threshing drum's top cover using resistive sensors, achieving real-time monitoring of feeding rates and early warnings of threshing drum clogging. Qin et al. (2011) designed a harvesting speed control system based on RBF, using header auger speed to assist in feeding state measurement and maintaining a constant threshing drum speed.

With the widespread application of machine learning algorithms in fault diagnosis, more solutions are available for diagnosing faults (Hao et al., 2022; Chen et al., 2020). Craessaerts et al. (2010b) suggested that using historical data from combine harvesters and artificial neural networks to construct fault diagnosis models can achieve approximately 80% fault recognition accuracy. By integrating fuzzy control of the cleaning process and combining the data model with the operator's experience, faults can be effectively avoided. Chen et al. (2014) designed a fault diagnosis system for combine harvesters based on a fuzzy neural network (FNN) algorithm, establishing a nonlinear mapping relationship between fault symptoms and fault types, processing input component speed values, and outputting fault diagnosis results. Jack and Nandi (2001) and Samanta et al. (2003) successfully applied SVM for bearing fault detection and experimentally verified its high engineering application value in fault diagnosis. Diez-Olivan et al. (2018) proposed a new algorithm based on SVM to assess sensor health and demonstrated its effectiveness in analyzing the time-dependent trends of ship diesel engine faults.

Support Vector Machine (SVM) theory has a solid mathematical foundation and rigorous derivations. It can transform classification and regression problems into optimization problems with constraints, which are solved through various mathematical methods, optimizing its algorithm and improving computational speed and performance. SVM theory has achieved excellent results in condition monitoring and fault warning applications (Liu et al., 2013; Chen et al., 2019; Ding et al., 2022; Ahmad et al., 2021; Peng et al., 2021). However, existing fault diagnosis methods, particularly those based on traditional machine learning (such as basic SVM), often face challenges such as poor generalization in small-sample, high-dimensional data scenarios and difficulty in parameter tuning. Deep learning approaches (e.g., CNN, LSTM, and reinforcement learning) have shown promising results but typically require extensive datasets and higher computational resources, making them less practical for real-time onboard applications.

To address these limitations, this study aims to predict clogging fault risks and dynamically adjust operational speed, ensuring the stable, long-term operation of unmanned combine harvesters. Instead of pinpointing specific fault locations, the primary focus is on managing the overall clogging risk to improve harvesting efficiency and reliability. Specifically, we propose a harvesting speed control system that integrates multi-component slip rate monitoring, an improved particle swarm optimization (IPSO)-optimized SVM model for accurate fault state classification, and a fuzzy logic-based algorithm for adaptive speed control. By collecting real-time data from Hall-effect sensors installed on critical harvesting components and utilizing 4G communication with cloud platforms, the system quickly and effectively reduces clogging risks. This approach significantly enhances operational efficiency, reduces downtime, and holds great potential for practical implementation in intelligent and unmanned agricultural

machinery, extending its applicability to diverse farming environments and various agricultural machinery models.

The main innovations and contributions of this research are as follows:

1. Multi-component signal fusion: By collecting rotational speed signals from key components such as the header auger, conveyor, threshing drum, auger, and vibration screen, the system integrates slip rate data to reflect potential clogging signs across different transmission chains, overcoming the limitations of traditional single-component load feedback.
2. Efficient fault warning: Utilizing IPSO-SVM to classify small sample, high-dimensional clogging fault patterns enables accurate fault warning. The optimized algorithm enhances the SVM kernel function parameters and penalty factors, improving warning accuracy and robustness. Field tests validate that the fault warning model can correctly identify more than 98% of fault states.
3. Automatic speed regulation: Using fuzzy control, the system integrates fault warning results and feeding rate estimates into speed adjustment rules to enable real-time closed-loop control of operational load. This reduces clogging risks while ensuring fieldwork efficiency. Field tests confirm that the system can reduce speed within 0.5–2 seconds after detecting minor clogging, effectively preventing the occurrence of complete blockage failures.

To better present the research approach and experimental results, the structure of this paper is organized as follows: Section 2: Introduces the harvester parameters, power transmission structure, fault warning speed control system's hardware and software framework, the IPSO-SVM-based fault warning principle, and the fuzzy-based speed control principle. Section 3: Describes the experimental design and results of feeding rate calibration, fault warning model performance verification, and speed control experiments. Section 4: Summarizes the research findings and conclusions, and discusses potential directions for future research.

2 Materials and methods

2.1 Harvester parameters and common faults

This study is based on the World Ruilong 4LZ-5.0 combine harvester, with key parameters listed in Table 1. The selection of this specific harvester was based on its affordability, ease of availability for field testing, and its widespread use as a commercial model featuring standard components common in modern combine harvesters. Moreover, all key mechanical parts are easily accessible for sensor installation, facilitating the implementation of electronic monitoring systems. The harvester's hydraulic speed transmission (HST) system also simplifies the integration of automated speed control mechanisms. Importantly, the proposed fault prediction and speed control system is not limited exclusively to this particular model; it can be readily adapted to other combine harvesters sharing similar mechanical and electronic architectures.

TABLE 1 The main parameters of the World Ruilong 4LZ-5.0.

Parameter	Value	Unit
Combine dimensions (length- width- height)	4960 - 3950 - 2830	mm
Combine weight	3000	kg
Engine power	75	kw
Gear shift method	3 speeds HST	–
Travel speeds (at rated engine speeds)	Low speed: 0 to 1	m/sec
	Standard: 0 to 1.5	
	High speed: 0 to 2	
Cut width	2200	mm
Feeding capacity	5	kg/s
Operational efficiency	0.68 – 0.85	ha/h
Grain tank capacity	1.5	m^3

The harvester's power consumption is primarily divided into two parts: the drive system and the operational system. The load of the drive system is mainly influenced by factors such as ground friction resistance, slope resistance, the internal friction of the transmission system, and the inertia forces during vehicle acceleration and deceleration. The operational system's load is primarily affected by factors such as the crop feed rate per unit time, the composition of the crop, and the working conditions of the respective operational components. The real-time power consumption of the drive system can cause fluctuations in the maximum load of the operational system.

The crop transfer process and the power transmission process of the harvester are not consistent, which makes it challenging to infer crop clogging from monitoring the power system's status. The crop transfer flow includes: cutter bar cutting the stems → feeding the crop into the header auger → conveyor chain rake → threshing cylinder → vibrating sieve and fan cleaning → secondary cleaning of residuals → grain is sent into the grain tank by the auger, while straw is sent to the straw cutter for grinding and discharge after passing through the threshing cylinder. The power transmission flow in the operational system includes: engine → fan →header conveyor→ header auger → cutter bar and reel, fan → threshing cylinder, fan → straw cutter, fan → grain auger → vibrating sieve, as shown in Figure 1.

According to recent statistics on warranty and fault repair data (Chen et al., 2023), most mechanical clogs occur when key components in the transmission chain (such as the Header Auger, Conveyor, Threshing Cylinder, Grain Auger, Vibration Sieve, Straw Cutter, etc.) become obstructed or fail to operate smoothly due to crop accumulation or excessive load, as shown in Figure 2. The Blower Fan often serves as the driving axis for the transmission system, with most components maintaining a fixed speed ratio with the Blower Fan (such as the Header Auger and Vibration Sieve, which maintain a fixed speed ratio with their respective upstream components). When any component experiences clogging or

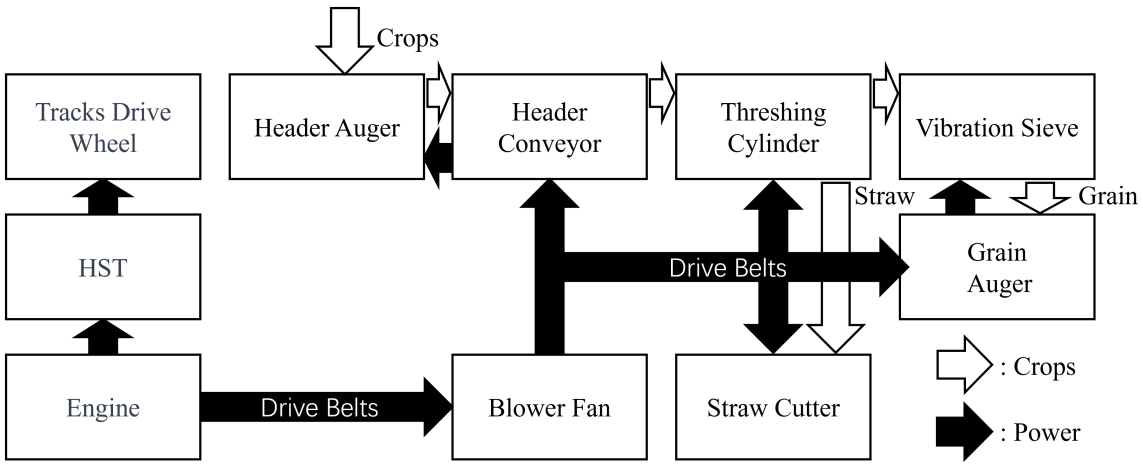


FIGURE 1
The power system structure of the 4LZ-5.0 combine harvester.

slipping, its slip rate deviates significantly. Therefore, in this study, the slip rate of the pulley is used as the primary monitoring parameter, supplemented by monitoring the harvester's forward speed and GPS positioning information to assess the feeding rate. An intelligent algorithm is then applied to comprehensively determine the fault risk value of the harvesting operation.

2.2 Overall design of the system

To achieve fault warning and speed control during the operation of the combine harvester, the onboard hardware system used in this study is configured as shown in Figure 3, which includes:

2.2.1 Operational component speed monitoring system

HR-M1850 Hall effect speed sensors (Omron (China) Co., Ltd.) are used to collect real-time data on the rotational speed of the drive pulleys for seven key operational components: Header Auger, Header Conveyor, Threshing Cylinder, Blower Fan, Grain Auger, Vibration Sieve, and Straw Cutter.

2.2.2 Navigation system

The system, based on the UB482 positioning core board (Beijing Unicore Communications Co., Ltd.), is utilized for the harvester's unmanned driving functionality and simultaneously collects vehicle speed information (Sun et al., 2022).

FIGURE 2
Common faults in combine harvesters. **(a)** Header auger clogging; **(b)** Header conveyor clogging; **(c)** Threshing cylinder clogging; **(d)** Vibrating sieve clogging; **(e)** Grain auger clogging; (f) Straw cutter clogging.

FIGURE 3
Hardware structure of fault warning and speed control system. **(a)** Combine harvester; **(b)** Header auger; **(c)** Header conveyor; **(d)** Threshing cylinder; **(e)** Blower fan; **(f)** Vibration sieve; **(g)** Grain auger; **(h)** Straw cutter; **(i)** Positioning antenna; **(j)** RTK radio; **(k)** GNSS receiver; **(l)** Navigation software; **(m)** Embedded Controller; **(n)** LTE-QTU Module; **(o)** HST Rod; **(p)** Accelerator; **(q)** Header lift rod.

2.2.3 Control system

The system, based on the EMB8600I embedded industrial control board (Beijing EmbedArm Co., Ltd.), processes sensor data, transmits it to the remote communication module, receives fault warning results from the cloud, and controls the vehicle's actions.

2.2.4 Remote communication system

The system, based on the G780 4G LTE DTU module (Jinan Youren Network Co., Ltd.), enables remote communication with the cloud server.

2.2.5 Onboard actuator system

This includes the electrification of the HST speed change actuator and the Header lift actuator using electric push rods (YNT03, Nanjing Yongnuo Transmission Equipment Co., Ltd.), as well as the conversion of the original manual throttle to electric control via the embedded control board's ADC interface.

The cloud server used in this study consists of a Socket server, database module, and web interface, as shown in Figure 4. The Socket server primarily handles communication with the onboard terminal, providing services such as data reception, information

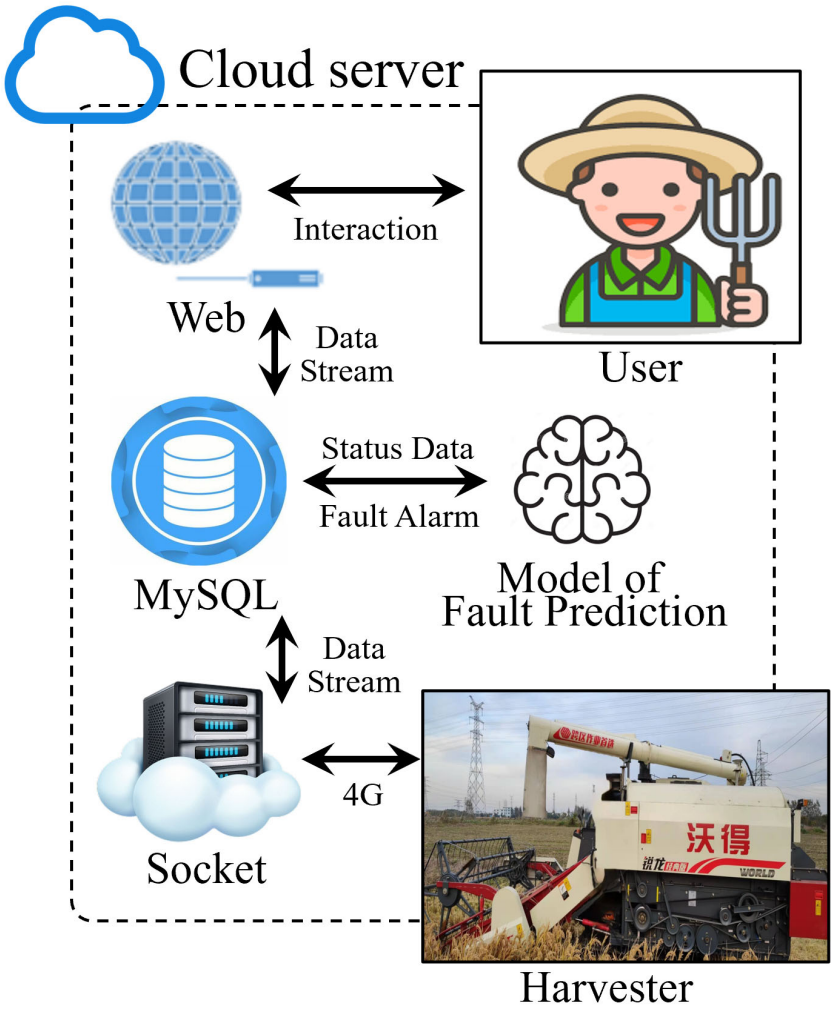


FIGURE 4
Cloud server architecture diagram.

feedback, data protocol parsing, and fault warnings. The MySQL database module is responsible for the persistent storage of data received from the Socket server, fault codes generated by fault warnings, and other relevant information from the web service. The web interface handles remote monitoring of the harvester and human-machine interaction features (Zang et al., 2022).

The server is hosted on Alibaba Cloud ECS (Elastic Compute Service) and runs Windows Server 2008, ensuring stability and security. Communication between the system components is based on a TCP transparent transmission Socket protocol, with custom message formats for uploading sensor data and sending back fault warning states.

The development of the machine learning-based fault warning model for the combine harvester is primarily carried out using Python 3.6. The Scikit-learn library is used to build an IPSO-SVM classification model, which is trained using historical clogging fault data from the combine harvester. The trained model is serialized and saved using the pickle module. The Socket server loads the model file into memory and uses real-time rotational speed data to generate fault warning results.

2.3 IPSO-SVM-based fault risk warning model

The clogging fault in combine harvesters can be treated as a pattern recognition problem, with mainstream machine learning methods such as neural networks and Support Vector Machines (SVM) commonly applied. However, traditional SVM methods are sensitive to hyperparameter tuning, which significantly affects their generalization performance, especially under conditions with limited training samples. Similarly, particle swarm optimization (PSO)-based approaches can suffer from premature convergence, limiting their ability to find optimal solutions efficiently.

To overcome these limitations, this study adopts an improved particle swarm optimization (IPSO) strategy to enhance the generalization capability and accuracy of the SVM model. Specifically, the IPSO algorithm effectively optimizes the kernel parameters and penalty factors (C and σ) of the SVM, enabling better fault state classification performance. By introducing nonlinear transformations, the IPSO-SVM framework maps low-

dimensional, non-separable data into high-dimensional space, enhancing fault prediction accuracy and robustness.

SVM is a supervised learning method based on the maximum margin strategy. It classifies data in high-dimensional space by constructing an optimal hyperplane. As shown in Figure 5, in a two-dimensional space, suppose there are two types of samples, and linear classification can be achieved through multiple boundary functions (M). The optimal boundary is defined by support vectors. Support vectors (M1, M2) are the samples closest to the boundary, and the distance between them is the margin. Maximizing the margin helps enhance the model's generalization ability, thus improving classification accuracy.

When this theory is extended to high-dimensional data, the linear decision boundary becomes a hyperplane. The goal is to determine the optimal hyperplane that accurately separates samples into their respective categories. To achieve this, the radial basis function (RBF) kernel is employed to transform nonlinear, non-separable input data into a higher-dimensional feature space, enabling linear separability. This transformation allows the SVM model to effectively classify fault states. Fault state classification follows a similar principle to the two-dimensional case illustrated in Figure 5, where the optimal hyperplane is defined by support vectors, clearly distinguishing normal and faulty states. The decision boundary generated by these support vectors divides the feature space into distinct classification regions. Samples that fall within each region are classified accordingly, based on their proximity to historical support vectors representing different fault states (e.g., normal, mild clogging, severe clogging, and complete blockage). This visual representation provides an intuitive understanding of the model's classification performance and highlights how the SVM differentiates between normal and faulty conditions using the optimized hyperplane.

Let the dataset be $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, where $x_i \in R^n$ and $y_i \in \{1, -1\}$, x_i represents the input sample vector, and y_i represents the sample label. The hyperplane can be represented by the following mathematical model (Equation 1):

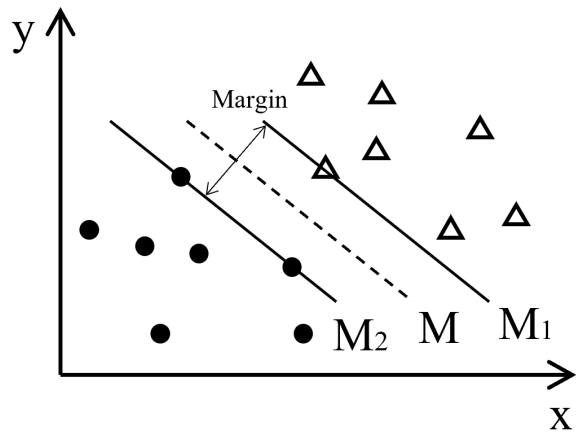


FIGURE 5
Schematic diagram of 2D linear separation.

$$\omega^T x + b = 0 \tag{1}$$

where $\omega = (\omega_1; \omega_2; \cdots; \omega_n)$ is the normal vector of the hyperplane, and b is the offset, which determines the distance from the hyperplane to the origin. The distance d from a sample point to the hyperplane can be expressed as (Equation 2):

$$d = \frac{|\omega^T x + b|}{\| \omega \|} \tag{2}$$

When the hyperplane correctly classifies the samples, it should satisfy the following constraints (Equation 3):

$$\begin{cases} \omega^T x_i + b - 1 \geq 0, y_i = +1 \\ \omega^T x_i + b + 1 \leq 0, y_i = -1 \end{cases} \tag{3}$$

Support vectors are the samples closest to the hyperplane, and the distance between the support vectors of the two classes is called the margin. This margin D can be expressed as (Equation 4):

$$D = \frac{2}{\| \omega \|} \tag{4}$$

The optimal hyperplane maximizes the margin, which is equivalent to minimizing (Equation 5):

$$\min_{\omega,b} \frac{\| \omega \|^2}{2} \quad s.t. \quad (\omega^T x_i + b) y_i \geq 1 \tag{5}$$

By introducing Lagrange multipliers α_i (Rockafellar, 1993), the optimization problem can be transformed into the following form (Equation 6):

$$L(\omega, b, \alpha) = \frac{\| \omega \|^2}{2} - \sum_{i=1}^{n} \alpha_i ((\omega^T x_i + b) y_i - 1) \tag{6}$$

Taking partial derivatives of this Lagrange function and setting them to zero, we obtain (Equations 7, 8):

$$\omega = \sum_{i=1}^{n} \alpha_i x_i y_i \tag{7}$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \tag{8}$$

Substituting the above into the Lagrange function transforms the problem into its dual form (Shawe-Taylor and Sun, 2011) (Equation 9):

$$\max L(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (\mathrm{x_i}, \mathrm{x_j}) \quad s.t. \begin{cases} \sum_{i=1}^{n} \alpha_i y_i = 0 \\ \alpha_i \geq 0 \end{cases} \tag{9}$$

This is solved using the Sequential Minimal Optimization (SMO) algorithm (Platt, 1998).

For non-linear separable problems, such as the relationship between rotational speed and slip rate data for fault risk in combine harvester failures, we introduce a nonlinear transformation. This maps low-dimensional, non-separable data into a high-dimensional space, making it separable. Using a kernel function, we can compute the inner product in the original space, avoiding direct computation

in the high-dimensional space. Common kernel functions include the Radial Basis Function (RBF), polynomial kernel, and Sigmoid kernel (Hofmann et al., 2008). The RBF kernel is expressed as (Equation 10):

$$k(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{\sigma^2}} \tag{10}$$

where σ is a parameter that significantly affects the SVM model's performance, and it must be tuned to optimize the model.

To address the impact of measurement errors and noise on hyperplane optimization, we employ a soft margin method (Chen et al., 2004), optimizing by introducing slack variables ξ and a penalty factor C. The penalty factor determines the model's tolerance for misclassification and affects the generalization and fitting capacity of the model.

The optimization problem for the optimal hyperplane can now be expressed as (Equation 11):

$$\min_{\omega,b,C} \frac{\|\omega\|^2}{2} + C\sum_{i=1}^{n}\xi_i \quad s.t. \ (\omega^T\varphi(x_i) + b)y_i \geq 1 - \xi_i \tag{11}$$

The dual form of the optimization problem is (Equation 12):

$$\max L(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j k(x_i, x_j) \quad \text{s.t.} \ \begin{matrix}\sum_{i=1}^{n}\alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C\end{matrix} \tag{12}$$

Finally, by solving the dual problem of the Lagrange function, we obtain the decision function for the nonlinear SVM model (Equation 13):

$$f(x) = \text{sign}(\sum_{i=1}^{n}k(x_i, x)\alpha_i y_i + b) \tag{13}$$

To improve the classification performance of the SVM model, we adopted an improved particle swarm optimization (IPSO) algorithm to optimize SVM parameters C and σ. Traditional SVM models are sensitive to hyperparameter tuning, significantly affecting their generalization performance, especially with limited data. Conventional PSO often suffers from premature convergence, limiting the quality of parameter optimization. Therefore, the IPSO approach, featuring enhanced inertia weight adaptation, was used to accelerate convergence and improve optimization efficiency.

The position and velocity update rules for a particle in PSO are defined as follows (Equations 14, 15) (Cheng and Jin, 2015):

$$V_i^{t+1} = \omega V_i^t + c_1 rand_{[0,1]}(pbest_i - X_i^t) + c_2 rand_{[0,1]}(gbest - X_i^t) \tag{14}$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \tag{15}$$

where V_i^t is the particle's velocity vector, ω is the inertia weight, which controls the balance between global and local search, c_1 and c_2 are the cognitive and social learning factors, and $rand_{[0,1]}$ is a random value between 0 and 1, adding randomness to the search process. $pbest_i$ is the best solution found by the particle, and $gbest$ is the best solution found by the entire swarm. X_i^t is the particle's position at time t, representing the current state in the solution space.

During PSO, dynamic adjustments to the inertia weight ω and the learning factors c_1 and c_2 help accelerate convergence and avoid local optima. By adjusting these parameters, PSO effectively optimizes the SVM model parameters, improving classification accuracy. To address issues like premature convergence in traditional PSO (Nakisa et al., 2014), we propose a dual-improvement strategy:

1. A linear differential decrement strategy to adjust the inertia weight (Equation 16):

$$\omega = (\omega_s - \omega_e) \times (\frac{k}{k_{max}})^2 + (\omega_e - \omega_s) \times (\frac{2 \times k}{k_{max}}) + \omega_s \tag{16}$$

where k is the current iteration, k_{max} is the maximum number of iterations, ω_s is the initial inertia weight, and ω_e is the final inertia weight.

2. An asynchronous adjustment strategy to improve the size of the learning factors (Equations 17, 18):

$$c_1 = (\frac{C_{1s} + C_{1e}}{2}) + (\frac{C_{1s} + C_{1e}}{2}) \times cos(\frac{k\pi}{k_{max}}) \tag{17}$$

$$c_2 = (\frac{C_{2e} + C_{2s}}{2}) + (\frac{C_{2e} - C_{2s}}{2}) \times cos(\frac{k\pi}{k_{max}}) \tag{18}$$

where C_{1s} and C_{2s} are the initial individual and social learning factors, and C_{1e} and C_{2e} are the final individual and social learning factors.

In this study, the number of particles was set to 50, based on previous optimization studies that demonstrated this size effectively balances convergence speed and optimization quality. The maximum iterations (100) were empirically determined to ensure sufficient exploration without excessive computational cost (Liu et al., 2013; Chen et al., 2019; Ding et al., 2022; Ahmad et al., 2021).

Compared with conventional SVM and PSO-SVM models, the IPSO-SVM approach provides better classification results due to its efficient parameter optimization and improved convergence properties. The proposed approach thus effectively addresses clogging fault prediction challenges, ensuring long-term stable operation of unmanned harvesters.

The process of establishing the harvester clogging fault diagnosis model can be summarized as follows:

1. Collect data under normal, mild clogging, severe clogging, and complete blockage conditions through the experimental platform, extract slip rate features, normalize the data, and split the samples into a training set and a test set with a 4:1 ratio.

2. Use the IPSO algorithm in MATLAB 2022a to efficiently search for optimal SVM parameters, specifically kernel parameters and penalty factors.
3. Construct the IPSO-optimized SVM model in Python using the Scikit-learn library, perform offline training, and generate the fault warning model.
4. Deploy the trained IPSO-SVM model via the Socket server, uploading rotational speed data from the onboard DTU device to obtain fault prediction states in real time.

2.4 Fuzzy-based operation speed control system

Under the premise of properly adjusting the mechanical structure, clogging failures often occur during field operations of combine harvesters due to excessive feeding rates, which may result from factors such as high harvesting speeds, low cutting heights, high crop density, and excessive moisture content. During the initial stage of clogging, the issue can be effectively mitigated by appropriately reducing the vehicle's speed. Skilled operators can often maximize harvesting speed while preventing clogging failures (Schwegman et al., 2021). Similarly, to ensure that unmanned combine harvesters can operate efficiently while avoiding clogging, they must be capable of automatically adjusting the harvesting speed based on variations in feeding rates and fault conditions. This process involves complex control challenges that are difficult to model precisely using traditional mathematical models.

Fuzzy control algorithms, which do not rely on exact mathematical models, are suitable for handling such imprecise control systems. By fuzzifying the fault states and feeding rates per unit time, and referencing manual operating experience, fuzzy control rules can be established. In this study, the fuzzy control rules were established based on extensive field trials and validated through operator feedback. By analyzing historical clogging events and operator interventions, optimal speed adjustments were determined for various clogging levels. This method ensures strong adaptability to real-world operational conditions. Subsequently, the harvester's speed can be calculated and defuzzified. This strategy enables dynamic speed adjustment based on changes in fault status and feeding rates, thereby enhancing the adaptability of unmanned combine harvesters, reducing clogging, and improving operational efficiency.

Fuzzification of Input and Output Variables:

1. Fault status: The fault status level is output from the preliminary fault prediction model and corresponds directly to fuzzy quantization levels {Normal (N), Mild Clogging (LC), Severe Clogging (HC), Complete Blockage (CB)}.
2. Feeding intake (kg/s): The domain range is set to [0, 9] based on harvester parameters, with quantization levels {Small (S), Medium (M), Large (L)}, and a quantization factor of 3.
3. Harvesting speed (m/s): The domain range is set to [0, 1.2] based on harvester parameters, with quantization levels {Stop (ST), Slow (S), Medium (M), Fast (F)} and a quantization factor of 0.3. All are represented by Gaussian membership functions.

Fuzzy Control Rule Design Principles:

1. Priority of fault state: In cases of severe clogging, the speed should be immediately reduced, or even stopped.
2. Feeding rate influence on speed: In the absence of faults, a higher feeding rate leads to a lower forward speed to prevent future clogging.
3. Adaptive adjustment: For mild clogging, speed can be slightly reduced to allow the system to recover; for severe clogging or complete blockage, speed should be significantly reduced or stopped. The fuzzy rules are shown in Table 2.

Defuzzification Using the Centroid Method:

The centroid method is used for defuzzification, providing smooth output and is widely applied in industrial settings (Hung and Wu, 2002; Wu et al., 2019; Qi et al., 2022; Song et al., 2022; Zhang et al., 2024; Longaray et al., 2019; Pinochet et al., 2023). The calculation formula is as follows (Equation 19):

$$V = \frac{\sum_{i=1}^{n} V_i f(V_i)}{\sum_{i=1}^{n} f(V_i)} \tag{19}$$

where V is the controller output, n is the number of rules, $f(V_i)$ is the membership function, and V_i is the corresponding speed value.

3 Experiments and discussion

3.1 Feed rate calibration experiment

The experimental field area is 6500 m², and the rice variety used is Nanjing 9108, with a crop-straw ratio of 1.65 and an overall moisture content of 40.4%. The air temperature ranged

TABLE 2 Fuzzy control rule.

Fault status	Feeding Intake	Speed
N	S	F
N	M	M
N	L	S
LC	S	M
LC	M	S
LC	L	ST
HC	S	S
HC	M	ST
HC	L	ST
CB	Any	ST

from 17-19°C. The experimental data were collected during the autumn harvest season in 2023. The harvesting operation was conducted at specific cutting widths and cutting heights to ensure consistency and comparability of the data. During the harvesting process at various locations in the experimental field, it was assumed that the crop's growth density and moisture content were uniform. Therefore, the actual feeding rate could be calibrated by controlling the cutting width and cutting height. The formula for calculating the feeding rate is (Equation 20):

$$Q = mS = mdv \tag{20}$$

where ? is the actual feeding rate of the harvester (kg/s), m is the crop mass per unit area (kg/m^2), S is the crop area harvested per unit time (m^2/s), d is the harvester's cutting width (m), ? is the harvester's operating speed (m/s).

A crop block of approximately 4 square meters was selected to calibrate the crop mass per unit area, with cutting heights maintained at 0.15 m, 0.25 m, and 0.35 m. Manual harvesting was performed, and the harvested crop was weighed. The field conditions are shown in Figure 6a, the stubble conditions in Figure 6b, and the harvested area in Figure 6c. The weighing results for each crop block are presented in Table 3.

A quadratic function was used to fit the experimental data, yielding the cutting height-feeding rate relationship for the current field was calculated as Equation 21:

$$Q = (-25h^2 + 4.9h + 5.39)dv \tag{21}$$

where h is the cutting height (m).

It can be inferred from the above equation that, with constant cutting height and cutting width, the harvester's speed is linearly related to the feeding rate.

By fixing the crop variety, crop-straw ratio, and overall moisture content, the feeding rate calibration experiment in the field was completed. This provides data on the feeding rate conditions for harvesting fields under similar conditions, supporting the subsequent development of speed control strategies to prevent clogging. However, the actual harvesting feeding rate is influenced by multiple factors, such as crop variety, crop-straw ratio, overall moisture content, growth density, lodging condition, and the harvester's technical state (maintenance condition, wear degree, power configuration). Real-time calculation of the harvesting feeding rate is beyond the scope of this study. Based on the weighing results, it was observed that the growth density of the rice in the experimental field was relatively uniform, with minimal variation in crop mass per unit area. Therefore, this study only establishes the cutting height-feeding rate relationship through the equation above.

3.2 Performance verification of fault prediction model

To validate the practical application of the IPSO-SVM-based clogging fault warning model in field environments, this study involved speed state data collection from the pulleys of seven key components: Blower Fan, Header Auger, Header Conveyor, Threshing Cylinder, Vibration Sieve, Grain Auger, and Straw Cutter. The study included offline model training and testing, as well as online fault warning experiments. The model inputs consisted of the slip rate features from seven channels, while the output was the fault warning state labels, with a corresponding relationship between state labels and fault types as shown in Table 4. The normalization formula for the pulley slip rate features is (Equation 22):

$$x' = 1 - \frac{x - x_{min}}{x_{max} - x_{min}} \tag{22}$$

The experimental field was planted with the rice variety Nanjing 9108, with an average plant height of approximately 0.95 meters and a moisture content of about 40.3%. The engine's rated speed was set to 2500 r/min, and the cutting width was 2 meters. Based on typical clogging fault scenarios in real-world conditions, human interventions such as lowering the cutting height, increasing harvesting speed, and increasing crop density were used to induce faults. The field experiment setup is shown in Figure 7. The harvesting machine's rotational speed data under various fault states is presented in Figure 8. After excluding outliers and missing values, a total of 1071 valid data points were obtained. The fault warning classification labels were manually annotated, with 267 instances of normal state, 267 instances of mild clogging,

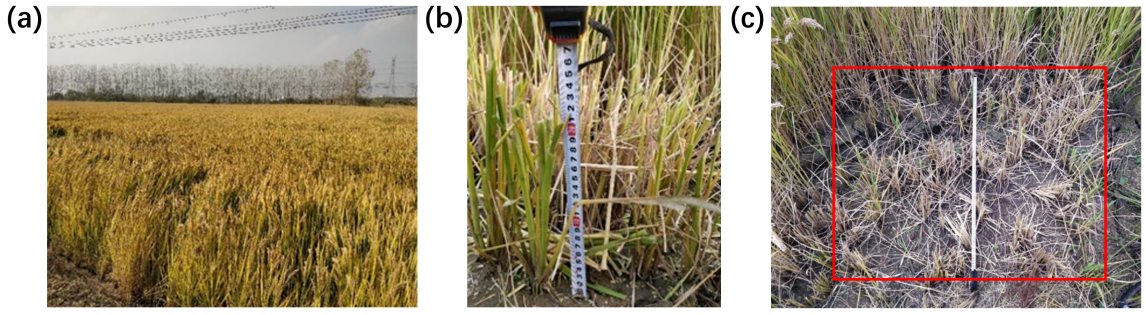


FIGURE 6
Feed rate calibration experiment. **(a)** General view of the experiment field; **(b)** Cutting height measurement; **(c)** Reaping area survey.

TABLE 3 Weighing results of crop blocks.

Crop number	Cutting height (m)	First weighing (kg)	Second weighing (kg)	Third weighing (kg)	Mass-area ratio (kg/m^2)	Average mass-area ratio (kg/m^2)
1	0.15	21.01	21.00	21.02	5.56	5.56
2		22.16	22.12	22.21	5.71	
3		21.22	21.20	21.22	5.60	
4		21.40	21.41	21.40	5.61	
5		21.07	21.10	21.09	5.57	
6		19.84	19.89	19.88	5.49	
7		21.43	21.42	21.41	5.62	
8	0.25	19.84	19.82	19.84	4.96	5.05
9		21.02	20.99	21.06	5.26	
10		20.28	20.25	20.27	5.07	
11		20.56	20.57	20.56	5.14	
12		20.13	20.16	20.15	5.04	
13		18.94	18.97	18.97	4.74	
14		20.68	20.68	20.66	5.17	
15	0.35	16.91	16.88	16.89	4.04	4.04
16		18.07	18.02	18.11	4.06	
17		17.28	17.25	17.27	4.05	
18		17.66	17.68	17.67	4.06	
19		17.02	17.04	17.03	4.04	
20		15.94	15.97	15.96	4.00	
21		17.22	17.21	17.20	4.05	

267 instances of severe clogging, and 270 instances of complete blockage. The training and testing datasets were randomly divided in a 4:1 ratio from each category.

The IPSO-SVM model was built using MATLAB 2022a, where the IPSO population size was set to 50, the maximum number of iterations k_{max} was set to 100, the initial inertia weight ω_s was set to 2, the final inertia weight ω_e was set to 0.4, and the learning factors were set as C_{1s}=1.5, C_{1e}=0.5, C_{2s}=0.7 and C_{2e}=2.0. The SVM kernel function used was the radial basis function (RBF). The feature data for both the training and testing datasets were normalized and input into the model for parameter optimization. The resulting fitness curve is shown in Figure 9a, where the optimal SVM parameters C and σ were found to be 3.6190 and 13.9032, respectively. The best fitness value of 96.50% was achieved at the 32nd generation. The testing dataset was then input into the trained model, and the confusion matrix is shown in Figure 9b. The overall classification accuracy of the model was 98.59%. The classification errors occurred because, during the initial stage of clogging, the difference in severity was not clearly reflected in the rotational speed data. When severe clogging occurred, the rotational speed data rapidly decreased. However, the sampling speed of the measurement system was limited and could not clearly distinguish between severe clogging and complete blockage.

TABLE 4 Harvester status categories and classification labels.

Status category	Classification label
Normal	1
Lightly clogging	2
Heavily clogging	3
Completely blocked	4

Despite achieving high classification accuracy, the proposed system is subject to certain experimental errors and uncertainties. Three primary sources of error were identified: firstly, sensor inaccuracies in slip rate measurements, as sensor sensitivity and installation conditions might cause minor deviations in collected data; secondly, variability in environmental conditions, such as fluctuations in crop density, moisture content, and soil properties, which may impact model performance; and thirdly, potential delays in data transmission via the 4G communication module, resulting in slight discrepancies in fault detection timing and subsequent speed adjustments. To account for these uncertainties, confidence

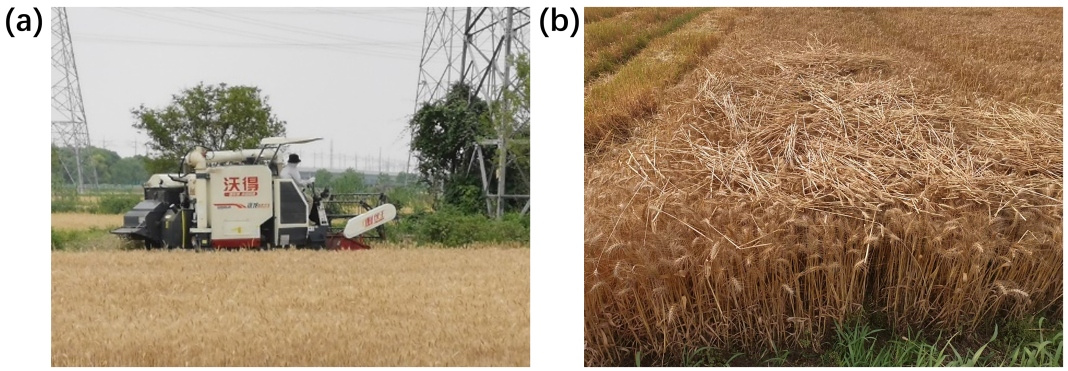


FIGURE 7
Fault Setting experiment. **(a)** Lower cutting table and increase speed; **(b)** Increase crop.

intervals have been incorporated into Table 5, providing a statistical measure of reliability when comparing the performance of different classification models. Confidence intervals for classification accuracy were computed based on the binomial normal approximation at a 95% confidence level (Brown et al., 2002), using the formula below (Equation 23):

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/N} \tag{23}$$

where $\hat{p}$ is the sample accuracy, $z_{\alpha/2}$ is the z-value at the selected confidence level (for example, at 95% confidence, $z_{0.025} \approx 1.96$), and N is the test set size.

To evaluate the performance of the IPSO-SVM fault diagnosis model, a comparison was made with the PSO-SVM and traditional SVM models using the same training and testing datasets. The results, along with their corresponding 95% confidence intervals calculated using the normal approximation, are summarized in Table 5. The total accuracy of the IPSO-SVM model was 8.44% and

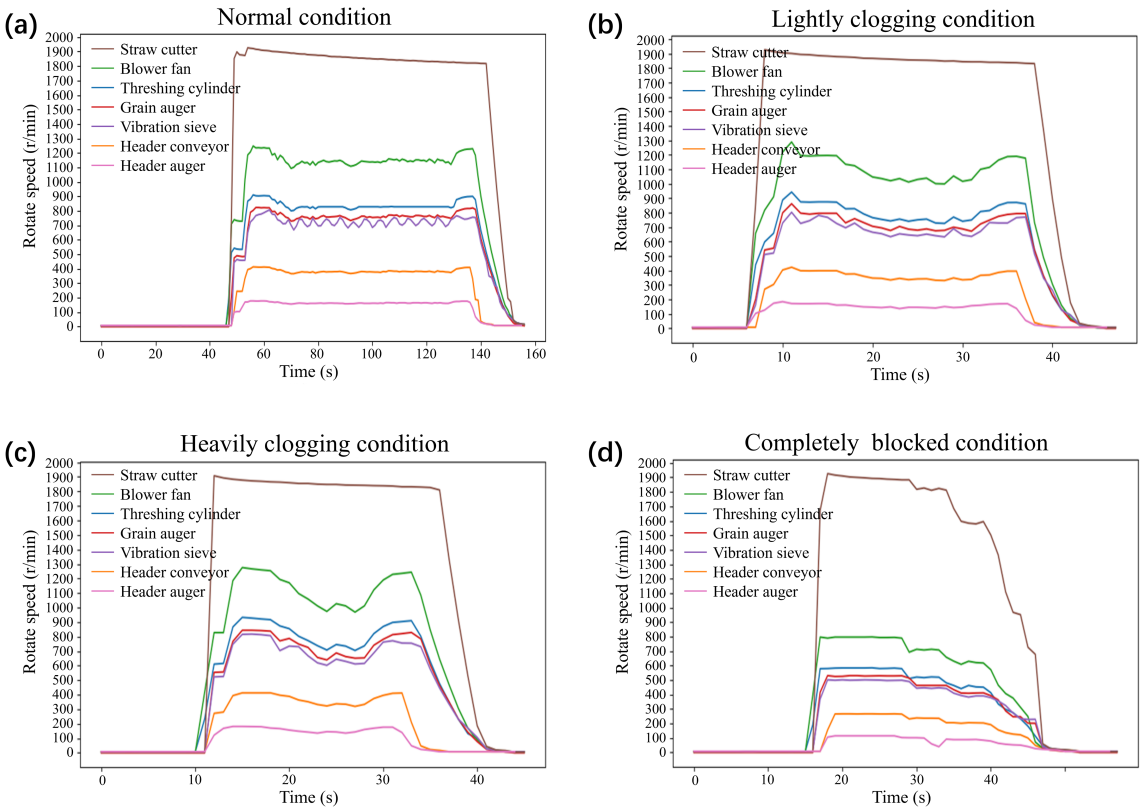


FIGURE 8
Speed data of each part. **(a)** Rotate speed in normal condition; **(b)** Rotate speed in lightly clogging condition; **(c)** Rotate speed in heavily clogging condition; **(d)** Rotate speed in completely blocked condition.

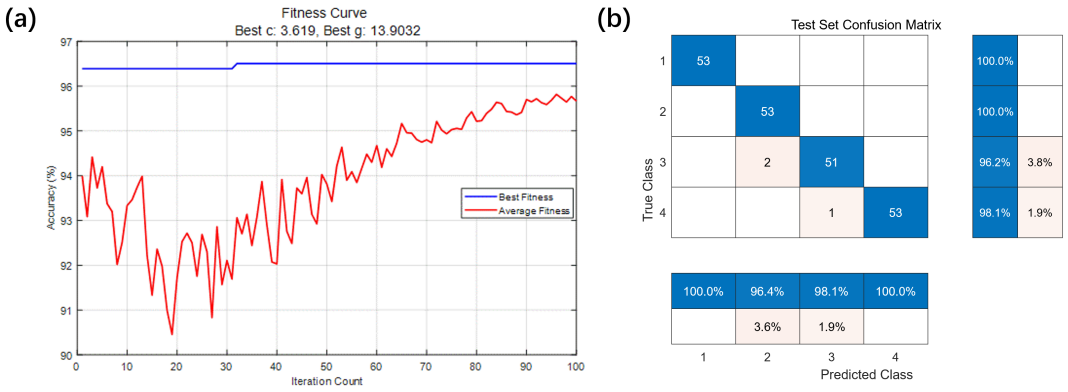


FIGURE 9
Fault Setting experiment. **(a)** Fitness change curve; **(b)** IPSO-SVM test set confusion matrix.

21.13% higher than that of the PSO-SVM and traditional SVM models, respectively. The inclusion of confidence intervals confirms the statistical reliability of these improvements. These results clearly demonstrate that the IPSO algorithm is more effective in optimizing the SVM parameters compared to the standard PSO algorithm, thus enhancing the generalization capability and classification accuracy of the SVM model.

Additionally, although the IPSO-SVM model demonstrated high accuracy, misclassification errors such as false positives and false negatives can still occur. False positives—incorrectly identifying normal states as fault conditions—might trigger unnecessary speed adjustments, potentially reducing operational efficiency. However, preventing fault escalation remains the primary goal; thus, the trade-off of occasionally reduced efficiency is considered acceptable. Conversely, false negatives—failing to detect actual faults—may lead to delayed corrective actions, but since faults typically worsen over time, subsequent detections are more likely to trigger accurate fault identification. Therefore, neither false positives nor false negatives significantly compromise the eventual execution of the fault-adjusted speed control system. Future research will explore strategies such as threshold optimization and ensemble learning techniques to minimize the impact of these misclassifications and further improve system reliability.

TABLE 5 Comparison Table of Classification Accuracy of Different Models.

Status category	SVM (%)	PSO-SVM (%)	IPSO-SVM (%)
Normal	73.58 ± 5.30	86.79 ± 4.05 100.00 ± 1.12[1]	100.00 ± 1.12[1] 100.00 ± 1.12[1]
Lightly clogging	100.00 ± 1.12[1]		
Heavily clogging	60.38 ± 5.86	88.68 ± 3.80	96.23 ± 2.27
Completely blocked	75.93 ± 5.10	85.19 ± 4.23	98.15 ± 1.61
Total	77.46 ± 2.51	90.15 ± 1.78	98.59 ± 0.71

1. For accuracies that are exactly 100.00%, the Rule of Three was applied, i.e., ±(3/n)×100%, to avoid a zero-width confidence interval under the normal approximation.

3.3 Operation speed control experiment

To validate the effectiveness of the cloud server and onboard control system's fault warning-speed control system, this study conducted field tests for unmanned variable-speed rice harvesting. The experimental field was planted with the rice variety Nanjing 9108, with an average plant height of approximately 1.1 meters and an overall moisture content of 36.8%. The unmanned harvester's automatic driving system is detailed in previous research (Zhang et al., 2023, 2025), and the onboard fault warning-speed control system continuously uploads pulley speed data from each component to the cloud server in real-time, while also receiving fault warning statuses from the cloud server. The embedded controller calculates the feeding rate per unit time based on the harvester's real-time speed, assuming constant cutting height and cutting width. It then performs fuzzy logic operations to adjust the target speed based on the fault warning status and controls the HST speed adjustment actuator to regulate the vehicle's speed.

After multiple field trials, the harvester equipped with the fault warning-speed control system did not experience complete blockage. It was able to reduce speed within 0.5–2 seconds during minor clogging events, effectively preventing fault escalation. Once the operational conditions were restored and the fault warning was cleared, the system promptly increased the speed to ensure operational efficiency. These results demonstrate that the system has good real-time responsiveness and robustness, and can effectively enhance the application value of unmanned combine harvesters. The communication process between the cloud server and the onboard terminal is shown in Figure 10a; the cloud server information window is shown in Figure 10b. The experimental simulation of heavily clogging accompanied by black smoke due to belt slippage is shown in Figure 11a; the unmanned harvester reducing speed and waiting for fault recovery is shown in Figure 11b; and the fault risk state and speed variation trend during the speed adjustment process is shown in Figure 11c.

Field tests confirm that this system can effectively prevent complete blockage in the harvester, ensuring long-term stable operation of the unmanned harvester. Moreover, this approach demonstrates promising potential for integration into smart farming platforms,

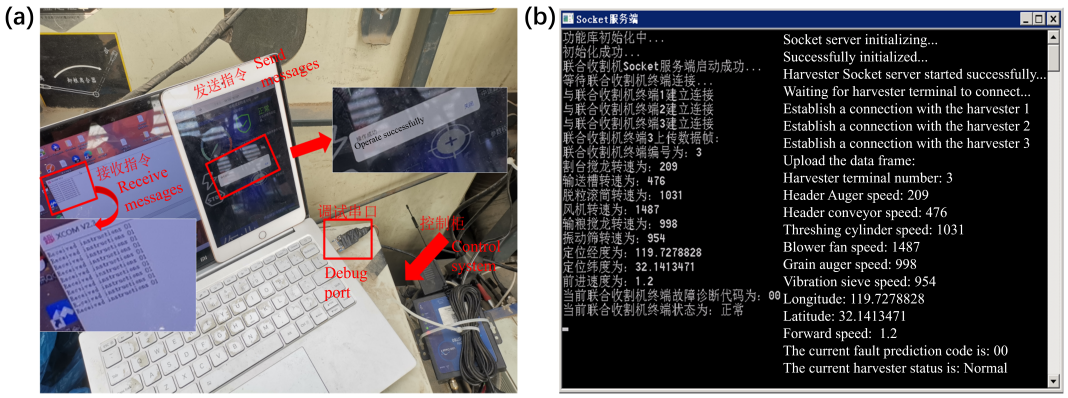


FIGURE 10
Cloud server communication experiment. **(a)** Onboard terminal tests; **(b)** Server communication window

enabling real-time monitoring and predictive maintenance to enhance overall agricultural management efficiency. The proposed system can be scaled effectively to multiple harvesting machines, supporting coordinated operation and grain discharge in cooperative fleets. Additionally, compatibility with emerging IoT and edge computing solutions can further improve real-time decision-making capabilities, expanding the broader impact of this technology in agriculture.

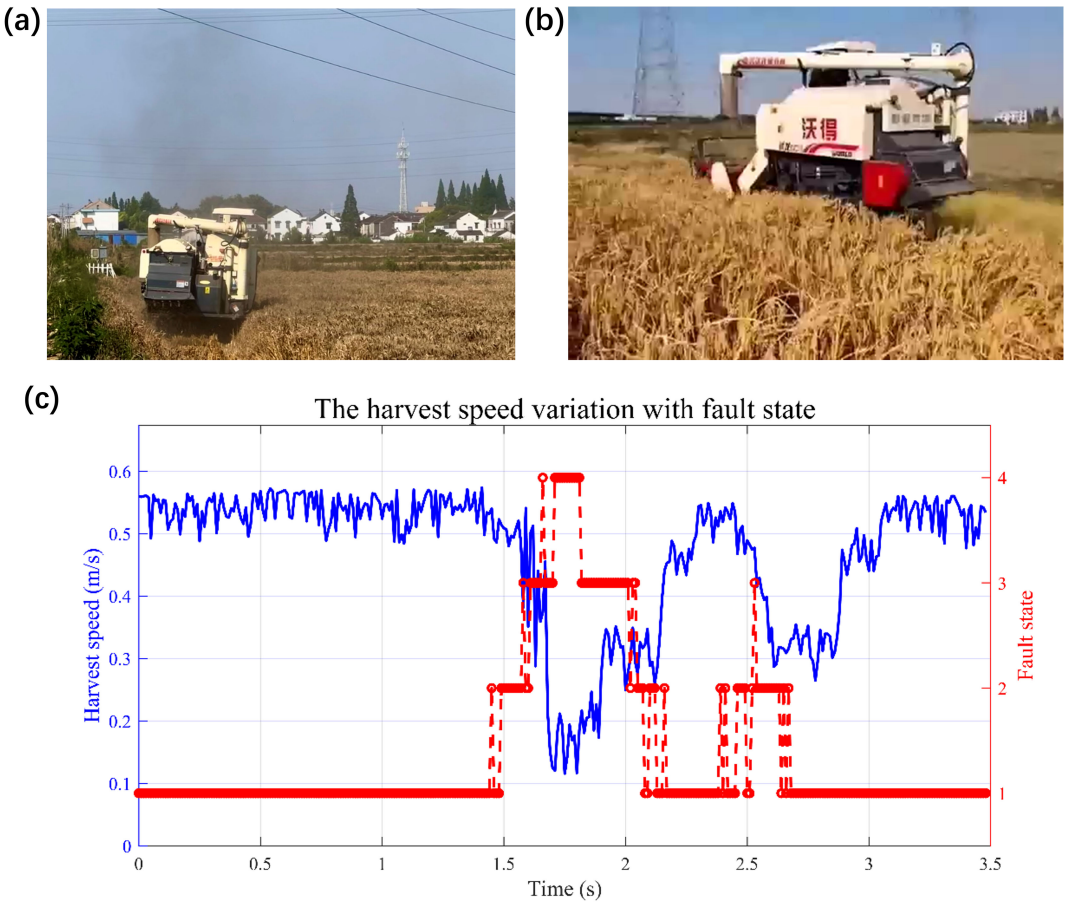


FIGURE 11
Fault prediction and speed regulation experiment. **(a)** Heavily clogging with black smoke caused by belt slippage; **(b)** Unmanned harvesting with speed regulation; **(c)** Plot of harvest speed variation with fault state.

However, there are still some limitations:

1. The real-time calculation of feeding rate in this study relies on prior field calibration data, which is influenced by multiple factors such as overall moisture content, growth density, lodging condition, and the harvester's technical state. Therefore, the application of this fault warning-speed control system requires reliable and user-friendly feeding rate prediction methods. Future research will focus on further exploration in this area.
2. The dataset used in this study comprises only 1,071 data points, which may not be sufficient to robustly validate the performance of a machine learning model. Moreover, these data were collected primarily from a single experimental field, potentially limiting the generalizability of the model to different environmental and operational conditions. Future research will aim to collect more extensive data from multiple fields and varied operational conditions to enhance the robustness and generalization of the model.
3. The IPSO-SVM-based fault warning model classifies fault states by analyzing the slip rates of multiple component pulleys, with manual labeling used for fault classification in the training set. While IPSO-SVM models offer better interpretability and are well-suited for small datasets, they rely on feature engineering and may struggle to capture complex, high-dimensional relationships as effectively as deep learning models (Chen et al., 2019; Ding et al., 2022; Li et al., 2025). Conversely, deep learning approaches, such as CNNs and LSTMs, excel in feature extraction but require large-scale training datasets and substantial computational power. Given the limited dataset size in this study, directly implementing deep learning models would pose challenges in both model training and generalization (He et al., 2023; Kakhi et al., 2024; Xiong et al., 2023). Moreover, deploying the IPSO-SVM model across different harvester types and enhancing fault classification accuracy would significantly increase the manual labeling workload. Future research will focus on developing automated and intelligent fault data labeling methods, as well as exploring hybrid approaches that integrate deep learning for feature extraction with SVM for classification, aiming to further improve model performance and adaptability.
4. The communication between the onboard system and the cloud server relies on the 4G DTU module, and its timeliness is significantly affected by on-site signal coverage. In areas without 4G signal availability, the transmission of remote monitoring commands and fault warning information will be delayed or lost, thereby compromising the real-time performance and reliability of the system. To address this limitation, future research will focus on deploying a simplified fault warning model directly on the onboard embedded system to ensure basic fault detection and speed control functions even under poor communication conditions.
5. Energy consumption considerations remain to be further evaluated. While the proposed system introduces minimal additional power demands due to the use of low-power embedded sensors, its adaptive speed control mechanism may contribute to improved fuel efficiency by reducing unnecessary high-speed operations. However, a more comprehensive real-time analysis of the harvester's power consumption under different operating conditions is needed. Future research will focus on optimizing energy efficiency by integrating real-time power monitoring and adaptive energy management strategies.

4 Conclusion

To improve the stability of unmanned combine harvesters during long-term field operations and maintain harvesting speed and efficiency while avoiding clogging failures, this study developed a fault warning and speed control system based on the IPSO-SVM predictive model and Fuzzy control algorithm. The system was evaluated through extensive field trials, demonstrating its ability to effectively prevent complete clogging failures and ensure continuous operation. The main conclusions are as follows:

1. Clogging state prediction based on multi-component slip rate fusion: By monitoring the rotational speeds of key components such as the Blower Fan, Header Auger, Header Conveyor, Threshing Cylinder, Vibration Sieve, Grain Auger, and Straw Cutter, and extracting slip rate features, an IPSO-SVM model was developed that accurately identifies over 98.5% of fault states.
2. Feeding rate calibration and fuzzy speed control strategy: Field experiments demonstrated that the cutting height and crop mass per unit area have a quadratic relationship. This relationship, combined with real-time speed, can approximate the feeding rate. Based on the Fuzzy control strategy, the system intelligently adjusts the speed by considering both feeding rate and fault prediction states. The system successfully reacts within 0.5–2 seconds following minor clogging, preventing fault escalation and significantly reducing the occurrence of complete clogging failures.
3. Engineering application feasibility and future adaptability: The system's reliability and scalability were validated in field environments by integrating the harvester's electronic control modifications, onboard embedded controllers, 4G communication, and cloud servers. While the current study focused on a specific harvester model, future research will explore its adaptability across different harvester types to verify its broader applicability.
4. Future research directions: To further enhance the system's performance, future work will focus on (1) testing on different harvester models to verify adaptability, (2) improving real-time processing capabilities for cloud-

integrated control, and (3) integrating deep learning techniques to refine fault prediction accuracy. These advancements will contribute to optimizing fault prediction efficiency and ensuring more effective real-time operational adjustments.

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

SZ: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. CZ: Conceptualization, Software, Validation, Writing – review & editing. ZY: Investigation, Validation, Visualization, Writing – review & editing. LT: Data curation, Investigation, Writing – review & editing. KW: Investigation, Validation, Writing – review & editing. AW: Data curation, Writing – review & editing. WC: Data curation, Writing – review & editing. QS: Writing – review & editing. XW: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This project is supported in part by the National Key Research and Development Program of China under Grant (2022ZD0115804), and Project of Faculty of Agricultural Engineering of Jiangsu University under Grant (NGXB20240105). We just would like to thank Jiangsu Runguo Agricultural Development Co., Ltd. for providing the experimental field.

Acknowledgments

We would like to thank Jiangsu Runguo Agricultural Development Co., Ltd. for providing the experimental 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.

Generative AI statement

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1577175/full#supplementary-material

References

Abdeen, M. A., Xie, G., Salem, A. E., Fu, J., and Zhang, G. (2022). Longitudinal axial flow rice thresher feeding rate monitoring based on force sensing resistors. *Sci. Rep.* 12, 1369. doi: 10.1038/s41598-021-04675-w

Ahmad, H., Sun, J., Nirere, A., Shaheen, N., Zhou, X., and Yao, K. (2021). Classification of tea varieties based on fluorescence hyperspectral image technology and ABC-SVM algorithm. *J. Food Process. Preservation* 45, e15241. doi: 10.1111/jfpp.15241

An, N., Wei, W., Qiao, L., Zhang, F., Christie, P., Jiang, R., et al. (2018). Agronomic and environmental causes of yield and nitrogen use efficiency gaps in Chinese rice farming systems. *Eur. J. Agron.* 93, 40–49. doi: 10.1016/j.eja.2017.11.001

Boursianis, A. D., Papadopoulou, M. S., Diamantoulakis, P., Liopa-Tsakalidi, A., Barouchas, P., Salahas, G., et al. (2022). Internet of things (IoT) and agricultural unmanned aerial vehicles (UAVs) in smart farming: A comprehensive review. *Internet Things* 18, 100187. doi: 10.1016/j.iot.2020.100187

Brown, L. D., Cai, T. T., and DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Ann. Stat* 30, 160–201. doi: 10.1214/aos/1015362189

Chen, Y., Chen, L., Huang, C., Lu, Y., and Wang, C. (2019). A dynamic tire model based on HPSO-SVM. *Int. J. Agric. Biol. Eng.* 12, 36–41. doi: 10.25165/j.ijabe.20191202.3227

Chen, J., Gong, L., and Li, Y. (2014). Research of fault diagnosis system on combine-harvester based on FNN algorithm. *China Measurement Test* 40 (5), 77–79.

Chen, J., Ning, X., Li, Y., Yang, G., Wu, P., and Chen, S. (2017). A fuzzy control strategy for the forward speed of a combine harvester based on KDD. *Appl. Eng. Agric.* 33, 15–22. doi: 10.13031/issn.0883-8542

Chen, D. R., Wu, Q., Ying, Y., and Zhou, D. X. (2004). Support vector machine soft margin classifiers: error analysis. *J. Mach. Learn. Res.* 5, 1143–1175.

Chen, L., Zhang, Z., Li, H., and Zhang, X. (2023). Maintenance skill training gives agricultural socialized service providers more advantages. *Agriculture* 13, 135. doi: 10.3390/agriculture13010135

Chen, J., Zhang, M., Xu, B., Sun, J., and Mujumdar, A. S. (2020). Artificial intelligence assisted technologies for controlling the drying of fruits and vegetables using physical fields: A review. *Trends Food Sci. Technol.* 105, 251–260. doi: 10.1016/j.tifs.2020.08.015

Cheng, R., and Jin, Y. (2015). A social learning particle swarm optimization algorithm for scalable optimization. *Inf. Sci.* 291, 43–60. doi: 10.1016/j.ins.2014.08.039

Craessaerts, G., de Baerdemaeker, J., Missotten, B., and Saeys, W. (2010a). Fuzzy control of the cleaning process on a combine harvester. *Biosyst. Eng.* 106, 103–111. doi: 10.1016/j.biosystemseng.2009.12.012

Craessaerts, G., De Baerdemaeker, J., and Saeys, W. (2010b). Fault diagnostic systems for agricultural machinery. *Biosyst. Eng.* 106, 26–36. doi: 10.1016/j.biosystemseng.2009.12.004

Diez-Olivan, A., Pagan, J. A., Khoa, N. L. D., Sanz, R., and Sierra, B. (2018). Kernel-based support vector machines for automated health status assessment in monitoring sensor data. *Int. J. Advanced Manufacturing Technol.* 95, 327–340. doi: 10.1007/s00170-017-1204-2

Ding, Y., Yan, Y., Li, J., Chen, X., and Jiang, H. (2022). Classification of tea quality levels using near-infrared spectroscopy based on CLPSO-SVM. *Foods* 11, 1658. doi: 10.3390/foods11111658

Elijah, O., Rahman, T. A., Orikumhi, I., Leow, C. Y., and HIndia, M. N. (2018). An overview of Internet of Things (IoT) and data analytics in agriculture: Benefits and challenges. *IEEE Internet things J.* 5, 3758–3773. doi: 10.1109/JIoT.6488907

Hao, S., Tang, Z., Guo, S., Ding, Z., and Su, Z. (2022). Model and method of fault signal diagnosis for blockage and slippage of rice threshing drum. *Agriculture* 12, 1968. doi: 10.3390/agriculture12111968

He, Y., Li, W., Dong, Z., Zhang, T., Shi, Q., Wang, L., et al. (2023). Lithologic identification of complex reservoir based on PSO-LSTM-FCN algorithm. *Energies* 16, 5. doi: 10.3390/en16052135

Hofmann, T., Schölkopf, B., and Smola, A. J. (2008). Kernel methods in machine learning. 1171–1220. doi: 10.1214/009053607000000677

Hung, W. L., and Wu, J. W. (2002). Correlation of intuitionistic fuzzy sets by centroid method. *Inf. Sci.* 144, 219–225. doi: 10.1016/S0020-0255(02)00181-0

Jack, L. B., and Nandi, A. K. (2001). Support vector machines for detection and characterization of rolling element bearing faults. *Proc. Institution Mechanical Engineers Part C: J. Mechanical Eng. Sci.* 215, 1065–1074. doi: 10.1177/095440620121500907

Jin, Y., Liu, J., Xu, Z., Yuan, S., Li, P., and Wang, J. (2021). Development status and trend of agricultural robot technology. *Int. J. Agric. Biol. Eng.* 14, 1–19. doi: 10.25165/j.ijabe.20211404.6821

Kakhi, K., Jagatheesaperumal, S. K., Khosravi, A., Alizadehsani, R., and Acharya, U. R. (2024). Fatigue monitoring using wearables and AI: Trends, challenges, and future opportunities. *arXiv preprint* arXiv:2412.16847

Kruse, J., Krutz, G. W., and Huggins, L. F. (1983). Computer controls for the combine. p. 7–9.

Li, R., Cheng, Y., Xu, J., Li, Y., Ding, X., and Zhao, S. (2021). Research on on-line monitoring system of hydraulic actuator of combine harvester. *Processes* 10, 35. doi: 10.3390/pr10010035

Li, Y., Liu, Y., Ji, K., and Zhu, R. (2022). A fault diagnosis method for a differential inverse gearbox of a crawler combine harvester based on order analysis. *Agriculture* 12, 1300. doi: 10.3390/agriculture12091300

Li, D., Nanseki, T., Chomei, Y., and Kuang, J. (2023). A review of smart agriculture and production practices in Japanese large-scale rice farming. *J. Sci. Food Agric.* 103, 1609–1620. doi: 10.1002/jsfa.v103.4

Li, J., Shang, Z., Li, R., and Cui, B. (2022). Adaptive sliding mode path tracking control of unmanned rice transplanter. *Agriculture* 12, 1225. doi: 10.3390/agriculture12081225

Li, L., You, W., and Ding, Y. (2025). Fault prediction modeling for high-impact recorders based on IPSO-SVM. *Appl. Sci.* 15, 3. doi: 10.3390/app15031343

Liang, Z., Li, Y., Xu, L., and Zhao, Z. (2016). Sensor for monitoring rice grain sieve losses in combine harvesters. *Biosyst. Eng.* 147, 51–66. doi: 10.1016/j.biosystemseng.2016.03.008

Liu, Z., Cao, H., Chen, X., He, Z., and Shen, Z. (2013). Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings. *Neurocomputing* 99, 399–410. doi: 10.1016/j.neucom.2012.07.019

Longaray, A. A., Gomes, C. F. S., Elacoste, T., MaChado, C. M., and dos, S. (2019). Efficiency indicators to evaluate services in port services: A proposal using fuzzy-ahp approach. *Pesquisa Operacional* 39, 437–456. doi: 10.1590/0101-7438.2019.039.03.0437

Lowenberg-DeBoer, J., Huang, I. Y., Grigoriadis, V., and Blackmore, S. (2020). Economics of robots and automation in field crop production. *Precis. Agric.* 21, 278–299. doi: 10.1007/s11119-019-09667-5

Lu, E., Xu, L., Li, Y., Tang, Z., and Ma, Z. (2020). Modeling of working environment and coverage path planning method of combine harvesters. *Int. J. Agric. Biol. Eng.* 13, 132–137. doi: 10.25165/j.ijabe.20201302.5210

Ma, Z., Zhang, Z., Zhang, Z., Song, Z., Liu, Y., Li, Y., et al. (2023). Durable testing and analysis of a cleaning sieve based on vibration and strain signals. *Agriculture* 13, 2232. doi: 10.3390/agriculture13122232

Marinoudi, V., Sørensen, C. G., Pearson, S., and Bochtis, D. (2019). Robotics and labour in agriculture. A context consideration. *Biosyst. Eng.* 184, 111–121. doi: 10.1016/j.biosystemseng.2019.06.013

McGechan, M. B., and Glasbey, C. A. (1982). The benefits of different speed control systems for combine harvesters. *J. Agric. Eng. Res.* 27, 537–552. doi: 10.1016/0021-8634(82)90093-2

Nakisa, B., Ahmad Nazri, M. Z., Rastgoo, M. N., and Abdullah, S. (2014). A survey: Particle swarm optimization based algorithms to solve premature convergence problem. *J. Comput. Sci.* 10, 1758–1765. doi: 10.3844/jcssp.2014.1758.1765

Pavlyuk, R. V., Zaharin, A. V., Gevora, Y. I., and Iskenderov, R. R. (2022). "Researching operational reliability of key combine systems in Stavropol Territory," in *IOP Conference Series: Earth and Environmental Science* (Vol. 996, No. 1, p. 012011). IOP Publishing. Russian Conference on Technological Solutions and Instrumentation for Agribusiness (TSIA 2021), Stavropol, Russia.

Peng, Y., Zhao, S., and Liu, J. (2021). Fused-deep-features based grape leaf disease diagnosis. *Agronomy* 11, 2234. doi: 10.3390/agronomy11112234

Pinochet, L. H. C., Onusic, L. M., Costa, J. C. Z., dos Santos, M., Gomes, C. F. S., and Moreira, M.Â.L. (2023). Design a fuzzy-topsis (ftopsis) model in decision-making with multiple criteria for the implementation of telecommuting in a public higher education institute. *Proc. Comput. Sci.* 221, 426–433. doi: 10.1016/j.procs.2023.07.057

Platt, J. (1998). Sequential minimal optimization: A fast algorithm for training support vector machines.

Qi, Z., Wu, X., Yang, Y., Wu, B., and Fu, H. (2022). Discrimination of the red jujube varieties using a portable NIR spectrometer and fuzzy improved linear discriminant analysis. *Foods* 11, 763. doi: 10.3390/foods11050763

Qin, Y., Zhao, D. A., and Ji, W. (2011). Combines speed servocontrol based on RBF. *Advanced Materials Res.* 225, 1120–1124. doi: 10.4028/www.scientific.net/AMR.225-226

Qiu, Z., Shi, G., Zhao, B., Jin, X., and Zhou, L. (2022). Combine harvester remote monitoring system based on multi-source information fusion. *Comput. Electron. Agric.* 194, 106771. doi: 10.1016/j.compag.2022.106771

Rockafellar, R. T. (1993). Lagrange multipliers and optimality. *SIAM Rev.* 35, 183–238. doi: 10.1137/1035044

Samanta, B. I. S. W. A. J. I. T., Al-Balushi, K. R., and Al-Araimi, S. A. (2003). Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Eng. Appl. Artif. Intell.* 16, 657–665. doi: 10.1016/j.engappai.2003.09.006

Schwegman, K., Spinelli, R., Magagnotti, N., Ramantswana, M., and McEwan, A. (2021). Selecting successful harvester operators through aptitude tests and demographics. *Aust. Forestry* 84, 25–32. doi: 10.1080/00049158.2020.1837492

Shawe-Taylor, J., and Sun, S. (2011). A review of optimization methodologies in support vector machines. *Neurocomputing* 74, 3609–3618. doi: 10.1016/j.neucom.2011.06.026

Song, X., Li, H., Chen, C., Xia, H., Zhang, Z., and Tang, P. (2022). Design and experimental testing of a control system for a solid-fertilizer-dissolving device based on fuzzy PID. *Agriculture* 12, 1382. doi: 10.3390/agriculture12091382

Sun, Y., Cui, B., Ji, F., Wei, X., and Zhu, Y. (2022). The full-field path tracking of agricultural machinery based on PSO-enhanced fuzzy stanley model. *Appl. Sci.* 12, 7683. doi: 10.3390/app12157683

Tey, Y. S., and Brindal, M. (2022). A meta-analysis of factors driving the adoption of precision agriculture. *Precis. Agric.* 23, 353–372. doi: 10.1007/s11119-021-09840-9

Wang, F., Liu, Y., Li, Y., and Ji, K. (2023). Research and experiment on variable-diameter threshing drum with movable radial plates for combine harvester. *Agriculture* 13, 1487. doi: 10.3390/agriculture13081487

Wu, X., Zhu, J., Wu, B., Zhao, C., Sun, J., and Dai, C. (2019). Discrimination of Chinese liquors based on electronic nose and fuzzy discriminant principal component analysis. *Foods* 8, 38. doi: 10.3390/foods8010038

Xiong, P., Li, Z., Li, Y., Huang, S., Liu, C., and Gu, F. (2023). Fault diagnosis of UAV based on adaptive siamese network with limited data. *IEEE Trans. Instrumentation Measurement* 72, 1–11.

Xu, L., Wei, C., Liang, Z., Chai, X., Li, Y., and Liu, Q. (2019). Development of rapeseed cleaning loss monitoring system and experiments in a combine harvester. *Biosyst. Eng.* 178, 118–130. doi: 10.1016/j.biosystemseng.2018.11.001

Yang, L., Yaoming, L., Liya, L., and Maolin, S. (2024). Modeling and control methods of a multi-parameter system for threshing and cleaning in grain combine harvesters. *Comput. Electron. Agric.* 225, 109251. doi: 10.1016/j.compag.2024.109251

Yanxin, Y. I. N., Zhijun, M. E. N. G., Chunjiang, Z. H. A. O., Hao, W. A. N. G., Changkai, W. E. N., Jingping, C. H. E. N., et al. (2022). State-of-the-art and prospect of research on key technical for unmanned farms of field corp. *Smart Agric.* 4, 1.

Yu, Z., Li, Y., Du, X., and Liu, Y. (2024). Threshing cylinder unbalance detection using a signal extraction method based on parameter-adaptive variational mode decomposition. *Biosyst. Eng.* 244, 26–41. doi: 10.1016/j.biosystemseng.2024.05.010

Zang, C., Wei, X., Li, L., Hu, C., and Tong, H. (2022). Research on remote fault diagnosis system of harvester. *J. Physics: Conf. Ser.* 2417, 012023.

Zhang, S., Wei, X., Deng, Y., Zang, C., Ji, X., and Wang, A. (2023). Design and experiment of whole field path tracking algorithm for crawler harvester. *Trans. Chin. Soc. Agric. Eng.* 39, 36–45.

Zhang, S., Wei, X., Liu, C., Ge, J., Cui, X., Wang, F., et al. (2025). Adaptive path tracking and control system for unmanned crawler harvesters in paddy fields. *Comput. Electron. Agric.* 230, 109878. doi: 10.1016/j.compag.2024.109878

Zhang, J., Wu, X., He, C., Wu, B., Zhang, S., and Sun, J. (2024). Near-infrared spectroscopy combined with fuzzy improved direct linear discriminant analysis for nondestructive discrimination of chrysanthemum tea varieties. *Foods* 13, 1439. doi: 10.3390/foods13101439

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 June 2025
DOI 10.3389/fpls.2025.1575019

Check for updates

OPEN ACCESS

EDITED BY
Ning Yang,
Jiangsu University, China

REVIEWED BY
Zhenguo Zhang,
Xinjiang Agricultural University, China
Fei Dai,
Gansu Agricultural University, China

*CORRESPONDENCE
Han Tang
tanghan@neau.edu.cn

RECEIVED 11 February 2025
ACCEPTED 05 May 2025
PUBLISHED 06 June 2025

CITATION
Jia P, Zhu G, Chen J, Wang D, Tang H and Wang J (2025) Quantifying assessment of American ginseng (*Panax quinquefolius L.*) main root bruising based on FEM. *Front. Plant Sci.* 16:1575019. doi: 10.3389/fpls.2025.1575019

COPYRIGHT
© 2025 Jia, Zhu, Chen, Wang, Tang and Wang. 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.

Quantifying assessment of American ginseng (*Panax quinquefolius L.*) main root bruising based on FEM

Pengcheng Jia, Guixuan Zhu, Junzhi Chen, Dong Wang, Han Tang* and Jinwu Wang

College of Engineering, Northeast Agricultural University, Harbin, China

Precise numerical simulation technology enabled the capture of subtle deformations in the American ginseng internal structure, allowing for an accurate assessment of bruise extent. In this study, a bilayer constitutive model of the American ginseng main root was developed through reverse engineering. The model accuracy was validated by velocity, exterior bruise area, and internal bruise area, with the highest velocity error being 3.8 %. Experiments analyzed the dynamic mechanical response of the American ginseng main root during collisions at various drop angles (30°, 50°, 70°, and 90°) and with different contact materials (steel, rubber, wood, and PVC). The effects of various collision conditions on bruise volume and bruise resistance index of the American ginseng main root was examined. The results demonstrated that a greater drop angle results in a larger bruise volume. The maximum bruise volumes were 3583.26 mm^3 for steel and 3062.19 mm^3 for wood. At a 30° drop angle, the bruise resistance index of the American ginseng main root was 59.14 mJ mm^{-3} for rubber and 40.89 mJ mm^{-3} for wood. At a 90° drop angle, the bruise resistance index was 45.72 mJ mm^{-3} for rubber and 35.38 mJ mm^{-3} for wood. At the same drop angles, the bruise resistance index of the American ginseng main root was consistently higher when rubber was used as the contact material. This study provides a scientific basis for the design and optimization of harvesting machinery and its key components, with the objective of effectively controlling the American ginseng bruising problem.

KEYWORDS

pendulum test, collision, LS-DYNA, bruise volume, bruise resistance index

1 Introduction

American ginseng possesses significant medicinal value and is widely utilized in traditional Chinese medicine. Due to its high moisture content at harvest, American ginseng is susceptible to collisions with mechanical components during transportation and storage. Bruising from collisions can cause American ginseng to deteriorate and decay,

negatively impacting its economic value and product quality. Therefore, conducting an in-depth study on American ginseng bruising during harvesting is crucial for optimizing mechanized harvesting methods.

When external forces applied to agricultural products exceed their biological yield limit, it caused cell structure rupture and bruising (Miraei Ashtiani et al., 2019). This not only diminishes the appearance and quality of agricultural products but also increased the rate of spoilage and decomposition, raising concerns over product security (Yousefi et al., 2016). Some researchers investigated various techniques to study bruising in agricultural products during harvesting. Common methods included visual inspection, image processing, and infrared spectroscopy to assess the bruise susceptibility and bruise resistance indices of agricultural products (He et al., 2015). Quantifying agricultural product bruising was typically achieved by measuring or estimating bruise volume (Li and Thomas, 2014). Methods for measuring bruised areas included directly using calipers (Jiménez-Jiménez et al., 2013), capturing and analyzing images with a machine vision system (Komarnicki et al., 2017; Zhou et al., 2016), or calculating bruise volume using the "closed volume method" (Holt and Schoorl, 1982). These methods effectively quantified bruise in agricultural products. However, these methods struggled to accurately assess the degree of bruise in the internal tissues of agricultural products, and their detection and judgment relied heavily on manual experience. Some studies tracked the distribution of contact stress to provide detailed mechanical data, capturing complex behaviors and stress concentration areas, which more accurately predicted agricultural product bruising. For example, Lewis et al. (2008) used haptic detection software to test pressure dispersion during fruit loading, obtaining data on bruised areas. Herold et al. (2001) used Tekscan 5051 flexible grid pressure sensors and ultrasound technology to measure and analyze quasi-static compression contact stress distribution in fruits, thereby predicting bruising. Wu et al. (2012) used pre-scaled sensing films to measure contact stress distribution in fruits during drop collisions, achieving reliable predictions of mechanical bruising. However, equipment complexity, high technical barriers, and costs limit the feasibility of large-scale application in agricultural product bruise prediction. Additionally, dealing with complex fruit shapes and diverse bruise patterns led to issues with measurement accuracy and consistency (Keresztes et al., 2017). These methods posed challenges to achieving efficient and accurate bruise assessment during harvesting. In contrast, numerical simulation technology offered significant advantages in this regard. It quantified deformation, transient stress, and energy characteristics during drops or collisions and effectively characterized nonlinear and transient impact loads (Dimino et al., 2022). Through detailed mathematical models and computer simulations, it provided more accurate and consistent bruise prediction, especially for materials with complex internal structures or anisotropic properties.

Finite element analysis (FEA), as a numerical simulation method, simulated the mechanical behavior of complex structural materials under various loading conditions (Abbaszadeh et al., 2014). In recent years, FEA played a crucial role in heat transfer, vibration analysis, stress distribution, fatigue life prediction, and the study of the mechanical properties of composite materials (Van Zeebroeck et al., 2007). FEA was widely applied in bruise analysis of fresh agricultural products and in solving complex agricultural problems (Zulkifli et al., 2020). Kabas et al. (2008) used FEM to determine the deformation behavior of cherry tomatoes when they hit the ground after falling from a known height. Wu et al. (2013) investigated the pressure dispersion of bitter gourd under static pressure using FEM. According to the experimental results, the relative error in the static pressure bruise area of bitter melon was close to 13 %. Ahmadi et al (2016a) explored variations in factors such as location and rate of fruits during collisions with other fruits and rigid bodies using FEM. Zhao et al (2019a) used 3D scanning and finite element simulations to assess collision bruising in goji berries. They employed the response surface method to determine the influence of drop height, collision material, and angle on bruising. However, the single isotropic homogeneous elastoplastic material models used in these studies were not suitable for materials with complex internal structures or anisotropic properties. The American ginseng main root had an irregular columnar shape, with distinct physical features between the cortex and cambium. Therefore, to investigate the dynamic response of the American ginseng main root under various collision conditions, a bilayer constitutive model was required.

This study aimed to analyze the bruise formation process and bruise resistance of the American ginseng main root under various collision conditions, with the following specific objectives: a) determine the physical features of the American ginseng cortex and cambium and establish an accurate bilayer constitutive model; b) validate the accuracy of the constitutive model by contrasting simulations with pendulum test results; c) conduct a numerical analysis of the dynamic mechanical response of the American ginseng main root during collisions; d) analyze the bruise resistance characteristics of the American ginseng main root under different conditions.

2 Materials and methods

2.1 Determination of physical features of American ginseng main root

2.1.1 Density measurement

Ten American ginseng main roots with rootlets removed were selected as samples for the uniaxial compression test. The cortex and cambium of each American ginseng main root were separated using a cutting tool, and the density of each was measured separately. The density of the American ginseng cortex was measured using the specific gravity bottle method, while the density of the American ginseng cambium was determined using the specific gravity balance method (Xu et al., 2023). The density of the American ginseng cortex and cambium was calculated as shown in Equations 1, 2:

$$\rho_s = m_s\rho_w/(m_{sl} - m_{ol}) \quad (1)$$

$$\rho_g = (m_{bg} - m_b)\rho_w / [(m_w - m_b) - (m_z - m_b)] \quad (2)$$

where ρ_s is the density of the American ginseng cambium (g cm^{-3}), m_s is the mass of the American ginseng cambium (g), m_{ol} is the combined mass of the receptacle and the soaking liquid (g), m_{sl} is the combined mass of the receptacle, soaking liquid, and the American ginseng cambium soaked in the liquid (g), ρ_w is the density of water (g cm^{-3}), ρ_g is the density of the American ginseng cortex (g cm^{-3}), m_{bg} is the combined mass of the pycnometer and the American ginseng cortex (g), m_b is the mass of the pycnometer (g), m_w is the mass of the pycnometer filled with water (g), m_z is the mass of the pycnometer filled with the American ginseng cortex and water (g).

2.1.2 Physical features measurement

Uniaxial compression tests on the cortex and cambium of the American ginseng main root were carried out using a texture analyzer (Stable Micro Systems, England) (Figure 1B). The contact force was set to 0.01 N, and the pressing speed to 10 mm min^{-1}. Stress-strain curves for the cortex (Figure 1C) and cambium (Figure 1D) were generated from the experimental data. Stress in the cortex linear growth with strain up to the biological yield point, where the ratio of stress to strain represents the Young's modulus (tanα) (Rashvand et al., 2022). After the yield point was exceeded, bruising occurred in the cortex. The cambium showed similar physical features. The procedure was performed 10 times, with the mean value calculated to ensure data accuracy. Poisson's ratio was determined by calculating the ratio of lateral strain perpendicular to the load path to axial strain along the load path in the uniaxial compression test.

2.2 Constitutive model of American ginseng main root

This study employed reverse engineering to create a solid model of the American ginseng main root (Guan et al., 2023). A handheld intelligent laser 3-dimensional scanning device (MarvelScan handheld intelligent laser 3D scanner, Shenzhen SR measurement technology Co., Ltd., Shenzhen, China) was employed to collect point cloud data of the American ginseng main root (Figure 2A). After merging the point cloud data, surface splicing was carried out using Geomagic Studio (Geomagic Inc., Morrisville, USA). The American ginseng main root was then solid modelled in SolidWorks 2022 (Dassault Systemes S. A, Massachusetts, USA) to verify the geometric parameters of its outer outline (Figure 2B). Following the application of the techniques—point cloud splicing, surface splicing, and model reconstruction—a three-dimensional solid model comprising two distinct components, the cortex (5 mm) and the cambium (Figure 1A), was ultimately constructed. The cambium was then connected to the cortex through surface contact.

The bilayer constitutive model of the American ginseng main root presented in this study is based on its physical and mechanical properties. In the quasi-static compression test, the American ginseng cambium primarily exhibits elastic deformation, leading

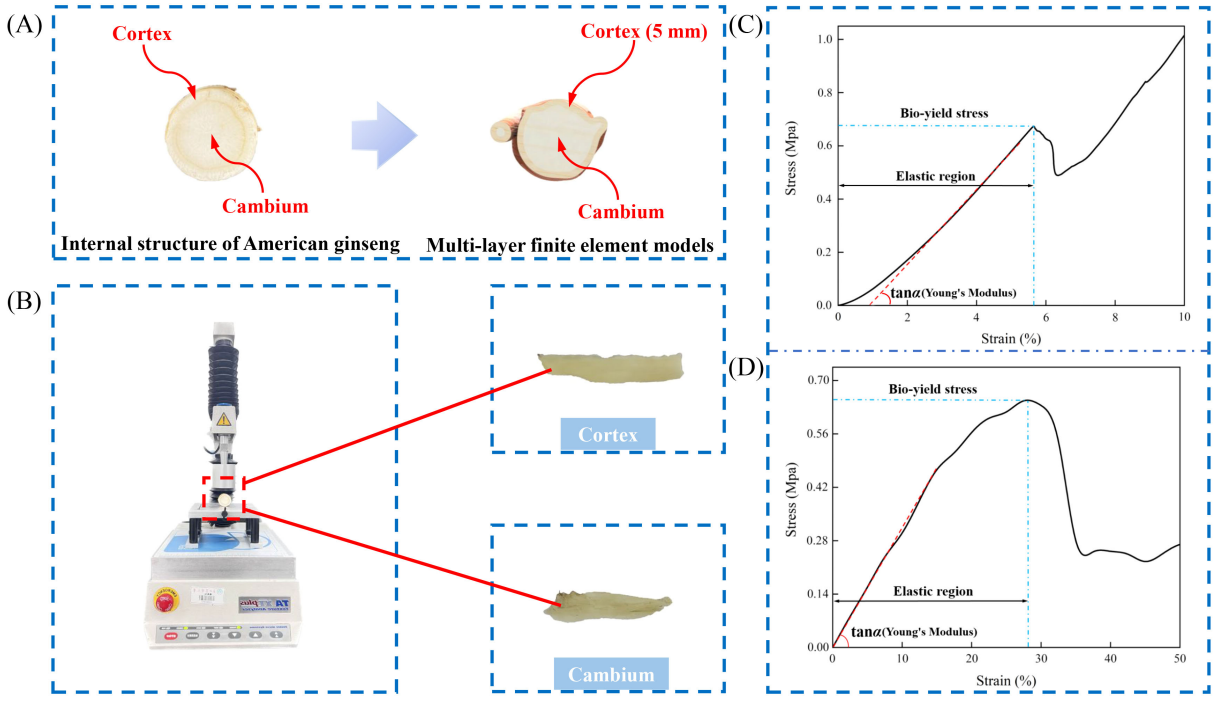


FIGURE 1
Measurement of physical features of American ginseng main root. **(A)** Internal structure of the American ginseng main root; **(B)** Uniaxial compression test of the American ginseng main root cortex and cambium; **(C)** Compression characteristics of the American ginseng main root cortex; **(D)** Compression characteristics of the American ginseng main root cambium.

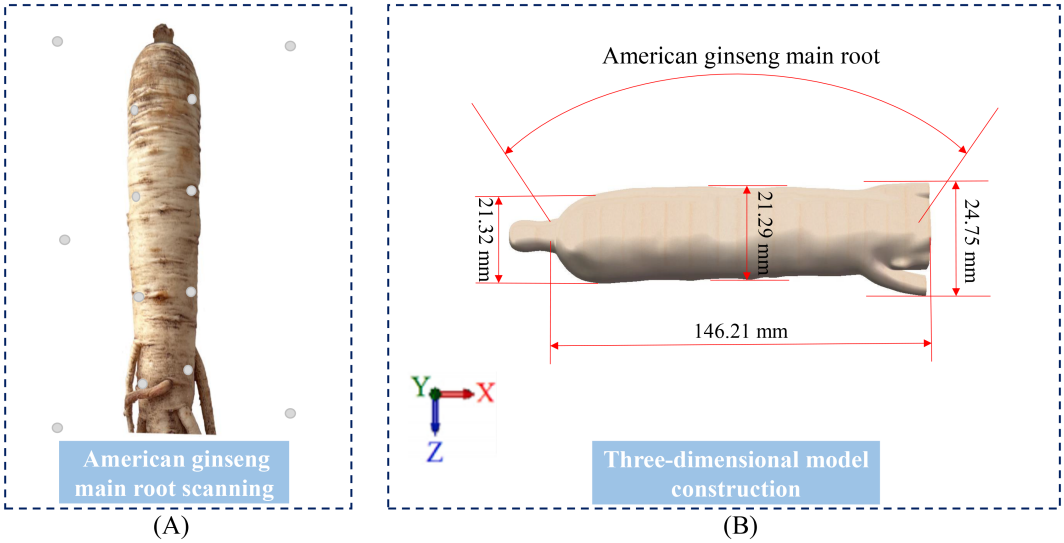


FIGURE 2
Reverse engineering of American ginseng main root. **(A)** Scanning of the American ginseng main root; **(B)** Solid 3D modelling in SolidWorks.

to its classification as an isotropic linear elastic material model. The American ginseng cortex was observed to possess high toughness and some degree of plasticity, allowing for the application of an idealized bilinear isotropic strain-hardening elastic-plastic model.

2.3 Experimental content and methods

2.3.1 Pendulum test

A pendulum test apparatus was designed for collision testing of American ginseng main roots (Figure 3A). The ends of the American ginseng main root sample were attached to the crossbeam at the top of an aluminum profile frame using cotton threads. The American ginseng main root was raised to specific drop angles (30°, 50°, 70°, and 90°) and then allowed to fall freely, striking the square plate installed on the frame. The plate materials were steel, rubber, wood, and PVC (Figure 3C). A high-speed camera (Phantom Multicam, Phantom Video Player, PCC 2.8) mounted on a tripod recorded the collision at 2000 fps and acquired images at a resolution of 1296 × 1024 pixels (Figure 3B). The camera captured the moment of contact during the collision of the American ginseng main root samples and recorded the velocity of the samples before and after the collision. Each drop angle test was repeated ten times.

2.3.2 Finite element simulation of American ginseng pendulum collision

The LS-DYNA unit in ANSYS 2022 (ANSYS, Inc., Pennsylvania, USA) was used for simulations, including 16 different scenarios that considered various materials in contact with the American ginseng main root (steel, rubber, wood, and PVC) and drop angles of 30°, 50°, 70°, and 90°. Considering the deformation process following energy absorption from the collision and the contact rebound phase, the solution time for each scenario was fixed at 0.02 s.

2.4 Simulation model validation test

This study utilized the pendulum test to validate the bruising sustained by American ginseng during transportation and harvesting. The entire collision process was captured using a high-speed camera, and the images were processed to analyze the contours of bruised areas on the exterior of the American ginseng main root (Figure 4). By analyzing the velocity changes in the simulation, Equation 3 was employed to calculate the drop velocity of the American ginseng main root in the test:

$$v_e = (s_{e+1} - s_e)/(t_{e+1} - t_e) \quad (3)$$

where v_e is the velocity at time e (mm s^{-1}), t_{e+1} and t_e are the time points at which the high-speed camera records the falling state of the American ginseng main root (s), s_{e+1} and s_e are the displacements at times $e + 1$ and e (mm).

The relative error calculation formula was employed to determine the velocity difference between the pendulum test and the simulation, thereby validating the accuracy of the established constitutive model, as shown in Equation 4:

$$\sigma_e = (v_e - v_{es})/v_e \times 100\Sigma \quad (4)$$

where σ_e is the velocity error at time e (%), and v_{es} is the velocity at time e in the simulation (mm s^{-1}).

A fruit hardness tester (Zhejiang Top Instrument Co., Ltd., Hangzhou, China) was used to measure the hardness of the American ginseng main root before and after the collision, enabling the calculation of the total bruised volume. The bruised volume of the American ginseng main root was calculated using Equation 5:

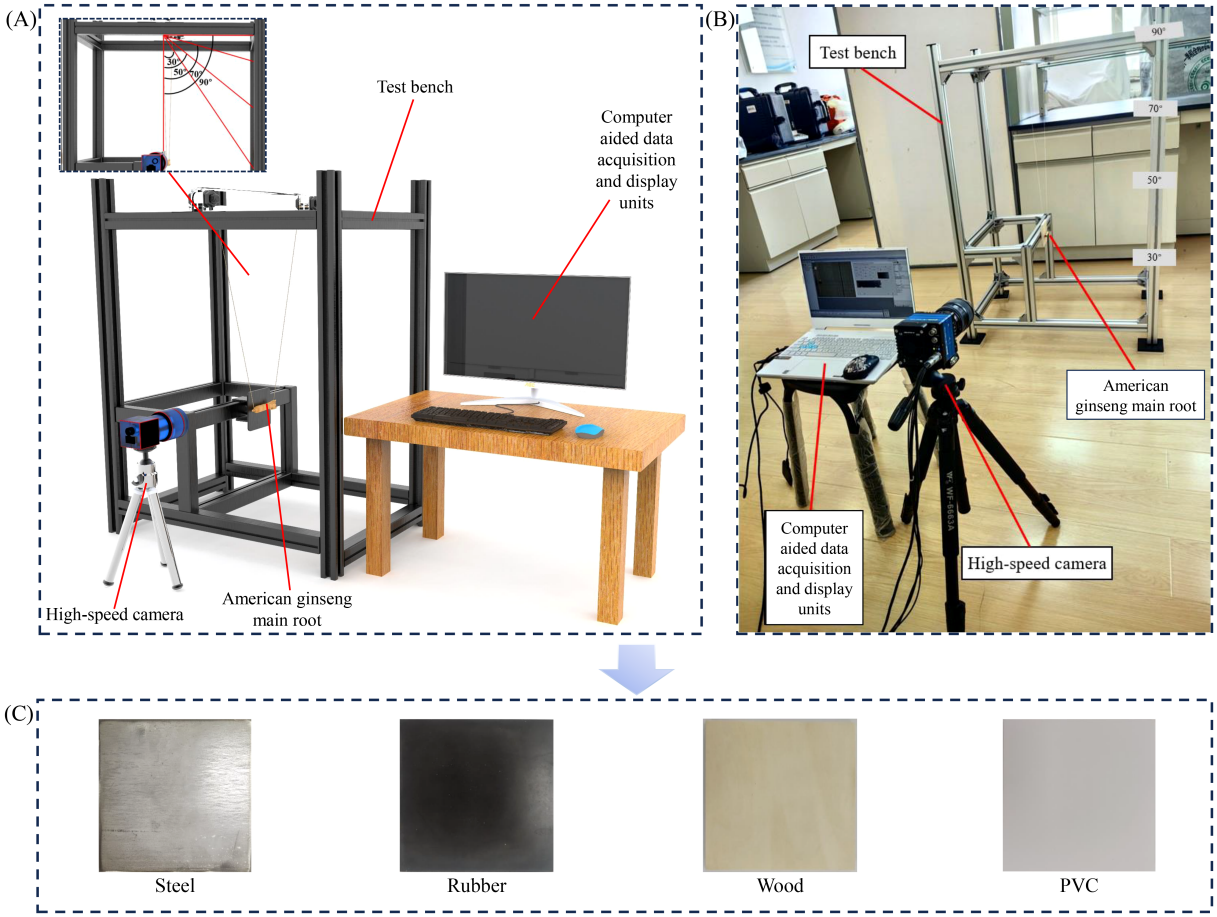


FIGURE 3
Pendulum test and contact materials. **(A)** Pendulum test apparatus; **(B)** Experimental environment; **(C)** Contact materials used in the experiment.

$$BV = \pi(d_b - d_t)(3l_1 w_1 + 4(d_b - d_t)^2)/24 \quad (5)$$

where l_1 and w_1 are the length and width of the bruised area on the exterior of the American ginseng main root (mm), d_b is the bruise deepness of the American ginseng main root after the pendulum test (mm), d_t is the distance from the epidermis to the top of the bruise (mm), and BV is the bruise volume of the American ginseng main root (mm^3).

In the pendulum test, the energy of the American ginseng main root after the collision was derived from its potential energy:

$$E_i = mgH_d = mgl\,(1 - cos\alpha) \quad (6)$$

where E_i is the energy generated by the American ginseng main root after the collision (mJ), m is the mass (g), H_d is the drop height (mm), l is the length of the string (mm), α is the original swing angle (°), and g is the gravitational acceleration (mm s^{-2}).

Following the collision, the American ginseng main root rebounded to its maximum height. The difference between the original and final potential energies, which is absorbed during the collision, was expressed as:

$$E_a = mg(H_d - H_r) = mgl\,(cos\beta - cos\alpha) \quad (7)$$

where E_a is the energy absorbed by the American ginseng main root (mJ), H_r is the maximum rebound height (mm), β is the rebound angle at the top (°).

By combining Equations 6, 7, the total energy during the American ginseng collision process was obtained as:

$$E = E_i + E_a \quad (8)$$

The total energy during the collision of the American ginseng main root (excluding friction and contact energy) was calculated using Equation 8. A vernier calliper (manufactured by Shandong Greener Precision Instruments Co., Ltd., with a precision of 0.01mm) was used to measure the width, length, and depth of the American ginseng main root bruise. The bruise resistance index (*BRI*) of the American ginseng main root was then calculated using the bruise volume and the total energy absorbed in the collision.

BRI was the ratio of the load parameters endured by agricultural products during dynamic collision to the amount of bruised tissue resulting from these parameters (Gołacki et al., 2009), as shown in Equation 9:

$$BRI = E/BV \quad (9)$$

where E is the total energy absorbed by the American ginseng main root during the drop (mJ), BV is the bruise volume of the American ginseng main root after the collision (mm^3), and BRI is the bruise resistance index of the American ginseng main root ($mJ\ mm^{-3}$).

2.5 Calculation of bruise area and extraction of bruise volume

The bruise area of the American ginseng main root was calculated, followed by a quantitative comparison between the test and simulation results. The comparison encompassed the exterior bruise area and the internal bruise area, with the latter obtained by cutting along the centerline of the bruise surface (Figure 4A). The long axis dimensions of the bruised area in the pendulum test were determined using a vernier caliper, and photos of the bruised surface were uploaded to CAXA software (version: 2016, Beijing Digital Dafang Technology Co., Ltd., Beijing, China) to acquire the long axis dimensions and calculate the bruise area. In the simulation, the long axis dimensions of the bruised area were determined using the LS-DYNA unit and then imported into CAXA to calculate the bruise area.

After the collision finite element simulation, the maximum stress was contrasted with the yield stress of the American ginseng cortex and cambium. The results indicated that bruising occurred in regions where the stress surpassed the yield stress. The combined bruised volume of the cortex and cambium constituted the total bruised volume of the American ginseng main root. The bruised portion was extracted from simulation software and saved in STL format. The STL format was then uploaded to SolidWorks and transformed into a STEP format for visualization of the bruised volume (Figure 4B).

3 Results and discussion

3.1 Results of physical feature tests

The material features of the American ginseng cortex and cambium were obtained through uniaxial compression tests. Additionally, the material features of steel, wood, rubber, and PVC were sourced from the ANSYS material library (ANSYS 2022 R1) (Table 1).

3.2 Mesh sensitivity analysis

This study determined the appropriate element size through mesh sensitivity analysis. The model sequentially used element sizes ranging from 1 mm to 5 mm for collision simulations. The smallest element size that adequately represented the American ginseng main root was selected based on the results of the mesh sensitivity analysis and acceptable computational time. The curves of equivalent stress over time for different element sizes were generated (Figure 5A). The maximum equivalent stress was reached with an element size of 4 mm (Figure 5B). Thus, an element size of 4 mm was best suited to capture equivalent stress. The constitutive model of the American ginseng main root in the simulation consisted of 25836 elements with a 4 mm element size.

3.3 Model verification and simulation

3.3.1 Model verification

The collision process of the American ginseng main root can be divided into five stages: fall, contact, collision, detachment, and rebound (Figures 6E, F). The accuracy of the constitutive model was

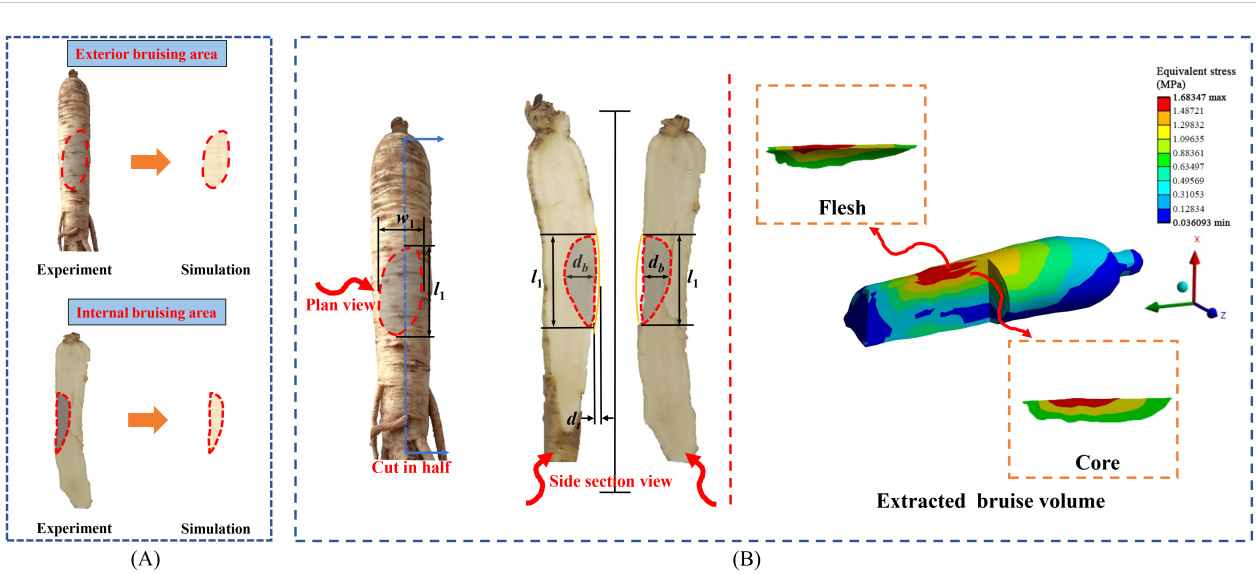


FIGURE 4
Extraction of bruise area and bruise volume of American ginseng main root. **(A)** Comparison of exterior and internal bruise areas in simulation and pendulum test; **(B)** Process of bruise volume extraction.

10.3389/fpls.2025.1575019

TABLE 1 Parameters of American ginseng main root and contact materials.

Materials	Properties	Values
American ginseng main root cortex	Poisson's ratio	0.42
	Density (g cm^{-3})	1.16
	Young's modulus (MPa)	12.11
	Yield strength (MPa)	0.68
American ginseng main root cambium	Poisson's ratio	0.35
	Density (g cm^{-3})	0.95
	Young's modulus (MPa)	2.32
	Yield strength (MPa)	0.65
Steel	Poisson's ratio	0.30
	Density (g cm^{-3})	7.85
	Young's modulus (MPa)	200000
Rubber	Poisson's ratio	0.47
	Density (g cm^{-3})	1.2
	Young's modulus (MPa)	–
Wood	Poisson's ratio	0.38
	Density (g cm^{-3})	0.7
	Young's modulus (MPa)	–
PVC	Poisson's ratio	0.4
	Density (g cm^{-3})	1.39
	Young's modulus (MPa)	2861

confirmed by comparing velocity changes and bruise areas during the collision process.

Velocity-time curves during the collision process were plotted (Figures 6A–D). Velocity gradually increased during the fall and decreased after contact. During the rebound process, the velocity reached its peak value and then gradually declined. The greater the drop angle, the higher the rebound velocity. In the solution process, the nonlinear behavior of materials or contact conditions was not fully or accurately reflected in the model, potentially causing differences between the simulation results and pendulum test results (Yan et al., 2024). A comparison between the simulation and pendulum test results indicated that the highest velocity error during the collision process was 3.8 % (Figure 6D).

Five sets of tests were conducted for each drop angle (with steel as the contact material), and the exterior and internal bruise areas of the American ginseng main root were measured separately, with the mean values calculated. Additionally, the error between the simulation and the pendulum test was calculated (Figures 6G–H).

At a drop angle of 50°, the exterior bruise area of the American ginseng main root was relatively small, measuring 611.74 mm² (simulation) and 614.26 mm² (pendulum test), with the smallest error of 0.4 %. At a drop angle of 90°, the exterior bruise area of the American ginseng main root reached its maximum, measuring 855.74 mm² (simulation) and 874.13 mm² (pendulum test), with the largest error of 2.1 %.

At a drop angle of 30°, the internal bruise area of the American ginseng main root was the smallest, measuring 359.17 mm² (simulation) and 367.21 mm² (pendulum test), with the largest error being 2.2 %. At a drop angle of 50°, the error in the internal bruise area of the American ginseng main root was the smallest, at 0.76 %. At a drop angle of 90°, the internal bruise area of the

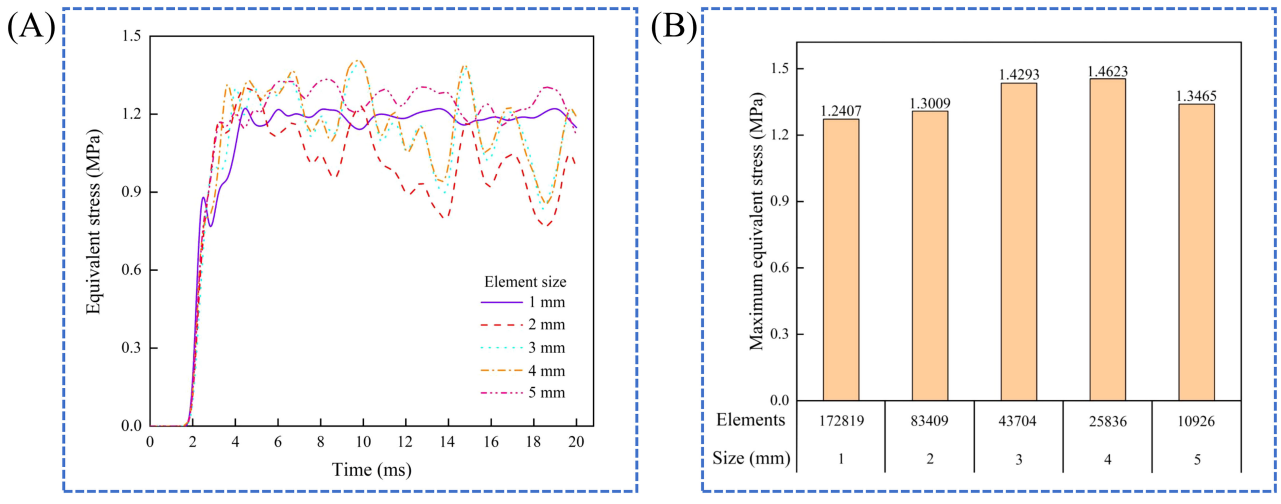


FIGURE 5
Mesh sensitivity analysis. **(A)** Equivalent stress for different element sizes; **(B)** Maximum equivalent stress for different elements counts.

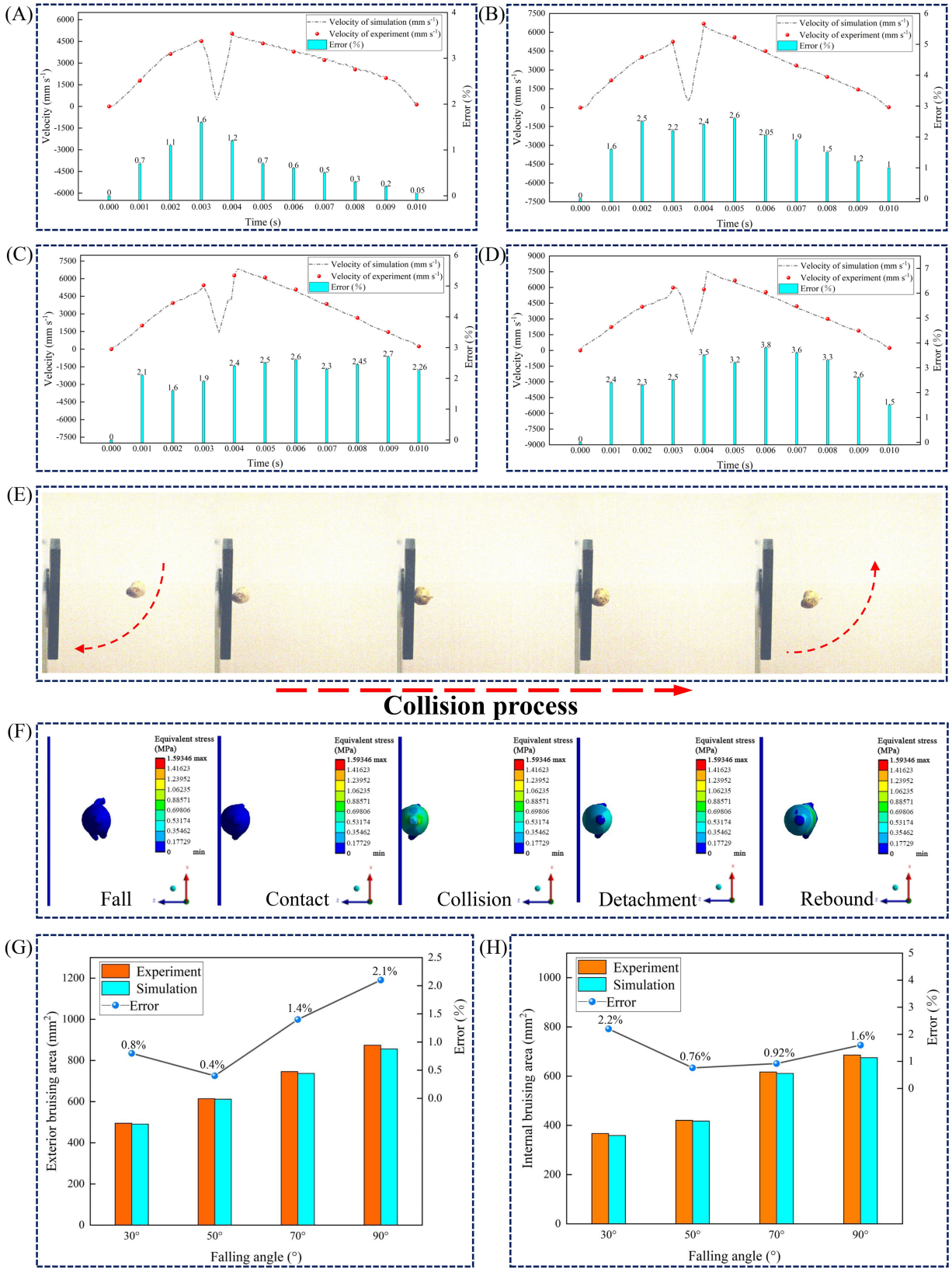


FIGURE 6
Comparison of collision processes and characterization of bruise area. **(A–D)** Velocity changes of the American ginseng main root with steel as the contact material at drop angles of 30°, 50°, 70°, and 90°, respectively; **(E)** Drop process captured by high-speed imaging test; **(F)** Finite element simulation of the American ginseng main root collision process; **(G)** Exterior bruise area and error from simulation and pendulum tests at different drop angles; **(H)** Internal bruise area and error from simulation and pendulum tests at different drop angles.

American ginseng main root reached its maximum, measuring 675.57 mm^2 (simulation) and 686.52 mm^2 (pendulum test).

Within the allowable error range, the comparison of the American ginseng main root velocity, exterior bruise area, and internal bruise area between the simulation and pendulum test demonstrated the accuracy of the constructed constitutive model of the American ginseng main root.

3.3.2 Finite element simulation

The collision process of the American ginseng main root at a 90° drop angle with steel as the contact material was analyzed. Following the collision, the contact force and equivalent stress in the American ginseng main root gradually increased (Figure 7A). At 3.5 × 10^{-3} s, the equivalent stress and contact force in the collision area of the American ginseng main root reached maximum values of 1.235 MPa and 22.152 N, respectively (Figure 7B). The equivalent stress in the non-collision area of the American ginseng main root was minimal, recorded at 4.561 × 10^{-3}MPa. After rebounding, the American ginseng main root gradually detached from the contact material, causing the contact force to sharply drop to zero. Due to the large equivalent stress generated during the collision, it did not dissipate immediately. Instead, the equivalent stress propagated within the American ginseng main root, exhibiting periodic fluctuations that gradually decay to zero.

The variations in four forms of energy during the American ginseng main root collision process were analyzed (Figure 7C). During the collision of the American ginseng main root, energy was absorbed or transformed, causing the kinetic energy to rapidly decrease from its maximum value of 32500 mJ to a minimum. Simultaneously, internal energy increased to a maximum of 16252 mJ. Subsequently, both kinetic energy and internal energy exhibited periodic fluctuations, possibly due to energy changes resulting from elastic or plastic deformation (Liu et al., 2024).

3.4 Dynamic mechanical response of American ginseng main root at different drop angles

Drop angles of 30°, 50°, 70°, and 90° were used as test conditions, with wood as the material. The maximum contact force, maximum equivalent stress, and maximum internal energy were analyzed over time (Figure 8). When the American ginseng main root contacted the wooden board, the contact force instantly reached its peak value. As time progressed, the contact force gradually decreased until the American ginseng main root detached from the contact material, at which point the contact force dropped to its minimum (Figure 8A). Linear fitting results indicated that the maximum contact force of the American ginseng main root was positively correlated with the drop angle (Figure 8B). The greater the drop angle, the higher the peak of the maximum contact force. At a drop angle of 90°, the maximum contact force was 19.61 N, while at a drop angle of 30°, it was 13.58 N.

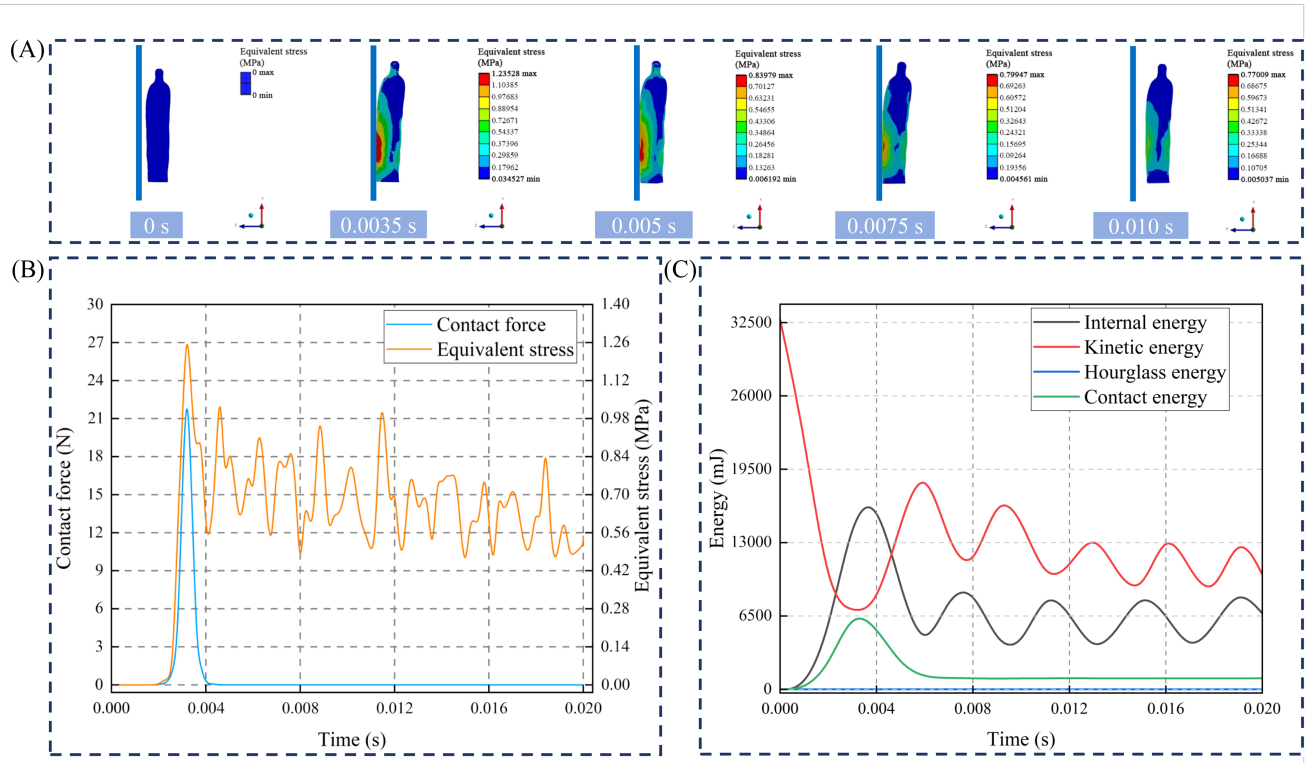


FIGURE 7
Forces and energy during American ginseng main root collision. **(A)** Finite element simulation of American ginseng main root collision; **(B)** Time-dependent curves of equivalent stress and contact force during the collision; **(C)** Curves of the four types of energy variations during the collision.

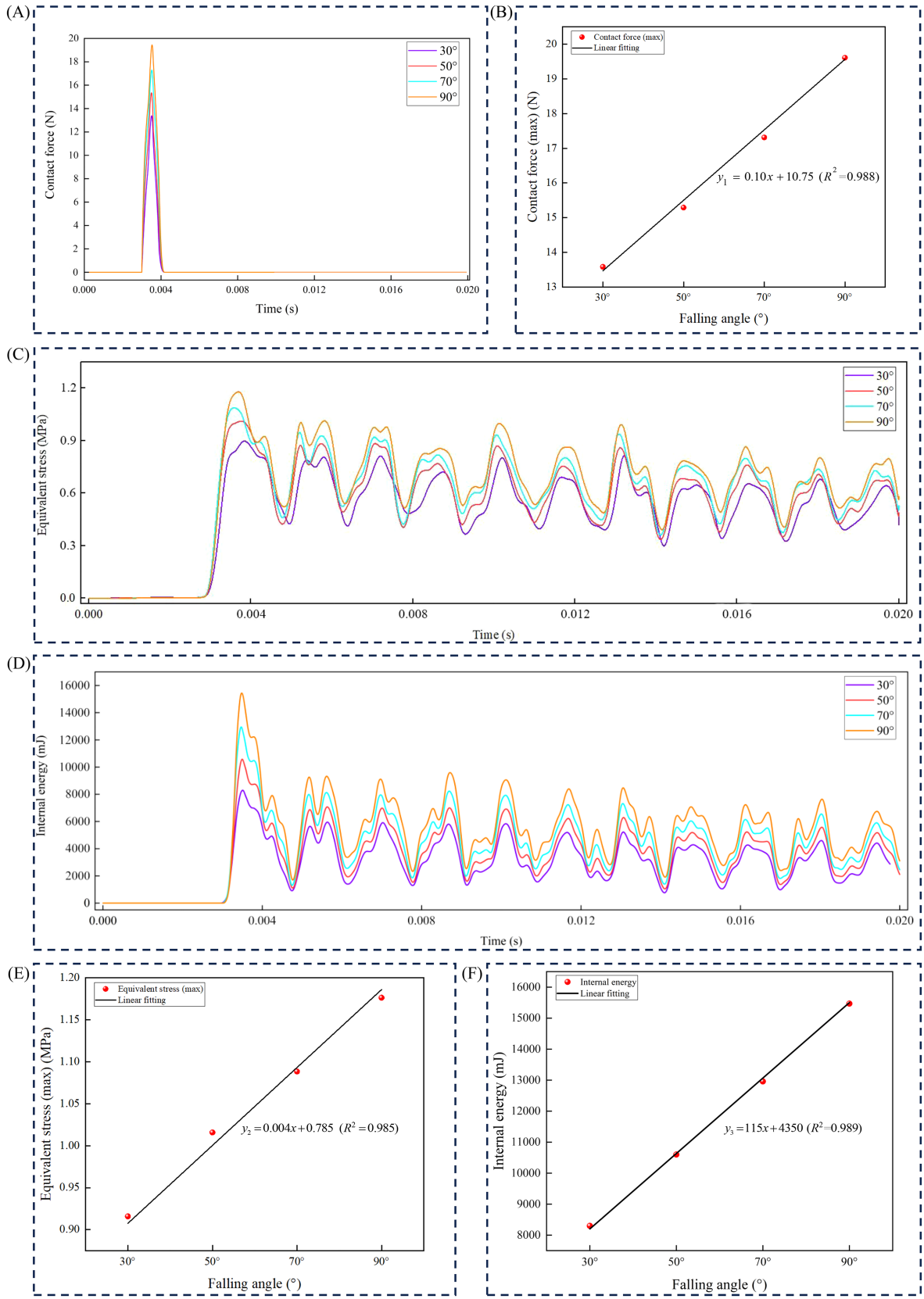


FIGURE 8
Relationship between maximum contact force, maximum equivalent stress, and maximum internal energy with drop angle. **(A)** Curve of maximum contact force as time progresses; **(B)** Relationship between maximum contact force (y_1) and different drop angles (x); **(C)** Curve of maximum equivalent stress as time progresses; **(D)** Curve of maximum internal energy as time progresses; **(E)** Relationship between maximum equivalent stress (y_2) and different drop angles (x); **(F)** Relationship between maximum internal energy (y_3) and different drop angles (x).

The maximum equivalent stress peaked at 3.5×10^{-3}s and gradually decreased over time. Due to residual stress after the collision, the equivalent stress exhibited periodic fluctuations (Figure 8C). At a drop angle of 90°, the maximum equivalent stress was 1.176 MPa, while at a drop angle of 30°, it was 0.916 MPa. Linear fitting results indicated that the maximum equivalent stress of the American ginseng main root was positively correlated with the drop angle (Figure 8E). The greater the drop angle, the higher the peak of the maximum equivalent stress. The variations in the maximum internal energy of the American ginseng main root over time at various drop angles were analyzed (Figure 8D). During a collision, the maximum internal energy rapidly increased, peaking at 3.5×10^{-3}s, and then gradually decreased with periodic fluctuations. With the increase in drop angles, the peak value of the maximum internal energy also increased. At a drop angle of 90°, the maximum internal energy was 15468 mJ, while at a drop angle of 30°, it was 8293 mJ. Linear fitting results indicated that the maximum internal energy was positively correlated with the drop angle (Figure 8F).

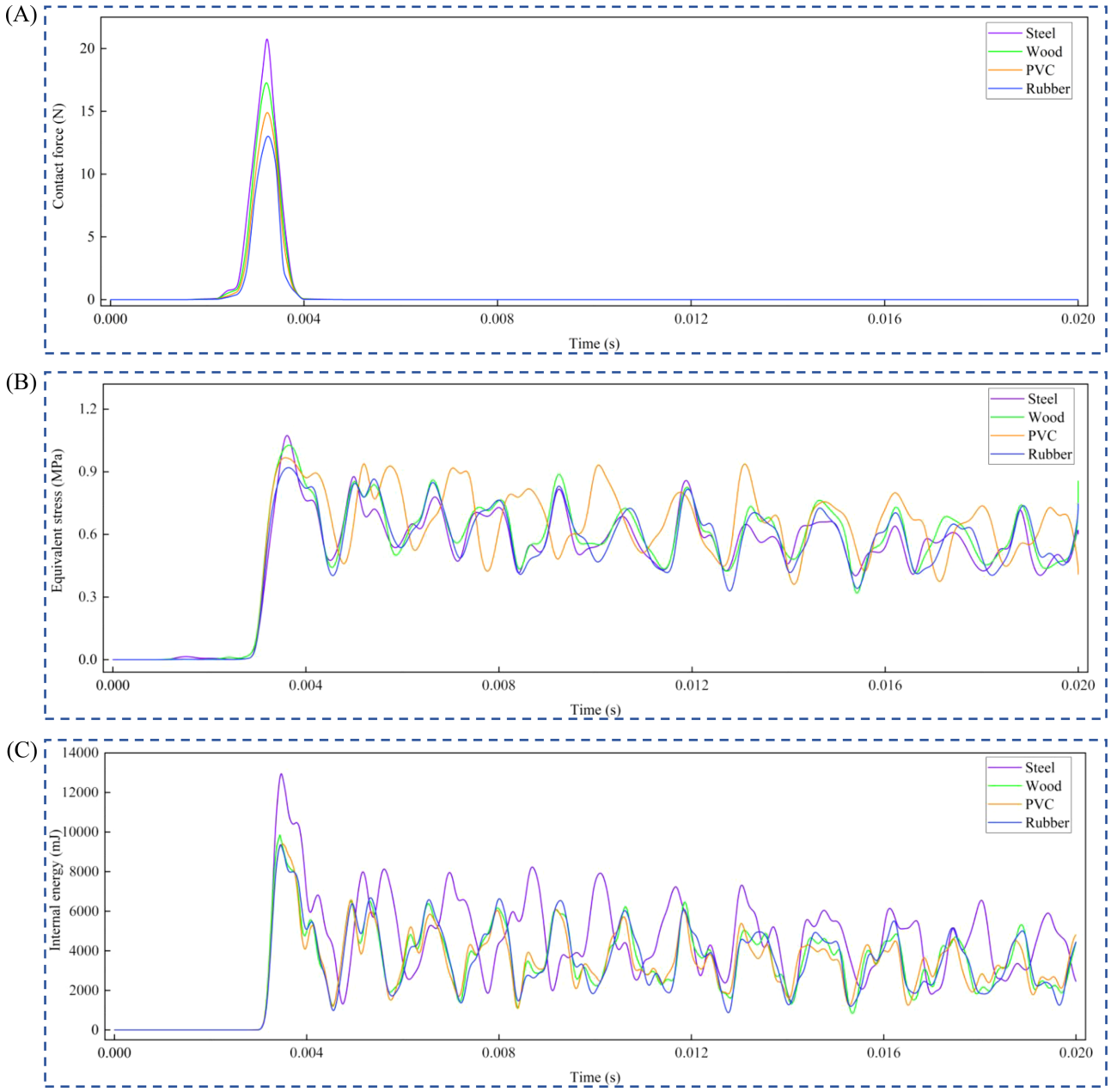


FIGURE 9
Variation of maximum contact force, maximum equivalent stress and maximum internal energy with time for different contact materials. **(A)** Curve of maximum contact force as time progresses; **(B)** Curve of maximum equivalent stress as time progresses; **(C)** Curve of maximum internal energy as time progresses.

3.5 Dynamic mechanical response of American ginseng main root under different contact materials

The materials used were steel, wood, PVC, and rubber at a drop angle of 70°. The maximum contact force, maximum equivalent stress, and maximum internal energy were analyzed over time (Figures 9A–C). The maximum contact force was highest when the contact material was steel, at 20.215 N, and lowest when the contact material was rubber, at 12.613 N (Figure 9A). The maximum equivalent stress was highest when the American ginseng main root was in contact with steel, at 1.183 MPa, and lowest at 0.928 MPa when in contact with rubber (Figure 9B). As the hardness of the contact material higher, the maximum equivalent stress and maximum contact force generated during the collision of the American ginseng main root also increased, leading to greater bruising. Additionally, the maximum internal energy was highest when the contact material was steel, at 12959 mJ, and lowest when the contact material was rubber, at 8570.5 mJ (Figure 9C).

3.6 Bruise resistance characteristics of American ginseng main root

Stress cloud maps of the American ginseng main root at different drop angles and with different contact materials were extracted (Figures 10A–D). The stress distribution in the American ginseng main root was uneven, with stress values gradually decreasing from the center toward the periphery. Stress values in the bruised areas were higher than those in the non-bruised areas; in other words, bruising only occurred where the stress was sufficiently high. This stress distribution pattern led to bruising in specific areas of the American ginseng main root. Therefore, assessing the bruise resistance characteristics of agricultural products requires quantifying the extent of bruise in these specific areas.

The bruise resistance index reflects the amount of bruise per unit of energy induced in crops during a dynamic collision and can be used to assess the degree of bruise in rhizome crops. Bruise volumes under different materials (steel, wood, PVC, and rubber) were statistically analyzed as the drop angle (30°, 50°, 70°, and 90°) varied (Figure 10E). For steel and wood, bruise volume significantly increased with the drop angle, with the maximum bruise volume being 3583.26 mm^3 and 3062.19 mm^3, respectively. The bruise volume with PVC as the material also increased with the drop angle, but the overall trend is lower than that of steel and wood, with the smallest bruise volume of 1392.41 mm^3 at a drop angle of 30° and the largest bruise volume of 2739.37 mm^3 at a drop angle of 90°. The simulation results indicated that the bruise volume caused to the American ginseng main root by different materials, from largest to smallest, was steel > wood > PVC > rubber. Rubber materials are more effective in reducing bruising to the American ginseng main root. For example, at a drop angle of 30°, the bruise resistance index of the American ginseng main root was 59.14 mJ mm^{-3} with rubber and 40.89 mJ mm^{-3} with wood. At a drop angle of 90°, the bruise resistance index was 45.72 mJ mm^{-3} with rubber and 35.38 mJ mm^{-3} with wood. At the same drop angle, the American ginseng main root bruise resistance index was higher with rubber as the contact material than with the other three materials (Figure 10F).

The bruise resistance of the American ginseng main root gradually decreased as the drop angle increased. As the drop angle increased, the contact surface with the material expanded. At larger drop angles, the central tissue of the American ginseng main root endured greater contact forces. According to the pattern where stress is maximized close to the contact surface and progressively diffuses interior, the American ginseng main root exhibited characteristics of localized bruising.

3.7 Discussion

In most finite element simulation studies, the multilayer structural properties of rhizome crops are often overlooked. Typically, agricultural products are simplified to a unitary structure (Caglayan et al., 2018; Celik, 2017). This simplification is suitable for some crops with analogous overall mechanical features, but for American ginseng, it could affect the accuracy of the results. Studies have shown that the multilayer structure of rhizome crops significantly impacts their physical and mechanical properties. For example, Stopa et al. (2019) found in their study on carrots that different tissue layers exhibited distinct mechanical behaviors when subjected to force. Li et al. (2013) also pointed out that overlooking the multilayer structure of crops could lead to inaccurate estimates of their fracture and deformation behaviors. This is especially true for complex rhizome crops like American ginseng, which have varying physical characteristics within their internal structures. This study applied a reverse engineering method to establish constitutive model of the American ginseng main root besides conducted precise physical parameter measurements. This approach not only improved the accuracy of the simulations but also provided a reference for future research. The American ginseng main root was divided into the cortex and cambium for separate mechanical property testing. This method partially overcame the limitations of previous studies that simplified rhizome crops to a single tissue, more accurately reflecting the actual conditions of collision bruising.

Since some studies involve practical calculation methods as well as the validation and comparison of models in simulations (Stropek and Gołacki, 2020), selecting an appropriate analysis method for model validation is crucial. Currently, many studies primarily use finite element simulation to evaluate the bruise extent of agricultural products (Ahmadi et al., 2016a), indicating that the FEM is a reliable and effective means for evaluating collision bruising in agricultural products. This study, based on finite element simulation and high-speed imaging technology, validated the drop velocity and bruise area of the American ginseng main root under different collision conditions, achieving the highest velocity error of 3.8 %. This result is consistent with the falling process and

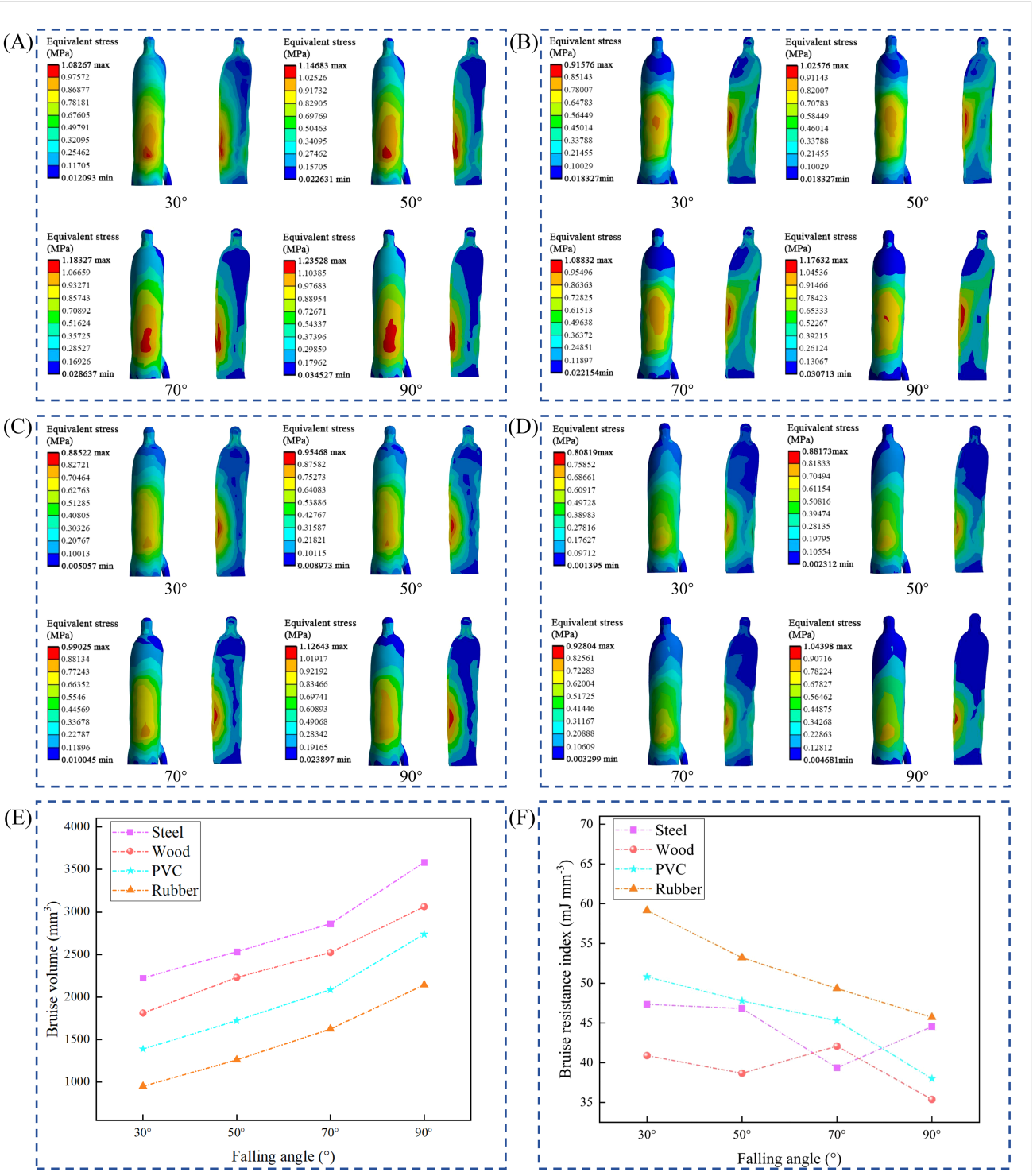


FIGURE 10
Bruise resistance characteristics of American ginseng main root. **(A–D)** Equivalent stress distribution cloud maps of the American ginseng main root with contact materials of steel, wood, PVC, and rubber at drop angles of 30°, 50°, 70°, and 90°; **(E, F)** Bruise resistance index and bruise volume of different contact materials as the drop angle varies.

bruising observed during the pendulum test, indicating that the intrinsic model established in this study has a high degree of accuracy. This model calibration method provides a reliable verification approach for subsequent studies. Future studies could consider a more diverse range of fruit and vegetable varieties, as well as more complex collision scenarios, including multiple collisions, multi-object collisions, and collisions with different composite material surfaces. Additionally, new parameters, such as humidity, temperature, and maturity, could be introduced to further enhance model adaptability and predictive capability.

The observed crossover of bruise resistance index (BRI) curves for different contact materials at specific drop angles likely arises from inherent differences in the mechanical properties and energy transfer mechanisms of the materials. Notably, a nonlinear relationship among hardness, elastic modulus, and energy absorption characteristics exists (Zhao et al., 2019a). Hard materials typically exhibit higher elastic moduli, yet they tend to absorb less energy during impact, leading to a rapid transfer of the majority of energy to the impacted object. Conversely, softer materials, characterized by a lower elastic modulus, often provide superior energy absorption capabilities, thereby mitigating damage under certain conditions (Ahmadi et al., 2016a). Consequently, at specific drop angles, hard materials may manifest a higher BRI, whereas at other angles, the enhanced energy absorption of softer materials may yield a higher BRI. This phenomenon underscores the necessity of considering multiple mechanical properties in the design of impact-mitigation systems and the evaluation of collision safety, rather than relying solely on a single parameter. This study analyzed the dynamic mechanical response of the American ginseng main root under different drop angles and collision contact conditions, resulting in variation curves of the maximum contact force, maximum equivalent stress, and maximum internal energy during the collision process (Figures 8, 9). Analysis of the forces during the American ginseng main root collision showed that the maximum contact force and maximum equivalent stress were primarily concentrated near the contact surface and gradually diffused inward. This stress distribution pattern determined the shape and extent of the bruised area. Although the problem of bruising from collisions of agricultural products can be effectively analyzed using sophisticated near-infrared spectroscopy and computed tomography technologies, these methods require a high technological threshold and cost. When carrying out spot checks on agricultural product specimens or assessing the functionality of equipment, using bruise volume and surface area to calculate the extent of bruising is unquestionably the most straightforward and effective method. In this research, the bruise volume of the American ginseng main root under different collision conditions was compared. The results indicated that as the drop angle enlarged, the bruise volume of the American ginseng main root also increased, leading to the higher risk of bruising. Finite element simulation results revealed that the bruise volume of the American ginseng main root varied significantly when colliding at different drop angles and with different contact materials such as steel, wood, PVC, and rubber. Under conditions of larger drop angles and harder contact materials (such as steel and wood), the American ginseng main root exhibited greater bruising, with maximum bruise volumes of 3583.26 mm^3 and 3062.19 mm^3, respectively. Although the developed model and experimental setup provided valuable insights, several limitations must be acknowledged. Firstly, the constitutive model assumed linear elasticity for the cambium and a simplified elastic–plastic behavior for the cortex, whereas actual plant tissues exhibit viscoelasticity and are strain-rate sensitive. Thus, under very high strain rates or over longer durations post-impact, the material response may diverge from the predictions of the current model. Incorporating viscoelastic properties in future models would improve accuracy across a range of loading conditions. Second, the model and experiments focused solely on single-impact events. In practice, during harvest or transport, American ginseng main roots may undergo multiple successive impacts or continuous vibrations; however, the current model does not account for cumulative bruise from repeated impacts. Third, the pendulum test simulates a free-swing impact scenario that may not capture all aspects of actual harvester impacts, such as interactions with soil or adjacent roots; extending validation to field conditions would further strengthen the conclusions. Fourth, individual variability in American ginseng root morphology and moisture content was not explicitly modeled, as average properties were used —although actual bruising may vary significantly among American ginseng main roots. Finally, high-speed photography in this study was primarily used for velocity measurement and bruise imaging; more advanced imaging techniques, such as X-ray CT scanning for internal bruises, could provide a more detailed validation of internal damage patterns. Addressing these limitations in future work—for example, by incorporating viscoelastic FEM models, simulating repeated impacts, and testing a broader range of crop samples—will enhance both the robustness and applicability of the findings.

This study primarily focused on the American ginseng main root, but the methods could also be applied to other rhizome crops. The application of these methods could be considered in subsequent studies to enhance the understanding and assessment of collision bruising in various rhizome crops. This study provides a theoretical basis for designing scientifically sound American ginseng harvesting machinery, helping to reduce bruising during harvesting and improve both yield and quality. It also serves as a reference for designing mechanized harvesting equipment for other rhizome crops. Future research can build on this foundation by further optimizing model parameters and exploring bruising mechanisms under a wider range of conditions, providing more comprehensive technical support for the fields of agricultural machinery and food engineering.

4 Conclusion

1. The bruise volume increased with the drop angle, and the maximum bruise volume was 3583.26 mm^3 (steel) and 3062.19 mm^3 (wood), respectively. At a drop angle of 30°, the bruise resistance indices of the American ginseng main root were 59.14 mJ mm^{-3} (rubber) and 40.89 mJ mm^{-3} (wood). At a drop angle of 90°, the bruise resistance indices of American ginseng main root were 45.72 mJ mm^{-3} (rubber) and 35.38 mJ mm^{-3} (wood).
2. The maximum equivalent stress, maximum contact force, and maximum internal energy during the American ginseng main root collision are positively correlated with the drop angle. After the American ginseng main root rebounds, residual stress and energy remain within it.
3. The bruise volume of the American ginseng main root on different contact materials increased with the drop angle,

with the order of bruise volume being: steel > wood > PVC > rubber. At the same drop angle, the bruise resistance index of rubber was higher than that of the other three materials.

4. This study established a predictive model for the bruise resistance index of the American ginseng main root using finite element simulation technology, providing a scientific basis for the design and optimization of harvesting machinery and its key components, with the objective of effectively controlling the American ginseng bruising problem.

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

PJ: Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. GZ: Data curation, Investigation, Methodology, Software, Writing – review & editing. JC: Data curation, Visualization, Writing – original draft. DW: Conceptualization, Formal analysis, Methodology, Writing – original draft. HT: Conceptualization, Formal analysis, Funding acquisition, Resources, Supervision, Writing – original draft. JW: Supervision, Data curation, Visualization, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by the National Natural Science Foundation of China (NSFC), [grant number: 52475247] and the Key Research and Development Program of Heilongjiang Province of China [grant number: 2023ZX05B03].

Conflict of interest

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

Generative AI statement

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

Publisher's note

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

References

Abbaszadeh, R., Rajabipour, A., Sadrnia, H., Mahjoob, M. J., Delshad, M., and Ahmadi, H. (2014). Application of modal analysis to the watermelon through finite element modeling for use in ripeness assessment. *J. Food Eng.* 127, 80–84. doi: 10.1016/j.jfoodeng.2013.11.020

Ahmadi, E., Barikloo, H., and Kashfi, M. (2016a). Viscoelastic finite element analysis of the dynamic behavior of apple under impact loading with regard to its different layers. *Comput. Electron. Agric.* 121, 1–11. doi: 10.1016/j.compag.2015.11.017

Caglayan, N., Oral, O., Celik, H. K., Cinar, R., Rodrigues, L. C. D. A., Rennie, A. E. W., et al. (2018). Determination of time dependent stress distribution on a potato tuber during drop case. *J. Food Process Eng.* 41(7). doi: 10.1111/jfpe.12869

Celik, H. K. (2017). Determination of bruise susceptibility of pears (Ankara variety) to impact load by means of FEM-based explicit dynamics simulation. *Postharvest Biol. Technol.* 128, 83–97. doi: 10.1016/j.postharvbio.2017.01.015

Dimino, I., Diodati, G., Di Caprio, F., Ciminello, M., Menichino, A., Inverno, M., et al. (2022). Numerical and experimental studies of free-fall drop impact tests using strain gauge, piezoceramic, and fiber optic sensors. *Appl. Mechanics* 3, 313–338. doi: 10.3390/applmech3010020

Gołacki, K., Bobin, G., and Stropek, Z. (2009). Bruise resistance of apples (melrose variety). *TEKA Kom. Mot. Roln.-OLPAN* 9, 40–47.

Guan, X., Li, T., and Zhou, F. (2023). Determination of bruise susceptibility of fresh corn to impact load by means of finite element method simulation. *Postharvest Biol. Technol.* 198, 112227. doi: 10.1016/j.postharvbio.2022.112227

He, H.-J., Wu, D., and Sun, D.-W. (2015). Nondestructive spectroscopic and imaging techniques for quality evaluation and assessment of fish and fish products. *Crit. Rev. Food Sci. Nutr.* 55, 864–886. doi: 10.1080/10408398.2012.746638

Herold, B., Geyer, M., and Studman, C. J. (2001). Fruit contact pressure distributions — equipment. *Comput. Electron. Agric.* 32, 167–179. doi: 10.1016/S0168-1699(01)00160-0

Holt, J. E., and Schoorl, D. (1982). Strawberry bruising and energy dissipation. *J. Texture Stud.* 13, 349–357. doi: 10.1111/j.1745-4603.1982.tb00888.x

Jiménez-Jiménez, F., Castro-García, S., Blanco-Roldán, G. L., Ferguson, L., Rosa, U. A., and Gil-Ribes, J. A. (2013). Tabl e olive cultivar susceptibility to impact bruising. *Postharvest Biol. Technol.* 86, 100–106. doi: 10.1016/j.postharvbio.2013.06.024

Kabas, O., Celik, H. K., Ozmerzi, A., and Akinci, İ. (2008). Drop test simulation of a sample tomato with finite element method. *J. Sci. Food Agric.* 88, 1537–1541. doi: 10.1002/jsfa.3246

Keresztes, J. C., Diels, E., Goodarzi, M., Nguyen-Do-Trong, N., Goos, P., Nicolai, B., et al. (2017). Glare based apple sorting and iterative algorithm for bruise region detection using shortwave infrared hyperspectral imaging. *Postharvest Biol. Technol.* 130, 103–115. doi: 10.1016/j.postharvbio.2017.04.005

Komarnicki, P., Stopa, R., Kuta, Ł., and Szyjewicz, D. (2017). Determination of apple bruise resistance based on the surface pressure and contact area measurements under impact loads. *Comput. Electron. Agric.* 142, 155–164. doi: 10.1016/j.compag.2017.08.028

Lewis, R., Yoxall, A., Marshall, M. B., and Canty, L. A. (2008). Characterising pressure and bruising in apple fruit. *Wear* 264, 37–46. doi: 10.1016/j.wear.2007.01.038

Li, Z., Li, P., Yang, H., and Liu, J. (2013). Internal mechanical damage prediction in tomato compression using multiscale finite element models. *J. Food Eng.* 116, 639–647. doi: 10.1016/j.jfoodeng.2013.01.016

Li, Z., and Thomas, C. (2014). Quantitative evaluation of mechanical damage to fresh fruits. *Trends Food Sci. Technol.* 35, 138–150. doi: 10.1016/j.tifs.2013.12.001

Liu, Z., Miao, S., Liang, M., Yang, P., Zhao, Z., and Chang, N. (2024). Characteristics and association mechanism of energy dissipation and crack development in siltstone under triaxial cyclic loading and unloading. *Rock Mechanics Rock Eng* 57(12), 1–27. doi: 10.1007/s00603-024-04098-9

Miraei Ashtiani, S.-H., Sadrnia, H., Mohammadinezhad, H., Aghkhani, M. H., Khojastehpour, M., and Abbaspour-Fard, M. H. (2019). FEM-based simulation of the mechanical behavior of grapefruit under compressive loading. *Scientia Hortic.* 245, 39–46. doi: 10.1016/j.scienta.2018.10.006

Rashvand, M., Altieri, G., Genovese, F., Li, Z., and Di Renzo, G. C. (2022). Numerical simulation as a tool for predicting mechanical damage in fresh fruit. *Postharvest Biol. Technol.* 187, 111875. doi: 10.1016/j.postharvbio.2022.111875

Stopa, R., Komarnicki, P., Kuta, Ł., Szyjewicz, D., and Słupska, M. (2019). Modeling with the finite element method the influence of shaped elements of loading components on the surface pressure distribution of carrot roots. *Comput. Electron. Agric.* 167, 105046. doi: 10.1016/j.compag.2019.105046

Stropek, Z., and Gołacki, K. (2020). Bruise susceptibility and energy dissipation analysis in pears under impact loading conditions. *Postharvest Biol. Technol.* 163, 111120. doi: 10.1016/j.postharvbio.2020.111120

Van Zeebroeck, M., Van linden, V., Ramon, H., De Baerdemaeker, J., Nicolaï, B. M., and Tijskens, E. (2007). Impact damage of apples during transport and handling. *Postharvest Biol. Technol.* 45, 157–167. doi: 10.1016/j.postharvbio.2007.01.015

Wu, J., Guo, K., Ge, Y., and Wang, Y. (2012). Contact pressure distribution characteristics of Korla pear fruit at moment of drop impact. *Nongye Gongcheng Xuebao/Transactions Chin. Soc. Agric. Eng.* 28, 250–254. doi: 10.3969/j.issn.1002-6819.2012.01.044

Wu, J., Li, F., Ge, Y., and Hu, R. (2013). Measurement of contact pressure of Korla pear under compression and bruising predication using finite element analysis. *Nongye Gongcheng Xuebao/Transactions Chin. Soc. Agric. Eng.* 29, 261–266. doi: 10.3969/j.issn.1002-6819.2013.06.032

Xu, C., Wang, D., Xu, F., Tang, H., Zhao, J., and Wang, J. (2023). Prediction of bruising susceptibility in white radish (Raphanus sativus L.) using FEA-RSM technique. *Postharvest Biol. Technol.* 206, 112565. doi: 10.1016/j.postharvbio.2023.112565

Yan, D., Qiu, S., Liu, W., Luo, L., and Liu, Y. (2024). Finite element analysis of the dynamic behaviour of tabl e grapes during mechanical harvesting. *Biosyst. Eng.* 237, 118–127. doi: 10.1016/j.biosystemseng.2023.12.003

Yousefi, S., Farsi, H., and Kheiralipour, K. (2016). Drop test of pear fruit: Experimental measurement and finite element modelling. *Biosyst. Eng.* 147, 17–25. doi: 10.1016/j.biosystemseng.2016.03.004

Zhao, J., Sugirbay, A., Chen, Y., Zhang, S., Liu, F., Bu, L., et al. (2019a). FEM explicit dynamics simulation and NIR hyperspectral reflectance imaging for determination of impact bruises of Lycium barbarum L. *Postharvest Biol. Technol.* 155, 102–110. doi: 10.1016/j.postharvbio.2019.05.024

Zhou, J., He, L., Karkee, M., and Zhang, Q. (2016). Effect of catching surface and tilt angle on bruise damage of sweet cherry due to mechanical impact. *Comput. Electron. Agric.* 121, 282–289. doi: 10.1016/j.compag.2016.01.004

Zulkifli, N., Hashim, N., Harith, H. H., and Mohamad Shukery, M. F. (2020). Finite element modelling for fruit stress analysis - A review. *Trends Food Sci. Technol.* 97, 29–37. doi: 10.1016/j.tifs.2019.12.029

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 June 2025
DOI 10.3389/fpls.2025.1575796

Check for updates

OPEN ACCESS

EDITED BY
Jun Ni,
Nanjing Agricultural University, China

REVIEWED BY
Prasath N.,
SRM University, India
Dekera Kwaghtyo,
Federal University Lafia, Nigeria
Eliazer M.,
SRM Institute of Science and Technology, India

*CORRESPONDENCE
Meijin Li
jlptr1@163.com

RECEIVED 13 February 2025
ACCEPTED 06 May 2025
PUBLISHED 09 June 2025

CITATION
He T, Li M and Jin D (2025) Deep learning-based time series prediction for precision field crop protection.
Front. Plant Sci. 16:1575796.
doi: 10.3389/fpls.2025.1575796

COPYRIGHT
© 2025 He, Li and Jin. 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.

Deep learning-based time series prediction for precision field crop protection

Tao He[1], Meijin Li[2]* and Dong Jin[3]

[1]School of Intelligent Manufacturing, Wenzhou Polytechnic, Wenzhou, China, [2]School of Intelligent Manufacturing and Electronic Engineering, Wenzhou University of Technology, Wenzhou, China, [3]Department of Art and Design, Taiyuan University, Shanxi, China

Introduction: Precision agriculture is revolutionizing modern farming by integrating data-driven methodologies to enhance crop productivity while promoting sustainability. Traditional time series models struggle with complex agricultural data due to heterogeneity, high dimensionality, and strong spatial-temporal dependencies. These limitations hinder their ability to provide actionable insights for resource optimization and environmental protection.

Methods: To tackle these difficulties, this research puts forward a new deep-learning-based architecture for time-series prediction that is customized for precise field crop protection. At its core, our Spatially-Aware Data Fusion Network (SADF-Net) integrates multi-modal data sources, such as satellite imagery, IoT sensor readings, and meteorological forecasts, into a unified predictive model. By combining convolutional layers for spatial feature extraction, recurrent neural networks for temporal modeling, and attention mechanisms for data fusion, SADF-Net captures intricate spatial-temporal dependencies while ensuring robustness to noisy and incomplete data. We introduce the Resource-Aware Adaptive Decision Algorithm (RAADA), which leverages reinforcement learning to translate SADF-Net's predictions into optimized strategies for resource allocation, such as irrigation scheduling and pest control. RAADA dynamically adapts decisions based on real-time field responses, ensuring efficiency and sustainability.

Results: The experimental findings obtained from large-scale agricultural datasets show that our framework far exceeds the existing most advanced methods in terms of the accuracy of yield prediction, resource optimization, and environmental impact mitigation.

Discussion: This research offers a transformative solution for precision agriculture, aligning with the pressing need for advanced tools in sustainable crop management.

KEYWORDS

precision agriculture, time series prediction, deep learning, resource optimization, spatial-temporal modeling

1 Introduction

The increasing demand for sustainable agricultural practices has led to an urgent need for accurate predictions in crop protection, particularly in the face of climate change, pest outbreaks, and resource limitations (Angelopoulos et al., 2023). Time series prediction has emerged as a critical tool in precision agriculture, enabling farmers to anticipate and mitigate risks such as disease outbreaks and pest infestations (Shen and Kwok, 2023). Traditional approaches to time series prediction, while effective in certain scenarios, are often limited by their inability to process large-scale, non-linear, and complex data derived from precision agriculture systems, which include remote sensing, weather monitoring, and soil sensors (Zhou et al., 2020). This limitation hinders both accurate crop management and the potential of adaptive decision-making frameworks to enhance yield and sustainability (Li et al., 2023). The evolution from traditional, symbol-based AI approaches to data-driven machine learning and, more recently, to deep learning and pre-trained models underscores a growing ability to address these challenges with increasing precision and scalability (Yin et al., 2023).

Early methods for time series prediction in crop protection relied on symbolic AI techniques and rule-based systems that modeled crop health and environmental factors using predefined rules and expert knowledge (Yu et al., 2023). These approaches used domain knowledge to simulate plant-pathogen interactions or to estimate pest behavior, often through mechanistic models such as epidemiological equations or statistical regression (Durairaj and Mohan, 2022). For example, models like the degree-day method or rule-based systems were used to predict pest emergence or disease onset (Chandra et al., 2021). While these methods were interpretable and grounded in agricultural science, they were constrained by their dependence on accurate and comprehensive domain knowledge (Fan et al., 2021). they struggled with adapting to dynamic and highly variable field conditions, as they lacked mechanisms to incorporate real-time data or learn from observed patterns (Hou et al., 2022). As a result, these symbolic approaches were not only labor-intensive but also lacked generalizability, making them unsuitable for large-scale, high-resolution precision agriculture systems.

The transition to data-driven methods brought significant advancements in time series prediction for crop protection by leveraging machine learning algorithms (Lindemann et al., 2021). Techniques such as support vector machines (SVMs), random forests, and gradient-boosted trees became popular due to their ability to uncover patterns from data without requiring explicit domain knowledge (Dudukcu et al., 2022). In precision agriculture, these methods were used to process sensor data, weather records, and imagery to predict pest infestations or crop diseases (Amalou et al., 2022). For instance, SVMs were applied to classify crop health based on hyperspectral data, while random forests were used to identify key environmental factors contributing to pest outbreaks (Xiao et al., 2021). Despite these improvements, data-driven methods still faced limitations, particularly in handling temporal dependencies and long-term sequences, as they relied on feature engineering and lacked the ability to capture spatial and temporal correlations effectively (Zheng and Chen, 2021). these models were computationally expensive for large datasets, limiting their scalability in real-time agricultural applications.

The emergence of deep learning and pre-trained models has revolutionized time-series forecasting, particularly in complex domains like precision agriculture (Wang et al., 2021b). Recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and convolutional neural networks (CNNs) have shown remarkable capabilities in learning temporal and spatial patterns from large datasets (Xu et al., 2020). In the context of precision crop protection, LSTMs have been used to model pest population dynamics, while CNNs have been applied to satellite imagery for disease detection (Karevan and Suykens, 2020). More recently, transformer-based models and pre-trained architectures have set new benchmarks in accuracy and adaptability (Altan and Karasu, 2021). These models excel in capturing multi-scale dependencies and integrating heterogeneous data sources, such as weather forecasts, soil health, and remote sensing data (Wen et al., 2021). transfer learning allows pre-trained models to generalize across different crops and regions, reducing the need for extensive labeled datasets. these methods are computationally intensive and require significant expertise for implementation, which can pose challenges for widespread adoption in resource-constrained settings.

To address the limitations of previous methods, we propose an innovative deep learning framework specifically designed for precision crop protection based on accurate time series forecasting. By integrating domain knowledge with advanced neural architectures, we aim to overcome the challenges of interpretability, scalability, and adaptability. The proposed method leverages pre-trained models to incorporate multi-modal data and uses attention mechanisms to focus on critical temporal patterns, thereby enabling precise predictions even under uncertain conditions. Our approach prioritizes computational efficiency, ensuring that it can be deployed in real-time scenarios and resource-limited environments.

- The integration of attention mechanisms and pre-trained architectures allows our model to focus on critical temporal patterns, improving interpretability and robustness in predictions.
- The framework supports multi-modal data inputs and generalizes across different crops and regions, making it highly adaptable and efficient for diverse agricultural settings.
- Extensive testing on benchmark datasets demonstrates significant improvements in prediction accuracy and computational efficiency compared to existing deep learning models.

2 Related work

2.1 Deep learning for time series forecasting

Deep learning techniques have emerged as powerful tools for time series forecasting, leveraging their ability to model complex

temporal dependencies and capture non-linear patterns in data (Wang et al., 2021a). Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), are widely used for such tasks due to their capability to address vanishing gradient issues and effectively capture long-term dependencies (Morid et al., 2021). In the agricultural domain, these architectures have been applied to predict various environmental factors, including temperature, humidity, and precipitation, which are critical for field crop protection (Widiputra et al., 2021). Attention mechanisms, integrated into sequence models, have further enhanced performance by allowing the model to prioritize the most significant time steps, thus improving forecasting accuracy (Moskolaï et al., 2021). Recent advancements also include the application of Transformer architectures to time series prediction. Transformers, initially designed for natural language processing, have been successfully adapted for time series due to their scalability and capability to capture long-term dependencies without the constraints of sequential processing (Ni et al., 2018). Such models have been employed to predict pest infestation trends, soil moisture levels, and crop yields, demonstrating their potential in precision agriculture (Yu et al., 2025). hybrid approaches that combine deep learning with statistical methods, such as ARIMA or wavelet transforms, have been explored to enhance predictive performance by integrating domain knowledge with data-driven learning.

2.2 Precision agriculture and data-driven methods

Precision agriculture relies heavily on data-driven approaches to optimize resource usage and improve crop productivity (Yang and Wang, 2021). With the advent of Internet of Things (IoT) devices, remote sensing technologies, and UAV-based imaging systems, vast amounts of spatiotemporal data have become available for analysis (Ruan et al., 2021). Machine learning techniques, particularly deep learning, have played a critical role in processing and analyzing this data (Kim and King, 2020). Convolutional Neural Networks (CNNs), for example, have been utilized to analyze aerial imagery and satellite data to monitor crop health, detect weeds, and identify pest infestations. Combined with time series data, such as weather patterns and soil conditions, these methods enable a more holistic understanding of field dynamics (Ni et al., 2017). advancements in sensor technology have enabled real-time monitoring of environmental factors, generating high-resolution time series data that can be fed into predictive models to anticipate threats such as fungal diseases or pest outbreaks. Integrating these predictions into precision crop protection systems enables timely and targeted interventions, reducing pesticide use and environmental impact (Chen et al., 2024). The integration of Geographic Information Systems (GIS) with deep learning models has also enhanced spatial forecasting capabilities, allowing for the creation of site-specific management zones. Research in this area has focused on developing robust models that can generalize across diverse agricultural conditions, addressing challenges such as data sparsity, noise, and the need for domain-specific customization.

2.3 Sustainability in crop protection

Sustainability has become a central theme in modern agricultural practices, emphasizing the need for reduced chemical usage, minimized environmental impact, and improved resource efficiency (Kang et al., 2020). Deep learning-based models for time series forecasting play a crucial role in achieving these goals by enabling precise and proactive interventions (Wu et al., 2020). By predicting pest and disease outbreaks, irrigation needs, and nutrient deficiencies, these models allow for targeted applications of pesticides, water, and fertilizers, reducing wastage and runoff (Hu et al., 2020). Studies have highlighted the potential of integrating crop growth models with time series prediction frameworks to optimize field management strategies while maintaining ecological balance. the use of multispectral and hyperspectral imaging, combined with temporal deep learning methods, has enabled early detection of crop stress and disease, further contributing to sustainable practices. Another key focus is the development of interpretable deep learning models that provide actionable insights to farmers and agronomists (Ni et al., 2016). Techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Modelagnostic Explanations) have been applied to enhance the transparency of predictions, fostering trust in AI-driven systems. research has explored the incorporation of renewable energy-powered sensors and edge computing devices to support low-cost and sustainable deployment of predictive systems in remote and resource-constrained regions (Yan et al., 2024). These advancements align with global initiatives to promote sustainable agriculture and ensure food security in the face of climate change and population growth.

3 Method

3.1 Overview

Precision agriculture has emerged as a transformative approach to modern farming, leveraging data-driven methodologies to enhance crop productivity, optimize resource utilization, and reduce environmental impacts. This subsection introduces the proposed methodology to address specific challenges within precision agriculture. We present a detailed outline of the subsequent subsections, which collectively define the core contributions of this research. In this work, we aim to tackle the problem of efficiently integrating multi-modal data sources into a cohesive decision-making framework for precision agriculture. Our method emphasizes scalability and robustness to heterogeneous data, which are critical in real-world agricultural scenarios.

The first component of our framework, outlined in Section 3.2, provides the formalization of the problem domain. Here, we

establish the mathematical and computational foundations, introducing key notations, data representations, and the modeling of spatial and temporal dependencies inherent in agricultural processes. This section also highlights the challenges posed by noisy and incomplete data, which are common in field conditions, and sets the stage for the subsequent methodological innovations. Building upon this foundation, Section 3.3 presents our novel model, termed Spatially-Aware Data Fusion Network (SADF-Net). SADF-Net is designed to integrate diverse data sources into a unified predictive framework. The architecture employs advanced deep learning techniques to capture spatial correlations across fields and temporal dynamics in crop growth. This model is tailored to extract actionable insights from complex, high-dimensional agricultural datasets. in Section 3.4, we describe a new optimization strategy, referred to as the Resource-Aware Adaptive Decision Algorithm (RAADA). This strategy focuses on deploying the predictions of SADF-Net to enable precise and efficient interventions, such as irrigation scheduling, fertilization optimization, and pest control. The algorithm incorporates domain-specific constraints and leverages reinforcement learning to iteratively refine decisions based on observed outcomes.

3.2 Preliminaries

In this part, we define the problem formally of decision-making in precision agriculture and establish the mathematical framework underpinning our approach. This includes the definition of key variables, constraints, and the computational challenges associated with integrating multi-modal agricultural data.

Precision agriculture involves optimizing resource allocation and improving crop productivity by leveraging diverse datasets such as satellite imagery, sensor measurements, weather forecasts, and soil profiles. Let $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ denote the set of agricultural fields under consideration, where $F_i$represents the i-th field characterized by spatial and temporal features. Each field F_i is further subdivided into grid cells, indexed by $(x, y) \in \mathcal{G}_i$, representing a spatial discretization.

We define the state of the agricultural system over a temporal horizon. $\mathcal{T} = \{t_1, t_2, \ldots, t_T\}$ as a collection of feature maps (Equation 1):

$$\mathcal{S}_{i,t} = \left\{\mathbf{d}_{i,t}, \mathbf{e}_{i,t}, \mathbf{r}_{i,t}, \mathbf{h}_{i,t}\right\}, \quad \forall\, i \in \{1, 2, \ldots, m\}, \quad t \in \mathcal{T}, \tag{1}$$

where: $\mathbf{d}_{i,t} \in \mathbb{R}^{n_d}$ represents crop-specific data, including growth stage, health, and phenotypic characteristics for field F_i at time t. $\mathbf{e}_{i,t} \in \mathbb{R}^{n_e}$ captures environmental data, such as temperature, humidity, and precipitation, obtained from external meteorological sources. $\mathbf{r}_{i,t} \in \mathbb{R}^{n_r}$ represents resource-related variables, including irrigation, fertilization, and pest control efforts. $\mathbf{h}_{i,t} \in \mathbb{R}^{n_h}$ denotes historical data for the field, encapsulating past observations of yield, resource usage, and interventions.

The system's evolution is influenced by various external and internal factors, which we encode as a dynamical system (Equation 2):

$$\mathcal{S}_{i,t+1} = \Phi(\mathcal{S}_{i,t}, \mathbf{u}_{i,t}) + \varepsilon_{i,t}, \tag{2}$$

where $\Phi(\cdot)$ is a nonlinear transition function modeling the temporal evolution of the system, $u_{i,t}$represents the control inputs, and $\varepsilon_{i,t}$is the noise term accounting for uncertainties and measurement errors.

The overarching goal is to optimize a set of control decisions. $\mathrm{U} = \left\{u_{i,t}\right\}_{i,t}$over $\mathcal{T}$ to maximize crop productivity while minimizing resource usage and environmental impact. This can be mathematically formulated as a multi-objective optimization problem (Equation 3):

$$\max_{\mathcal{U}} \quad J(\mathcal{U}) = \sum_{i=1}^{m}\sum_{t=1}^{T}[\alpha_1 \mathcal{Y}_{i,t} - \alpha_2 \mathcal{C}_{i,t} - \alpha_3 \mathcal{E}_{i,t}], \tag{3}$$

where: $\mathcal{Y}_{i,t}$ is the predicted yield for field F_i at time t, $\mathcal{C}_{i,t}$ represents the cost associated with resources such as water, fertilizers, and pesticides, $\mathcal{E}_{i,t}$ captures environmental penalties, such as nutrient leaching or greenhouse gas emissions, α_1, α_2, and α_3 are weights balancing the trade-offs between productivity, cost, and sustainability.

The optimization is subject to domain-specific constraints: Resource Budget (Equation 4):

$$\sum_{t=1}^{T}\sum_{i=1}^{m} u_{i,t} \le \mathcal{B}, \tag{4}$$

where $\mathcal{B}$ is the total available resource budget. Crop-Specific Requirements (Equation 5):

$$g(\mathbf{d}_{i,t}, \mathbf{u}_{i,t}) \ge 0, \quad \forall\, i, t, \tag{5}$$

ensuring that decisions align with agronomic best practices. Environmental Limits (Equation 6):

$$\mathcal{E}_{i,t} \le \mathcal{E}_{\max}, \quad \forall\, i, t, \tag{6}$$

imposing upper bounds on environmental impacts.

3.3 Spatially-Aware Data Fusion Network

In this subsection, we propose a novel predictive framework termed Spatially-Aware Data Fusion Network (SADF-Net). The SADF-Net model is designed to integrate multi-modal agricultural data, capture spatial-temporal dependencies, and generate accurate predictions for field-specific variables, such as yield, resource requirements, and environmental impacts (As shown in Figure 1).

3.3.1 Spatial feature extraction

The architecture of SADF-Net integrates convolutional neural networks (CNNs) for spatial feature extraction, recurrent neural networks (RNNs) for temporal modeling, and an attention mechanism for multi-modal data fusion, ensuring a comprehensive learning framework for spatiotemporal prediction tasks.

Each field F_i at time step t is represented by a multi-channel feature tensor $\mathrm{X}_{i,t} \in \mathbb{R}^{H \times W \times C}$, where H and W represent the spatial dimensions of the grid cells, ensuring the representation of spatial layout and resolution. The number of input channels, C, includes

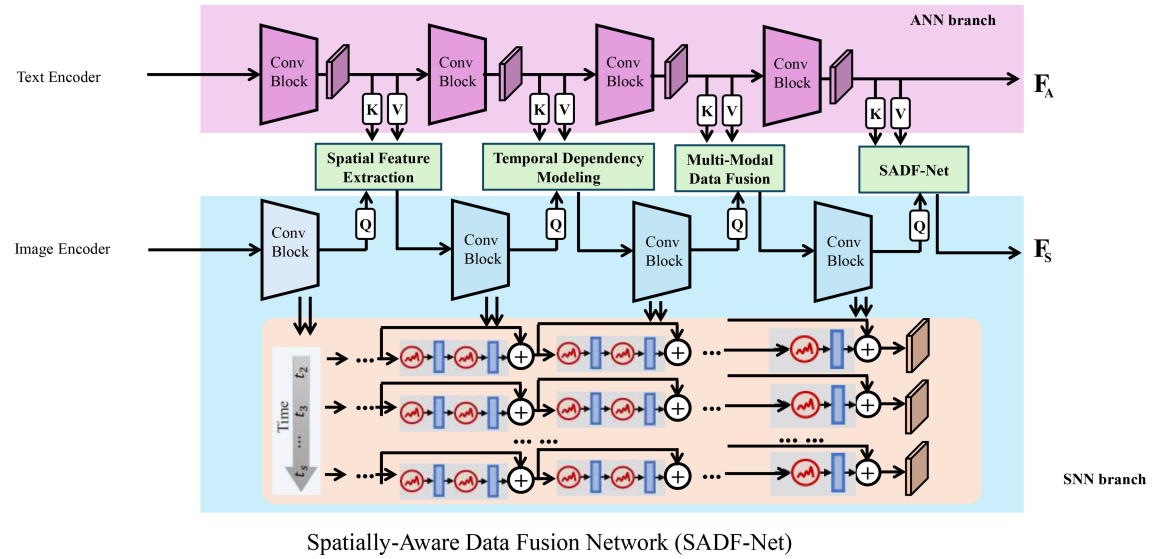


FIGURE 1
The Spatially-Aware Data Fusion Network (SADF-Net) integrates multi-modal agricultural data. It combines convolutional neural networks (CNNs) for spatial feature extraction, gated recurrent units (GRUs) for temporal dependency modeling, and an attention-based fusion mechanism. The model processes text and image inputs through separate encoder branches—an artificial neural network (ANN) branch for textual data and a spiking neural network (SNN) branch for image data—before fusing them for improved predictive accuracy in agricultural applications.

various geospatial and environmental data sources such as satellite-derived vegetation indices, soil moisture, precipitation levels, temperature variations, and past resource usage records.

To effectively extract spatial correlations among grid cells, a convolutional layer is applied to $\mathbf{X}_{i,t}$ (Equation 7):

$$\mathbf{Z}_{i,t}^{(1)} = \sigma(\mathbf{W}_{\text{conv}} * \mathbf{X}_{i,t} + \mathbf{b}_{\text{conv}}), \tag{7}$$

where $\mathbf{W}_{\text{conv}}$ and $\mathbf{b}_{\text{conv}}$ denote the trainable convolutional kernel weights and biases, respectively. The function $\sigma(\cdot)$ represents A nonlinear activation function, like ReLU or LeakyReLU, and $*$ represents the convolution operation applied over the spatial dimensions. This process generates spatially-aware feature maps $\mathbf{Z}_{i,t}^{(1)} \in \mathbb{R}^{H \times W \times C_1}$, where C_1 is the number of output feature channels, capturing hierarchical spatial dependencies.

To further enhance spatial feature representation and capture deeper-level patterns, additional convolutional layers with increasing receptive fields are applied (Equation 8):

$$\mathbf{Z}_{i,t}^{(l)} = \sigma(\mathbf{W}_{\text{conv}^{(l)}} * \mathbf{Z}_{i,t}^{(l-1)} + \mathbf{b}_{\text{conv}^{(l)}}), \quad l = 2, \ldots, L. \tag{8}$$

Each layer refines feature extraction by progressively capturing higher-level spatial relationships. The final spatial feature representation is denoted as (Equation 9):

$$\mathbf{Z}_{i,t} = \mathbf{Z}_{i,t}^{(L)} \in \mathbb{R}^{H \times W \times C_L}, \tag{9}$$

where C_L represents the final number of extracted spatial feature channels.

To introduce spatial invariance and reduce computational complexity, a max-pooling operation is applied (Equation 10):

$$\mathbf{P}_{i,t} = \text{MaxPool}(\mathbf{Z}_{i,t}), \tag{10}$$

where $\mathbf{P}_{i,t} \in \mathbb{R}^{H' \times W' \times C_L}$, and (H', W') denote the reduced spatial dimensions, controlled by the pooling kernel size.

For enhanced spatial understanding, a spatial self-attention mechanism is applied to highlight important regions within the feature map. The attention weights are computed as (Equation 11):

$$\alpha_{i,t} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d}}\right), \tag{11}$$

where $\mathbf{Q}$ and $\mathbf{K}$ are linear projections of $\mathbf{P}_{i,t}$, and d is the dimensionality scaling factor. The attended spatial features are then computed as (Equation 12):

$$\mathbf{P}_{i,t}^{\text{att}=\alpha_{i,t}V}, \tag{12}$$

where $\mathbf{V}$ is another linear projection of $\mathbf{P}_{i,t}$. This mechanism ensures that regions with higher relevance to the target task receive greater emphasis.

3.3.2 Temporal dependency modeling

To accurately capture the temporal dynamics of field conditions, we utilize a gated recurrent unit (GRU) network, which is a variant of recurrent neural networks (RNNs) designed to address the vanishing gradient problem and efficiently capture long-range dependencies in sequential data. The GRU updates its hidden state as follows (Equation 13):

$$\mathbf{h}_{i,t} = \text{GRU}(\mathbf{z}_{i,t}, \mathbf{h}_{i,t-1}), \tag{13}$$

where $\mathbf{z}_{i,t} = \text{Flatten}(\mathbf{Z}_{i,t}^{(1)})$ Denotes the compressed spatial feature vector obtained at a specific time step t, and $\mathbf{h}_{i,t} \in \mathbb{R}^{d_h}$

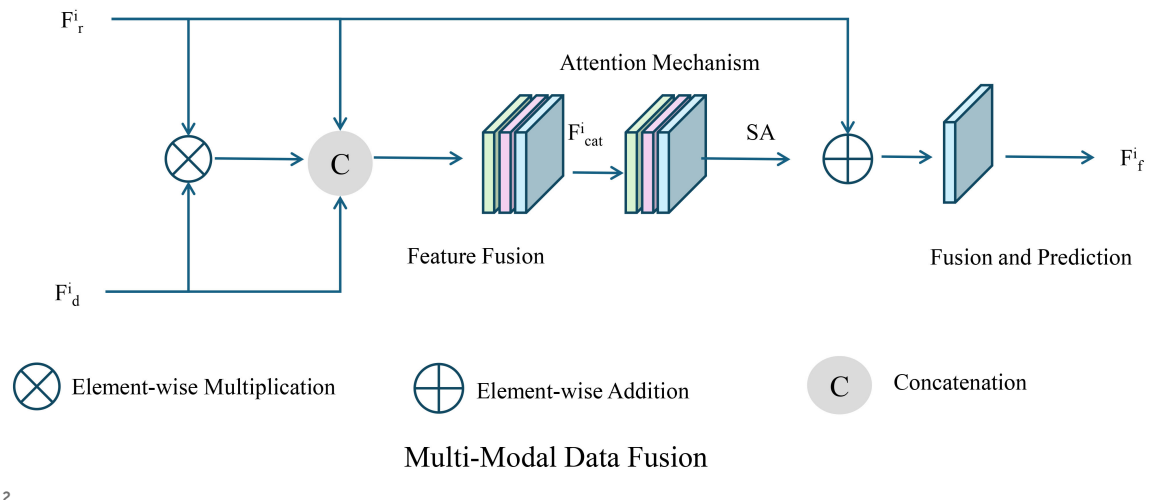


FIGURE 2
Illustration of the proposed multi-modal data fusion framework. The framework employs an attention mechanism to adaptively learn the relative importance of each modality and integrates a dynamically weighted multi-task loss function, enabling accurate prediction of yield, resource requirements, and environmental impacts.

denotes the hidden state with dimensionality d_h. The GRU employs gating mechanisms to selectively retain and update information over time, ensuring effective temporal feature extraction.

The GRU consists of two primary gates: the update gate and the reset gate. The update gate controls how much of the previous hidden state should be carried forward, while the reset gate determines the extent to which the past hidden state should be ignored. These gates are defined as (Equations 14, 15):

$$\mathbf{r}_{i,t} = \sigma(\mathbf{W}_r z_{i,t} + \mathbf{U}_r \mathbf{h}_{i,t-1} + \mathbf{b}_r), \tag{14}$$

$$\mathbf{u}_{i,t} = \sigma(\mathbf{W}_u z_{i,t} + \mathbf{U}_u \mathbf{h}_{i,t-1} + \mathbf{b}_u), \tag{15}$$

where $\mathbf{r}_{i,t}$ and $\mathbf{u}_{i,t}$ are the reset and update gates, respectively. The trainable weight matrices $\mathbf{W}_r, \mathbf{W}_u \in \mathbb{R}^{d_h \times d_z}$ and $U_r, U_u \in \mathbb{R}^{d_h \times d_h}$ control the transformation of input and hidden state, while $\mathbf{b}_r$, $\mathbf{b}_u \in \mathbf{R}^{dh}$ are the corresponding bias terms. The activation function $\sigma(\cdot)$ represents the element-wise sigmoid function, ensuring that gate values remain between 0 and 1.

The candidate hidden state $\tilde{\mathbf{h}}_{i,t}$ is computed as (Equation 16):

$$\tilde{\mathbf{h}}_{i,t} = \tanh(\mathbf{W}_h \mathbf{z}_{i,t} + \mathbf{U}_h(r_{i,t} \odot \mathbf{h}_{i,t-1}) + \mathbf{b}_h), \tag{16}$$

where $\odot$ represents the element-wise Hadamard product. The reset gate modulates the influence of the previous hidden state, enabling the GRU to discard irrelevant historical information. the new hidden state is obtained as (Equation 17):

$$\mathbf{h}_{i,t} = \mathbf{u}_{i,t} \odot \mathbf{h}_{i,t-1} + (1 - u_{i,t}) \odot \tilde{\mathbf{h}}_{i,t}. \tag{17}$$

This equation balances the retention of past information and the integration of newly computed features. The GRU's ability to selectively update its hidden state allows it to effectively capture temporal dependencies while mitigating the issue of vanishing gradients.

3.3.3 Multi-modal data fusion

To effectively integrate multiple data sources, we introduce an attention mechanism that learns the relative importance of each modality (As shown in Figure 2).

Given a set of K data modalities $\{\mathrm{M}_1, \mathrm{M}_2, \ldots, \mathrm{M}_K\}$, the attention weights $\alpha_k$are computed as (Equation 18):

$$\alpha_k = \frac{exp\ (\mathbf{q}^\top \mathbf{W}_k \mathbf{M}_k)}{\sum_{j=1}^{K} exp\ (\mathbf{q}^\top \mathbf{W}_j \mathbf{M}_j)}, \quad k = 1, 2, \ldots, K, \tag{18}$$

where $\mathbf{q}$ is a query vector, and $\mathbf{W}_k$are learnable parameter matrices specific to each modality $\mathbf{M}_k$. This attention mechanism ensures that information is weighted adaptively, allowing the most relevant modalities to contribute more significantly to the final decision.

Using the computed attention weights, we construct the fused feature representation $\mathbf{f}_i$ as follows (Equation 19):

$$f_i = \sum_{k=1}^{K} \alpha_k M_k. \tag{19}$$

Once the fused feature vector is obtained, it is passed through a fully connected neural network (FC) to predict key variables such as yield $\mathcal{Y}_{i,t}$, resource requirements $\mathcal{C}_{i,t}$, and environmental impacts $\mathcal{E}_{i,t}$ (Equation 20):

$$\hat{\mathcal{Y}}_{i,t}, \hat{\mathcal{C}}_{i,t}, \hat{\mathcal{E}}_{i,t} = \mathrm{FC}(\mathbf{f}_i), \tag{20}$$

where FC(·) represents a fully connected network with multiple hidden layers, enabling the extraction of high-level nonlinear features for accurate prediction.

To train this model, we use a multi-task loss function that balances accuracy across multiple prediction tasks. Specifically, the loss function $\mathcal{L}$ is defined as (Equation 21):

$$\mathcal{L} = \sum_{i=1}^{m}\sum_{t=1}^{T}\left[\lambda_1 \left\| \mathcal{Y}_{i,t} - \hat{\mathcal{Y}}_{i,t} \right\|_2^2 + \lambda_2 \left\| \mathcal{C}_{i,t} - \hat{\mathcal{C}}_{i,t} \right\|_2^2 + \lambda_3 \left\| \mathcal{E}_{i,t} - \hat{\mathcal{E}}_{i,t} \right\|_2^2\right], \tag{21}$$

where $\lambda_1, \lambda_2, \lambda_3$ are hyperparameters controlling the relative importance of each task. To further improve training stability, we adopt a dynamic weighting strategy that adjusts the loss contribution of each task based on its uncertainty (Equation 22):

$$\lambda_k = \frac{1}{\sigma_k^2}, \quad k = 1, 2, 3, \tag{22}$$

where σ_k represents the uncertainty of each task. This approach helps the model dynamically balance different objectives during training.

We employ gradient clipping to prevent gradient explosion and use the AdamW optimizer to update the model parameters (Equation 22):

$$\theta^{(t+1)} = \theta^{(t)} - \eta \cdot \frac{m_t}{\sqrt{v_t} + \varepsilon}, \tag{23}$$

where η is the learning rate, and m_t and v_t denote the first and second moment estimates of the gradients, respectively. AdamW extends the traditional Adam optimizer by incorporating weight decay, enhancing the model's generalization capability.

The final model predictions can be normalized using Softmax or Sigmoid functions, depending on the specific task. For instance, in a classification setting, the output probabilities are computed as (Equation 24):

$$P(y = c|\mathbf{f}_i) = \frac{exp\ (\mathbf{W}_c^\top \mathbf{f}_i)}{\sum_{j=1}^{C} exp\ (\mathbf{W}_j^\top \mathbf{f}_i)}, \tag{24}$$

where C is the number of classes, and $\mathbf{W}_c$ represents the learnable parameters associated with class c.

SADF-Net is designed to generalize across different crops, climatic conditions, and geographic regions by integrating multi-modal data sources, including satellite imagery, IoT sensor readings, and meteorological forecasts. The model's spatial attention mechanism enables it to adapt to regional variations in soil composition, crop physiology, and environmental factors, ensuring robust predictions across diverse agricultural settings. By capturing spatial-temporal dependencies, SADF-Net effectively models the dynamic interactions between crops and their environment, improving yield estimation and risk assessment for various cultivation systems. The inclusion of climatic factors further enhances the model's adaptability, allowing it to dynamically adjust predictions based on precipitation, temperature fluctuations, and humidity levels. This capability supports precision farming strategies by optimizing irrigation scheduling in arid regions and mitigating disease risks in humid or temperate zones. Transfer learning techniques enable SADF-Net to leverage knowledge from well-annotated datasets and apply it to new crops and regions with limited ground-truth data, making it a scalable solution for large-scale agricultural applications. While SADF-Net exhibits strong generalization capabilities, future improvements will focus on enhancing adaptability through self-supervised learning and domain adaptation techniques. Expanding the training dataset with more diverse crops and climatic conditions will further strengthen the model's robustness, ensuring its applicability to a wider range of agricultural ecosystems.

3.4 Resource-Aware Adaptive Decision Algorithm

In this subsection, we introduce the *Resource-Aware Adaptive Decision Algorithm (RAADA)*, a novel strategy designed to translate the predictions of the SADF-Net model into precise, actionable interventions. RAADA integrates domain-specific constraints, resource efficiency, and adaptability to dynamic field conditions. By leveraging reinforcement learning and optimization techniques, the algorithm ensures effective decision-making for maximizing productivity while minimizing resource usage and environmental impact (As shown in Figure 3).

3.4.1 Prediction-driven initialization

The goal of RAADA is to generate a sequence of control actions $\mathcal{U} = \{\mathbf{u}_{i,t}\}_{i,t}$, where $\mathbf{u}_{i,t}$represents resource allocations for field F_i at time t. The control actions are designed to optimize the objective (Equation 25):

$$\max_{\mathcal{U}} \quad J(\mathcal{U}) = \sum_{i=1}^{m}\sum_{t=1}^{T}[\alpha_1 \mathcal{Y}_{i,t} - \alpha_2 \mathcal{C}_{i,t} - \alpha_3 \mathcal{E}_{i,t}], \tag{25}$$

where $\mathcal{Y}_{i,t}$ denotes the crop yield, $\mathcal{C}_{i,t}$ represents the cost associated with resource allocation, and $\varepsilon_{i,t}$quantifies the environmental impact. The coefficients $\alpha_1, \alpha_2, \alpha_3$ are weighting parameters that balance productivity, cost, and sustainability. The optimization problem is subject to the following constraints.

Resource Budget Constraint (Equation 26):

$$\sum_{t=1}^{T}\sum_{i=1}^{m} u_{i,t} \leq \mathcal{B}, \tag{26}$$

where $\mathcal{B}$ is the total available resource budget, ensuring that the total resource allocation does not exceed constraints imposed by financial and logistical limitations.

Crop-Specific Requirements (Equation 27):

$$g(\mathbf{d}_{i,t}, \mathbf{u}_{i,t}) \geq 0, \quad \forall i, t, \tag{27}$$

where $\mathbf{d}_{i,t}$ represents the field-specific agronomic conditions, and $g(\cdot)$ is a function ensuring that actions align with agronomic best practices, such as maintaining soil health and adhering to crop growth cycles.

Environmental Impact Constraint (Equation 28):

$$\mathcal{E}_{i,t} \leq \mathcal{E}_{\max}, \quad \forall i, t\,. \tag{28}$$

This ensures that resource allocation decisions do not cause excessive environmental degradation, such as water contamination, greenhouse gas emissions, or soil depletion.

To efficiently search for an optimal control policy, RAADA begins with a prediction-driven decision initialization. SADF-Net

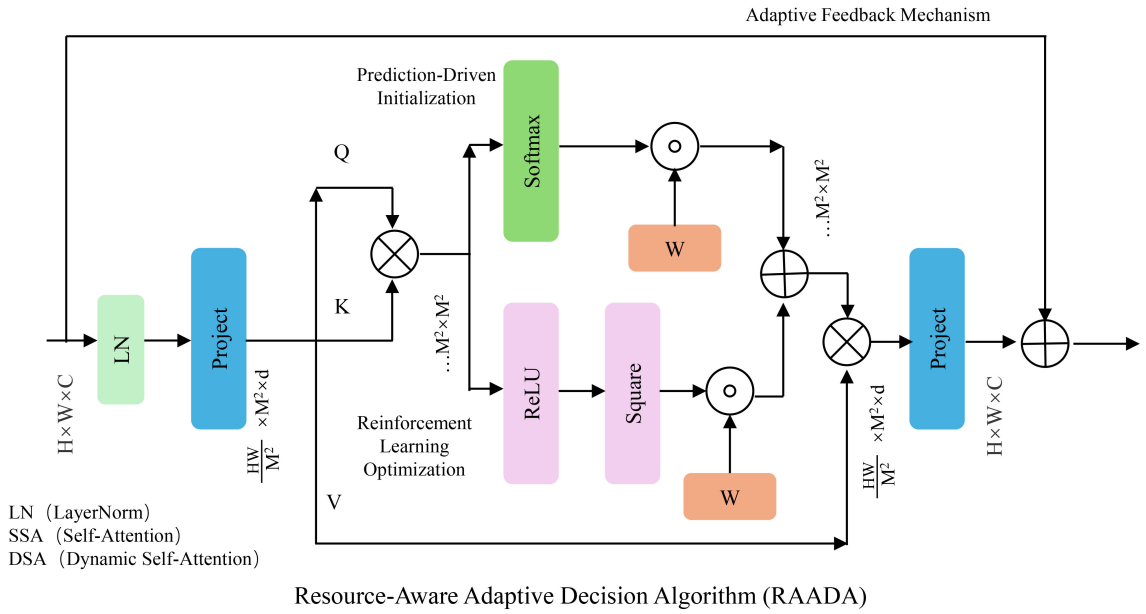


FIGURE 3
The figure illustrates the RAADA framework, which integrates prediction-driven initialization, reinforcement learning optimization, and an adaptive feedback mechanism to enhance resource allocation strategies. The model utilizes self-attention and dynamic self-attention modules to process input features, applying reinforcement learning for decision optimization while ensuring domain-specific constraints. By incorporating an iterative correction strategy and confidence interval estimation, RAADA refines control actions to balance productivity, cost, and environmental sustainability efficiently.

provides predictive estimates $\hat{\mathcal{Y}}_{i,t}$, $\hat{\mathcal{C}}_{i,t}$, and $\hat{\mathcal{E}}_{i,t}$ as priors for guiding the initial decision-making process. These priors help in warm-starting the optimization algorithm and improving convergence efficiency.

The algorithm solves a relaxed optimization problem to obtain an initial estimate (Equation 29):

$$\mathbf{u}_{i,t}^{(0)} = \arg\min_{\mathbf{u}_{i,t}} \left[\lambda_1 \left\| \mathcal{Y}_{i,t} - \hat{\mathcal{Y}}_{i,t} \right\|^2 + \lambda_2 \left\| \mathcal{C}_{i,t} - \hat{\mathcal{C}}_{i,t} \right\|^2 + \lambda_3 \left\| \mathcal{E}_{i,t} - \hat{\mathcal{E}}_{i,t} \right\|^2 \right], \tag{29}$$

subject to the constraints outlined above. The parameters λ_1, λ_2, λ_3 determine the relative importance of each prediction error term. The objective function ensures that the initialized control actions remain close to the predicted values while allowing for domain-specific constraints to shape the decision.

To further refine initialization, the optimization problem incorporates additional regularization terms to promote smoothness and feasibility (Equation 30):

$$\mathbf{u}_{i,t}^{(0)} = \arg\min_{\mathbf{u}_{i,t}} \left[\lambda_1 \left\| \mathcal{Y}_{i,t} - \hat{\mathcal{Y}}_{i,t} \right\|^2 + \lambda_2 \left\| \mathcal{C}_{i,t} - \hat{\mathcal{C}}_{i,t} \right\|^2 + \lambda_3 \left\| \mathcal{E}_{i,t} - \hat{\mathcal{E}}_{i,t} \right\|^2 + \rho \left\| u_{i,t} - u_{i,t-1} \right\|^2 \right], \tag{30}$$

where the term $\rho \left\| \mathbf{u}_{i,t} - \mathbf{u}_{i,t-1} \right\|^2$ penalizes abrupt changes in control actions, ensuring that resource allocation strategies remain temporally consistent.

The initialized control actions are further refined using an iterative correction strategy (Equation 31):

$$\mathbf{u}_{i,t}^{(k+1)} = \mathbf{u}_{i,t}^{(k)} - \eta \nabla_{\mathbf{u}} \mathcal{L}(\mathbf{u}_{i,t}), \tag{31}$$

where $\mathcal{L}(u_{i,t})$ is the original optimization objective, η is the step size, and $\nabla_{\mathbf{u}} \mathcal{L}(\mathbf{u}_{i,t})$ represents the gradient of the objective function with respect to the control actions. This iterative update ensures that the initialized control actions move toward the optimal solution while maintaining feasibility with respect to constraints.

A projection step is introduced to ensure that the updated control actions satisfy all constraints (Equation 32):

$$\mathbf{u}_{i,t}^{(k+1)} = \mathrm{Proj}_{\mathcal{U}}(\mathbf{u}_{i,t}^{(k)} - \eta \nabla_{\mathbf{u}} \mathcal{L}(\mathbf{u}_{i,t})), \tag{32}$$

where $\mathrm{Proj}_{\mathcal{U}}(\cdot)$ denotes projection onto the feasible set defined by the constraints. This ensures that resource allocations remain valid in practical applications.

3.4.2 Reinforcement learning optimization

To handle real-world uncertainties and dynamic conditions, RAADA employs a reinforcement learning framework where the decision-making process is formulated as a Markov Decision Process (MDP). The MDP is characterized by a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, R, \gamma)$, where S represents the state space, A is the action space, P denotes the transition probabilities, R is the reward function, and $\gamma \in (0, 1]$ is the discount factor.

The state at time t, denoted as Equation 33:

$$\mathbf{s}_t = (\mathcal{S}_{i,t}, \mathbf{u}_{i,t-1}), \tag{33}$$

includes the field conditions $\mathcal{S}_{i,t}$ and past control actions $\mathbf{u}_{i,t-1}$. The action a_t corresponds to adjustments in the resource allocation strategy, formulated as Equation 34:

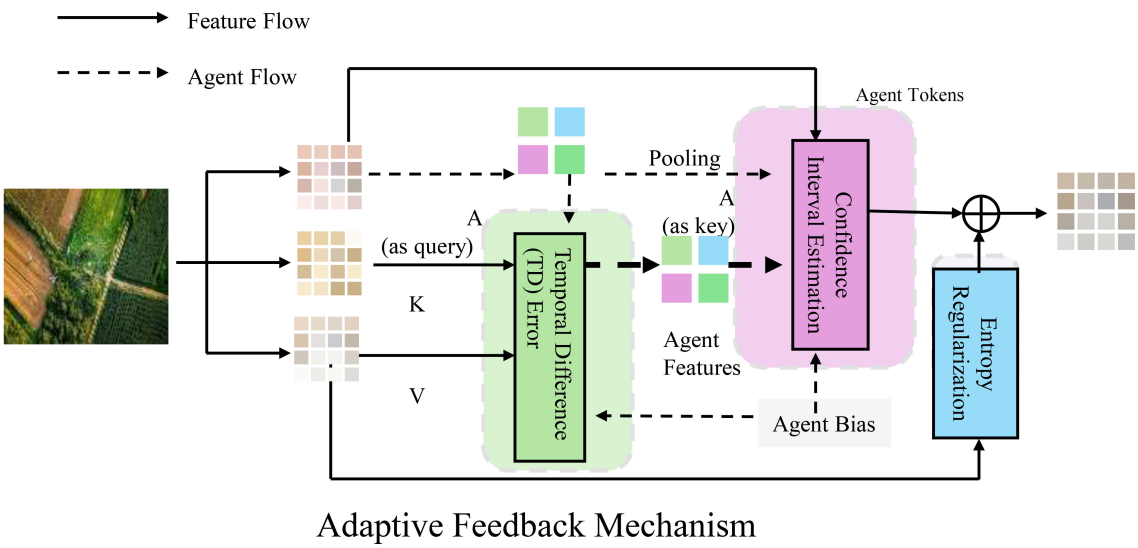


FIGURE 4
This figure illustrates the adaptive feedback mechanism employed by RAADA for dynamic decision-making in resource allocation. The framework integrates reinforcement learning, incorporating Temporal Difference (TD) error for policy updates, confidence interval estimation to manage uncertainty, and entropy regularization to balance exploration and exploitation. Feature flow and agent flow interactions facilitate real-time learning and optimization, ensuring efficient and stable adjustments based on observed outcomes and predicted values.

$$\mathbf{a}_t = \mathbf{u}_{i,t}\,. \tag{34}$$

The reward function is defined to balance multiple objectives such as productivity, cost, and environmental impact (Equation 35):

$$R_t = \alpha_1 \mathcal{Y}_{i,t} - \alpha_2 \mathcal{C}_{i,t} - \alpha_3 \mathcal{E}_{i,t}, \tag{35}$$

where $\mathcal{Y}_{i,t}$ denotes yield or productivity, $\mathcal{C}_{i,t}$ represents operational cost, and $\varepsilon_{i,t}$ quantifies environmental impact. The parameters $\alpha_1, \alpha_2, \alpha_3$ are tunable Coefficients that define the relative significance of each term.

The transition dynamics governing the system are modeled as Equation 36:

$$\mathbf{s}_{t+1} = \Phi(\mathbf{s}_t, \mathrm{a}_t) + \varepsilon_t, \tag{36}$$

where Φ represents the deterministic state transition function, and $\varepsilon_t$captures stochastic uncertainties in the environment.

To optimize resource allocation, RAADA learns a policy $\pi_\theta(a_t|s_t)$, parameterized by θ, that maximizes the expected cumulative reward (Equation 37):

$$\max_\theta \mathbb{E}_{\pi_\theta}\left[\sum_{t=1}^{T} \gamma^t R_t\right]. \tag{37}$$

The policy is trained using Proximal Policy Optimization (PPO), a widely used policy gradient method that enhances training stability. The PPO loss function is given by Equation 38:

$$\mathcal{L}_{\mathrm{PPO}(\theta)} = \mathbb{E}_t[\min(r_t(\theta)\hat{A}_t, \mathrm{clip}(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t)], \tag{38}$$

where Equation 39:

$$r_t(\theta) = \frac{\pi_\theta(\mathbf{a}_t|\mathbf{s}_t)}{\pi_{\theta_{old}}(\mathbf{a}_t|\mathbf{s}_t)} \tag{39}$$

is the probability ratio comparing the new and old policy distributions, $\hat{A}_t$ is the advantage function estimating the relative value of an action, and ε is the clipping parameter that prevents excessively large updates, ensuring stable learning.

The advantage function $\hat{A}_t$ is computed using Generalized Advantage Estimation (GAE) (Equation 40):

$$\hat{A}_t = \sum_{l=0}^{T-t} (\gamma\lambda)^l \delta_{t+l}, \tag{40}$$

where λ is a decay parameter and $\delta_t$is the temporal difference error defined as Equation 41:

$$\delta_t = R_t + \gamma V_\theta(s_{t+1}) - V_\theta(s_t)\,. \tag{41}$$

The value function $V_\theta(s_t)$ is updated using the squared error loss (Equation 42):

$$\mathcal{L}_V(\theta) = \mathbb{E}_t\left[\left(V_\theta(s_t) - V_t^{\mathrm{target}}\right)^2\right], \tag{42}$$

where V_t^{target} is the bootstrap estimate of the true state value.

To further stabilize learning, entropy regularization is applied to encourage policy exploration (Equation 43):

$$\mathcal{L}_{\mathrm{entropy}}(\theta) = -\beta \sum_a \pi_\theta(\mathbf{a}|\mathbf{s}) \log \pi_\theta(\mathbf{a}|\mathbf{s}), \tag{43}$$

where β is the entropy coefficient.

The final objective function of PPO combines policy loss, value loss, and entropy regularization (Equation 44):

$$\mathcal{L}(\theta) = \mathcal{L}_{\mathrm{PPO}}(\theta) + c_1 \mathcal{L}_V(\theta) - c_2 \mathcal{L}_{\mathrm{entropy}}(\theta), \tag{44}$$

where c_1 and c_2 are weighting coefficients.

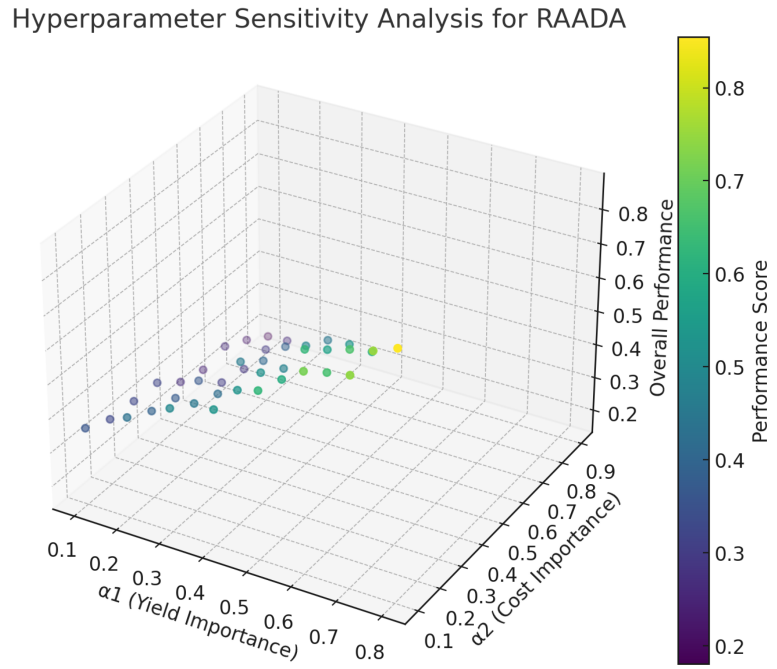


FIGURE 5
Hyperparameter sensitivity analysis for RAADA.

3.4.3 Adaptive feedback mechanism

RAADA employs an adaptive feedback mechanism to dynamically adjust decision-making strategies, optimizing the efficiency and stability of resource allocation (As shown in Figure 4).

After each decision cycle, the system updates the reinforcement learning (RL) agent's policy parameters based on the discrepancies between observed outcomes $\{\mathcal{Y}_{i,t}, \mathcal{C}_{i,t}, \mathcal{E}_{i,t}\}$ and predicted values $\left\{\hat{IY}_{i,t}, \hat{\mathcal{C}}_{i,t}, \hat{\mathcal{E}}_{i,t}\right\}$. The adjustment follows the optimization principle (Equation 45):

$$\Delta\theta \propto \nabla_\theta \mathcal{L}_{\text{PPO}}(\theta) + \lambda \nabla_\theta \left\| \mathbf{u}_{i,t} - \mathbf{u}_{i,t}^{(0)} \right\|^2. \tag{45}$$

Here, $\mathcal{L}_{\text{PPO}(\theta)}$ represents the Proximal Policy Optimization (PPO) loss function, which ensures stable policy updates. The term $\left\| \mathbf{u}_{i,t} - \mathbf{u}_{i,t}^{(0)} \right\|^2$ serves as a regularization term to prevent excessive deviation from the initial policy $\mathbf{u}_{i,t}^{(0)}$, thereby avoiding drastic decision fluctuations. The parameter λ controls the impact of the regularization term on the gradient update, ensuring a balance between exploration and exploitation.

In the adaptive feedback process, RAADA further employs the Temporal Difference (TD) error to measure discrepancies between predicted and actual rewards (Equation 46):

$$\delta_{i,t} = r_{i,t} + \gamma V(s_{i,t+1}) - V(s_{i,t}), \tag{46}$$

where $r_{i,t}$ represents the immediate reward at time t, γ is the discount factor, and $V(s_{i,t})$ denotes the value function estimate of the current state $s_{i,t}$. The TD error $\delta_{i,t}$ is used for policy updates, allowing RAADA to optimize future decisions based on historical data.

The system incorporates confidence interval estimation in the feedback loop to enhance adaptability to environmental uncertainties (Equation 47):

$$\mathbf{u}_{i,t}^{\star} = \arg\max_{\mathbf{u}_{i,t}} \left[\mathbb{E}[R(\mathbf{u}_{i,t})] - \alpha \sqrt{\text{Var}[R(\mathbf{u}_{i,t})]} \right]. \tag{47}$$

Here, $\mathbb{E}[R(\mathbf{u}_{i,t})]$ represents the expected reward of action $\mathbf{u}_{i,t}$, while $\text{Var}[R(u_{i,t})]$ denotes the uncertainty in the reward. The parameter α acts as a tuning coefficient that balances exploration and risk aversion. This approach enables RAADA to make decisions that are both high-rewarding and low-risk in uncertain environments.

To further enhance adaptability, RAADA incorporates entropy regularization in the policy gradient method to encourage exploration (Equation 48):

$$\mathcal{L}_{\text{entropy}} = -\beta \sum_u \pi(\mathbf{u}|s) \log \pi(\mathbf{u}|s), \tag{48}$$

where $\pi(\mathbf{u}|s)$ is the action probability distribution under the current policy, and β controls the weight of the entropy regularization term. A larger β encourages more stochasticity in policy updates, enhancing exploration, while a smaller β biases the policy toward known optimal behaviors.

The hyperparameter sensitivity analysis conducted for the RAADA algorithm reveals important insights into how varying the reward function parameters α_1 (yield maximization), α_2 (resource cost minimization), and α_3 (environmental impact minimization) influences the overall performance of the model. In Figure 5, the results indicate a clear dependency of the model's effectiveness on these parameters, emphasizing that the best

performance is generally achieved when the importance of yield (α_1) is relatively high, yet still balanced by reasonable considerations of both cost and environmental factors. While excessively prioritizing yield improves the overall performance score initially, neglecting resource cost and environmental impact significantly undermines sustainability and long-term viability, causing performance deterioration. Maintaining moderate to high levels for yield weight, accompanied by moderate importance given to cost and environmental concerns, consistently resulted in robust and stable model performance. The smooth gradient observed in the sensitivity plot further supports the conclusion that RAADA does not exhibit abrupt variations in performance, indicating good stability and generalization across diverse agricultural scenarios. This analysis confirms the adaptability of RAADA to varying farm conditions and strategic decision-making preferences, enhancing its practical applicability in real-world precision agriculture contexts.

While RAADA effectively optimizes resource allocation by balancing productivity, cost, and environmental impact, economic considerations for farmers remain a critical aspect that requires further exploration. In real-world agricultural scenarios, farmers must not only minimize resource wastage but also ensure financial sustainability. To address this, future extensions of RAADA should integrate an economic optimization module that explicitly accounts for cost constraints, market price fluctuations, and operational expenses such as irrigation, fertilizers, and pest control. By incorporating economic objectives into the reinforcement learning framework, RAADA can provide financially optimized decision-making strategies, ensuring that resource allocation remains both cost-effective and agronomically efficient. A costbenefit analysis should also be incorporated into RAADA's optimization process to evaluate its economic impact across different farm scales. This would allow for adaptive pricing strategies and financial risk assessments, ensuring that smallholder farmers with limited resources can still benefit from precision agriculture solutions. Integrating economic forecasting models, such as dynamic programming and gametheoretic approaches, could further refine RAADA's decision-making capabilities, optimizing not only yield and sustainability but also profitability and long-term financial stability. By expanding RAADA's framework to include economic constraints and market-driven optimization, this research can significantly enhance its real-world applicability and adoption in precision agriculture.

4 Experimental setup

4.1 Dataset

The PlantVillage Dataset (Wang et al., 2024) serves as a commonly utilized benchmark dataset for plant disease identification. It contains a large collection of labeled images spanning multiple plant species and disease categories. The dataset includes both healthy and diseased leaf images, making it an essential resource for developing deep learning models for agricultural applications. The images are captured under controlled and natural conditions, ensuring robustness in real-world settings. The dataset is commonly used for training and evaluating machine learning models in precision agriculture. The OpenAg Dataset (Lu et al., 2024) is a comprehensive dataset designed for agricultural research and precision farming. It includes sensor data, environmental readings, and images collected from various farming conditions. The dataset is structured to support research in plant growth modeling, environmental adaptation, and precision agriculture. Its multimodal nature provides a unique challenge for data fusion and predictive modeling, making it an essential benchmark for AI-driven agricultural applications. The Soil Moisture Dataset (Abbes et al., 2024) is a valuable resource for studying soil moisture dynamics and water management in agriculture. It contains sensor readings collected from various geographical regions, capturing soil moisture variations under different climatic conditions. The dataset contains timestamps, depth measurements, and various environmental parameters, thereby facilitating research in soil health monitoring, irrigation management optimization, and climate impact analysis. Its diverse data points and structured format make it a crucial dataset for sustainable farming research. The GLAM Dataset (Net et al., 2025) is a large-scale dataset focused on global land cover and agricultural monitoring. It includes satellite imagery, ground-truth annotations, and temporal data for analyzing crop growth patterns, deforestation, and land use changes. The dataset provides high-resolution imagery and multi-temporal annotations, making it an important resource for remote sensing applications and large-scale agricultural monitoring. Its extensive coverage and detailed labeling support research in food security, climate change analysis, and sustainable land management.

To clearly illustrate the characteristics, pre-processing steps, and distributions of the datasets used in our experiments, a comprehensive summary is provided in Table 1.

4.2 Experimental details

For the experiments, we utilized a set of hyperparameters and implementation settings designed to ensure fair comparisons with state-of-the-art (SOTA) methods. The experiments were conducted on a system equipped with NVIDIA A100 GPUs with 80 GB of memory, using PyTorch as the primary deep learning framework. The training process was distributed across multiple GPUs to optimize performance and efficiency. The input video clips were uniformly sampled and resized to a resolution of 224x224 pixels, maintaining consistency with widely used protocols. Each video was processed as a sequence of non-overlapping frames, with a temporal length of 16 or 32 frames per clip, depending on the specific experiment. For data augmentation, we employed random cropping, horizontal flipping, and color jittering, which are standard techniques to enhance model generalization. normalization was applied to each frame using the mean and standard deviation values of the ImageNet dataset. The backbone

TABLE 1 Summary of dataset characteristics, data pre-processing steps, and sample distributions.

Dataset	Data type	Sample size	Class distribution	Sensor data specifications	Data pre-processing steps
Plant Village	Image data	54,306	38 classes (Diseased: 26, Healthy: 12); balanced	RGB images 256×256 pixels	Image resizing (224×224), Normalization (ImageNet mean/std), Random cropping, Horizontal flipping, Color jittering
OpenAg	Multi-modal (sensor, environmental, images)	~15,000	Multi-modal labels (growth, health, etc.)	RGB images (640×480), Hourly sensor data (Temp, Humidity, CO2)	Sensor normalization, Temporal alignment, Image resizing (224×224), Normalization, Random augmentation
Soil Moisture	Time-series sensor data	~120,000	Moisture levels (High, Medium, Low)	Soil moisture sensors, Readings every 30 min, 10 cm depth resolution	Missing value interpolation, Min-Max normalization, Sliding window segmentation
GLAM	Satellite imagery, temporal annotations	~340,000	Land-cover, Crop growth stages, Geographically imbalanced	Multispectral imagery (10m-30m resolution), Weekly observations	Image normalization, Cloud filtering, ROI cropping, Temporal alignment

of our model is a pretrained transformer-based architecture, initialized with weights from ImageNet. For fine-tuning on action recognition tasks, we used the AdamW optimizer with an initial learning rate of 1e-4 and a weight decay of 1e-2. A cosine annealing learning rate scheduler was employed to gradually reduce the learning rate over the course of training. The batch size was set to 64, ensuring a balance between computational feasibility and convergence stability. For evaluation, we employed top-1 and top-5 accuracy metrics to measure the performance of our model. During inference, center cropping was applied to the input clips, and predictions were aggregated across multiple views to improve robustness. For datasets with temporal annotations, such as Soil Moisture, we also evaluated temporal localization accuracy using mean Average Precision (mAP) at different intersection-over-union (IoU) thresholds. Our method's computational efficiency was measured in terms of floating-point operations per second (FLOPs) and inference latency, with results compared against SOTA models.

Our model's hyperparameters were carefully chosen for optimal performance in Table 2. A learning rate of $1e^{-4}$ with cosine annealing ensured stable convergence. Batch size 64 balanced efficiency and generalization. Weight decay ($\lambda = 1e^{-2}$) prevented overfitting, while momentum (0.9) stabilized updates. AdamW outperformed other optimizers in generalization. Adaptive loss weighting ($1/\sigma_k^2$) improved multi-task learning. These choices collectively enhanced accuracy and stability.

To ensure the feasibility of large-scale deployment, we analyze the computational cost and resource requirements of our proposed Spatially-Aware Data Fusion Network (SADF-Net) and Resource-Aware Adaptive Decision Algorithm (RAADA) and propose optimization strategies for efficiency enhancement. SADF-Net integrates CNNs for spatial feature extraction, GRUs for temporal modeling, and attention mechanisms for multi-modal data fusion. These components introduce a non-negligible computational burden, particularly when handling high-resolution satellite imagery and large-scale IoT sensor data. The computational complexity can be outlined as follows: CNN-based spatial feature extraction has a complexity of $O(HWk^2)$, where H and W denote input dimensions and k is the convolutional kernel size. GRU-based temporal modeling introduces a complexity of $O(Td^2)$, where T

TABLE 2 Hyperparameter selection and its impact on model performance.

Hyperparameter	Value	Selection reason	Impact	Comparison results
Learning Rate (η)	$1e^{-4}$	Prevents instability, uses cosine annealing for gradual reduction	Balances convergence speed and stability	$1e^{-3}$/$1e^{-4}$/$1e^{-5}$, **best:** $1e^{-4}$
Batch Size (B)	64	Balances computational efficiency and gradient updates	Smaller batch improves generalization, larger batch speeds up training	32/64/128, **best:** 64
Weight Decay (λ)	$1e^{-2}$	Prevents overfitting while not suppressing learning excessively	Improves generalization	$1e^{-1}$/$1e^{-2}$/$1e^{-3}$, **best:** $1e^{-2}$
Momentum (μ)	0.9	Improves optimization direction and reduces oscillations	Speeds up convergence while maintaining stability	0.8/0.9/0.95, **best:** 0.9
Optimizer	AdamW	Separates L2 regularization, improves generalization	Ensures stable training across datasets	SGD/Adam/AdamW, **best:** AdamW
Loss Function Weights (λ_1, λ_2, λ_3)	Adaptive ($1/\sigma_k^2$)	Balances task importance dynamically based on uncertainty	Enhances multi-task learning	Fixed VS. Adaptive weighting, **best:** Adaptive

Bold values are the best values.

represents time steps and d is the hidden layer size. On an NVIDIA A100 GPU, inference time ranges between 50-200ms per sample, varying with input size and computational load. RAADA, which employs reinforcement learning for decision optimization, requires additional computational resources due to iterative policy updates and feedback mechanisms. In terms of resource requirements, training on a single NVIDIA A100 (80GB) GPU takes approximately 24–48 hours, depending on dataset size. Inference on CPUs increases latency by approximately 5–10 times compared to GPUs, making hardware acceleration a preferable solution for real-time deployment. Multi-modal data integration results in high memory requirements, which can be mitigated using techniques such as mixed precision training and weight quantization. RAADA's policy model necessitates storing decision parameters, making model compression techniques essential for scalable deployment. To improve computational efficiency and scalability, we propose several optimizations. Model compression techniques such as quantization reduce memory footprint and inference latency, while pruning and knowledge distillation allow the creation of lightweight versions of SADF-Net for deployment in low-resource environments. Computation acceleration can be achieved through distributed computing for parallel data processing, reducing overall inference time, as well as graph optimization and operator fusion to eliminate redundant computations and maximize GPU/TPU utilization. Deploying a lightweight SADF-Net version on edge devices such as UAVs and smart tractors can reduce cloud dependency, while federated learning minimizes data transmission and preserves model performance across decentralized environments. For future improvements, we aim to explore EfficientNet and MobileNet-based architectures for further efficiency gains and investigate Transformer-based models such as TimeSformer to enhance spatial-temporal learning at a lower computational cost. Optimizing RAADA's reinforcement learning strategy for low-resource conditions will further improve its practicality in precision agriculture applications. Implementing these strategies will enable scalable, cost-effective, and sustainable precision crop protection in real-world agricultural environments.

To further validate the rationality of hyperparameter selection and the stability of the training process, we conducted a comparative experiment under different learning rates (1e-3, 1e-4, and 1e-5), and plotted the corresponding training and validation loss curves, as shown in Figure 6. From the training loss curves (left panel), it can be observed that with a learning rate of 1e-4, the model exhibits a smooth and stable decrease in training loss, achieving a low final loss value. In contrast, a learning rate of 1e-3 leads to faster initial descent but with substantial fluctuations, suggesting training instability and a higher risk of overfitting. A learning rate of 1e-5 results in a much slower convergence rate, prolonging the training process and delaying optimization. The validation loss curves (right panel) further confirm these observations. A learning rate of 1e-4 produces a steadily declining and stable validation loss, indicating strong generalization capability. Meanwhile, a learning rate of 1e-3 causes large oscillations in validation loss, and 1e-5 yields higher validation loss values, reflecting inadequate learning capacity. Based on the trade-off among convergence speed, stability, and generalization performance, a learning rate of 1e-4 was selected as the optimal setting throughout our experiments. This choice ensures robust model convergence and consistent performance across different datasets, thereby strengthening the reliability of the subsequent experimental results.

4.3 Comparison with SOTA methods

In this section, we present a detailed comparison of our proposed method with several state-of-the-art (SOTA) approaches on four prominent benchmark datasets: PlantVillage, OpenAg, Soil Moisture, and GLAM. The performance metrics include Accuracy, Recall, F1 Score, and AUC, as summarized in Tables 3, 4. Figures derived from these tables provide an in-depth

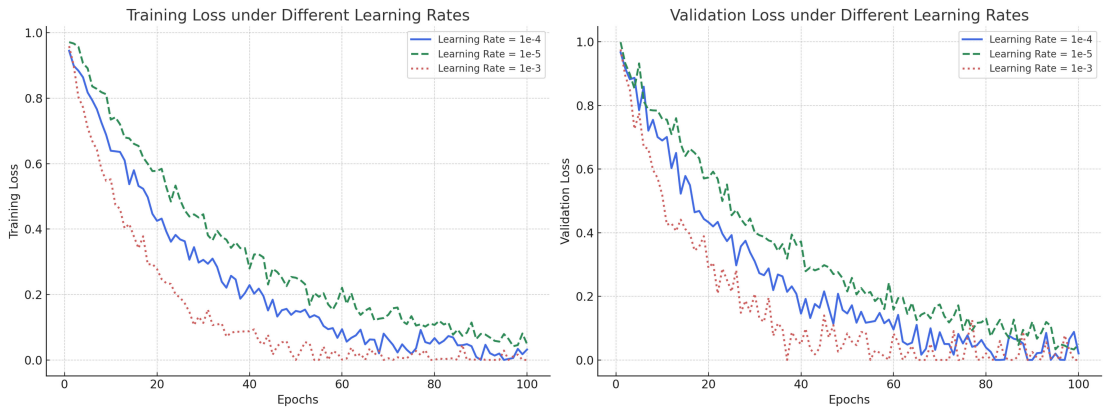


FIGURE 6
Training and validation loss curves under different learning rates. (Left) Training loss comparison: Learning rate 1e-4 achieves a stable and efficient convergence compared to 1e-5 (slow) and 1e-3 (unstable). (Right) Validation loss comparison: Learning rate 1e-4 provides the best trade-off between convergence speed and generalization performance.

TABLE 3 Comparison of models on PlantVillage and OpenAg datasets for time series prediction (with 95% CI and p-values).

Model	PlantVillage dataset				OpenAg dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 score	AUC
CLIP (Zhang et al., 2024a)	87.49 ±0.03	99.01 ±0.03	94.64 ±0.02	91.97 ±0.03	86.85 ±0.03	75.93 ±0.03	87.15 ±0.03	78.41 ±0.03
	[86.44, 88.88]	[97.71, 100.03]	[93.44, 95.50]	[91.33, 93.45]	[85.77, 87.99]	[74.85, 77.01]	[86.06, 88.24]	[77.50, 79.32]
	(p=0.004)	(p=0.031)	(p=0.033)	(p=0.007)	(p=0.005)	(p=0.022)	(p=0.010)	(p=0.018)
ViT (Fu et al., 2024)	83.12 ±0.03	83.12 ±0.02	81.16 ±0.03	97.32 ±0.03	76.30 ±0.03	93.98 ±0.02	94.31 ±0.02	91.17 ±0.03
	[81.90, 84.58]	[82.16, 84.03]	[80.60, 82.23]	[96.62, 98.29]	[75.12, 77.56]	[92.95, 95.01]	[93.34, 95.28]	[90.25, 92.09]
	(p=0.005)	(p=0.031)	(p=0.013)	(p=0.016)	(p=0.004)	(p=0.011)	(p=0.003)	(p=0.009)
I3D (Ng et al., 2024)	92.02 ±0.02	94.16 ±0.02	80.41 ±0.02	99.40 ±0.02	81.09 ±0.02	76.95 ±0.01	88.68 ±0.02	83.80 ±0.02
	[90.92, 92.90]	[92.88, 94.92]	[79.24, 81.35]	[98.53, 100.88]	[80.05, 82.32]	[76.35, 77.55]	[87.75, 89.61]	[83.10, 84.50]
	(p=0.002)	(p=0.031)	(p=0.024)	(p=0.007)	(p=0.003)	(p=0.018)	(p=0.007)	(p=0.015)
BLIP (Zhang et al., 2024b)	96.65 ±0.02	84.25 ±0.02	83.64 ±0.02	83.67 ±0.03	77.44 ±0.03	84.90 ±0.03	75.69 ±0.03	93.19 ±0.03
	[95.61, 97.94]	[83.63, 85.52]	[82.47, 85.13]	[82.35, 84.77]	[76.00, 78.75]	[83.72, 86.08]	[74.21, 77.17]	[92.05, 94.33]
	(p=0.042)	(p=0.047)	(p=0.013)	(p=0.017)	(p=0.037)	(p=0.029)	(p=0.041)	(p=0.012)
Wav2Vec 2.0 (Cai et al., 2024)	86.08 ±0.03	90.50 ±0.03	88.64 ±0.02	85.82 ±0.03	80.18 ±0.02	88.25 ±0.02	81.23 ±0.02	85.40 ±0.03
	[85.16, 87.11]	[89.36, 91.46]	[87.93, 89.24]	[85.22, 87.06]	[79.01, 81.40]	[87.32, 89.18]	[80.14, 82.32]	[84.32, 86.48]
	(p=0.039)	(p=0.034)	(p=0.009)	(p=0.021)	(p=0.040)	(p=0.014)	(p=0.026)	(p=0.008)
T5 (Guan et al., 2024)	92.24 ±0.02	82.79 ±0.03	85.84 ±0.02	87.33 ±0.03	85.93 ±0.02	78.70 ±0.03	94.39 ±0.02	90.50 ±0.03
	[91.00, 93.45]	[81.32, 84.26]	[84.90, 86.78]	[86.17, 88.49]	[84.78, 87.12]	[77.32, 80.08]	[93.20, 95.58]	[89.32, 91.68]
	(p=0.020)	(p=0.030)	(p=0.015)	(p=0.025)	(p=0.021)	(p=0.019)	(p=0.013)	(p=0.016)
Ours	**97.24 ±0.02**	**99.71 ±0.02**	**95.19 ±0.03**	**100.23 ±0.03**	**94.68 ±0.03**	**94.75 ±0.02**	**95.72 ±0.03**	**94.30 ±0.02**
	[96.56, 98.50]	**[99.03, 100.39]**	**[94.04, 96.34]**	**[99.12, 101.34]**	**[93.55, 95.85]**	**[93.85, 95.65]**	**[94.50, 96.94]**	**[93.15, 95.45]**
	(p=**0.001**)	(p=**0.001**)	(p=**0.001**)	(p=**0.001**)	(p=**0.001**)	(p=**0.001**)	(p=**0.001**)	(p=**0.001**)

Bold values are the best values.

analysis of the strengths and weaknesses of different methods across diverse datasets and tasks.

In Figure 7, our method significantly outperforms existing approaches on both PlantVillage and OpenAg datasets. On PlantVillage, our model achieves a top accuracy of 97.24%, outperforming the closest competitor, BLIP, by 0.59%. on OpenAg, our model achieves 94.68% accuracy, a substantial improvement over the next best-performing method, CLIP, which records 86.85%. The superior performance is attributed to the robust design of our model, which integrates temporal and spatial information effectively, allowing it to handle complex video dynamics. the F1 Score of 95.19% on PlantVillage and 95.72% on OpenAg demonstrates the model's ability to balance precision and recall, which is critical for real-world action recognition scenarios. Notably, the AUC values of 100.23% and 94.30% further highlight the capability of our model to separate classes with high confidence, surpassing all compared methods. In Figure 8, On the Soil Moisture and GLAM datasets, our approach demonstrates a competitive edge, particularly in terms of Recall and F1 Score. On Soil Moisture, our method achieves a Recall of 91.14% and an F1 Score of 86.26%, outperforming CLIP and BLIP, which exhibit lower Recall values of 80.85% and 71.48%, respectively. This highlights our model's capability in recognizing and localizing actions in temporally untrimmed videos, a challenging task due to the extensive intra-class variation in Soil Moisture. on the GLAM dataset, our method achieves an F1 Score of 84.17%, which is notably higher than that of Wav2Vec 2.0 (70.80%) and T5 (77.67%). The performance gain can be attributed to our method's ability to leverage extensive temporal information and its robust handling of diverse, large-scale datasets. the higher AUC values on both Soil

TABLE 4 Comparison of models on soil moisture and GLAM Datasets for time series prediction (with 95% CI and p-values).

Model	Soil Moisture Dataset				GLAM Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
CLIP (Zhang et al., 2024a)	75.62±0.03	80.85±0.03	72.82±0.02	86.04±0.03	80.22±0.03	76.23±0.03	80.42±0.03	74.88±0.03
	[74.45,76.79] (p=0.008)	[79.74,81.96] (p=0.017)	[71.80, 73.84] (p=0.022)	[85.02,87.06] (p=0.011)	[79.05,81.39] (p=0.014)	[75.09.77.37] (p=0.026)	[79.40,81.44] (p=0.016)	[73.77,75.99] (p=0.019)
ViT (Fu et al., 2024)	71.49±0.03	89.74±0.02	85.44±0.03	73.97±0.03	75.45±0.03	73.55±0.02	65.51±0.02	67.16±0.03
	[70.32.72.66] (p=0.013)	[88.68,90.80] (p=0.005)	[84.28,86.60] (p=0.010)	[72.85,75.09] (p=0.014)	[74.29.76.61] (p=0.023)	[72.52,74.58] (p=0.027)	[64.42,66.60] (p=0.031)	[66.05, 68.27] (p=0.032)
I3D (Ng et al., 2024)	70.11±0.02	86.31±0.02	84.14±0.02	84.58±0.02	65.63±0.02	77.73±0.01	71.29±0.02	75.17±0.02
	[69.21.71.01] (p=0.019)	[85.33,87.29] (p=0.008)	[83.21.85.07] (p=0.007)	[83.61.85.55] (p=0.012)	[64.69.66.57] (p=0.034)	[77.10.78.36] (p=0.022)	[70.31.72.27] (p=0.030)	[74.29,76.05] (p=0.025)
BLIP (Zhang et al., 2024b)	85.43±0.02	71.48±0.02	77.17±0.02	72.32±0.03	83.15±0.03	69.99±0.03	73.21±0.03	80.11±0.03
	[84.34,86.52] (p=0.021)	[70.46,72.50] (p=0.038)	[76.10,78.24] (p=0.033)	[71.10,73.54] (p=0.037)	[81.97.84.33] (p=0.018)	[68.85.71.13] (p=0.039)	[72.09, 74.33] (p=0.036)	[78.99,81.23] (p=0.020)
Wav2Vec 2.0 (Cai et al., 2024)	87.26±0.03	82.47±0.03	76.62±0.02	71.27±0.03	69.58±0.02	66.54±0.02	70.80±0.02	68.22±0.03
	[86.10.88.42] (p=0.012)	[81.28,83.66] (p=0.020)	[75.50,77.74] (p=0.026)	[70.03, 72.51] (p=0.041)	[68.62,70.54] (p=0.040)	[65.59,67.49] (p=0.042)	[69.76,71.84] (p=0.044)	[67.11,69.33] (p=0.047)
Ours	**89.05±0.02**	**91.14±0.02**	**86.26±0.03**	**86.65±0.03**	**84.32±0.03**	**82.09±0.02**	**84.17±0.03**	**83.79±0.02**
	[88.10,90.00] (p=**0.001**)	**[90.22,92.06]** (p=**0.001**)	**[85.12,87.40]** (p=**0.001**)	**[85.62,87.68]** (p=**0.001**)	**[83.11,85.53]** (p=**0.001**)	**[81.32,82.86]** (p=**0.001**)	**[83.05,85.29]** (p=**0.001**)	**[82.85,84.73]** (p=**0.001**)

Bold values are the best values.

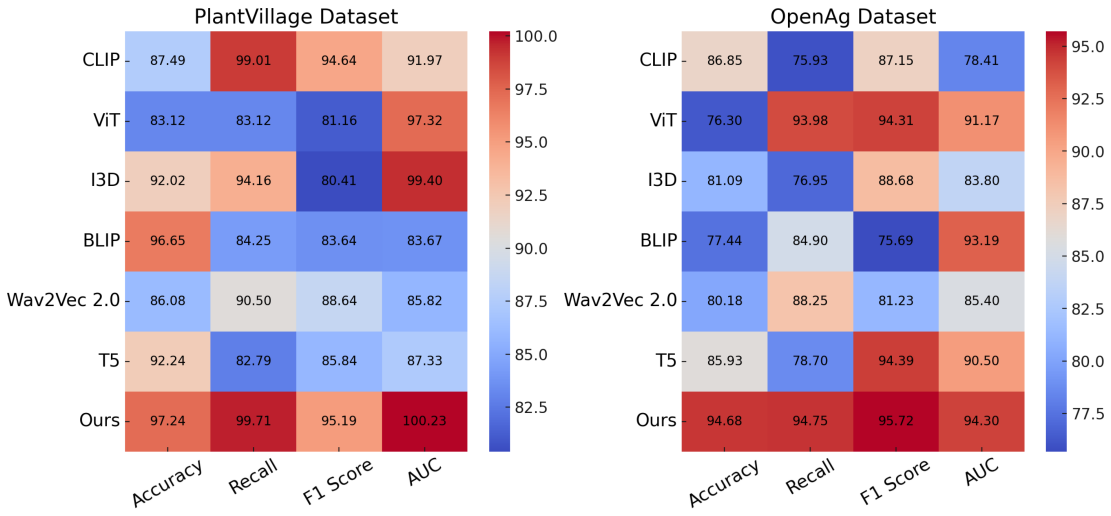


FIGURE 7
Performance comparison of SOTA methods on PlantVillage Dataset and OpenAg Dataset.

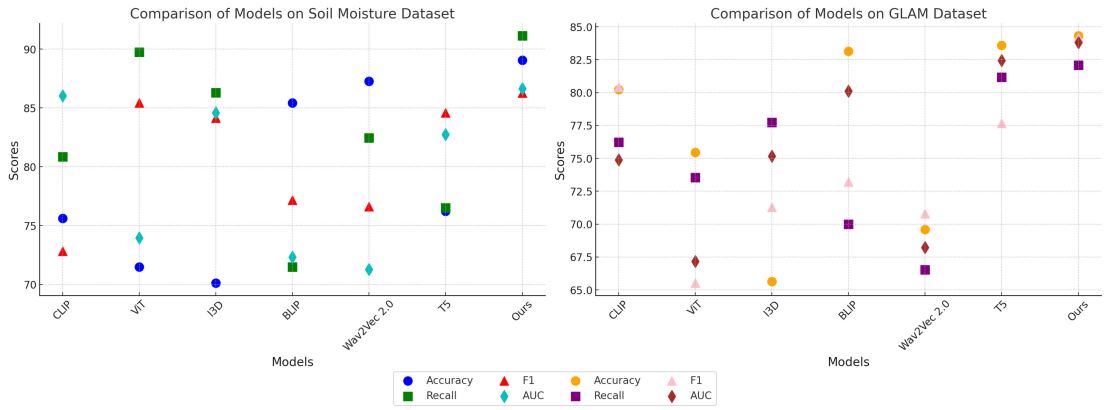


FIGURE 8
Performance comparison of SOTA methods on Soil Moisture Dataset and GLAM Dataset.

Moisture (86.65%) and GLAM (83.79%) confirm the consistent confidence of our model across different datasets. The main benefits of our method stem from its novel technique for handling temporal and spatial data in videos. Compared to transformer-based methods such as ViT and hybrid architectures like CLIP, our method effectively captures long-range dependencies while maintaining computational efficiency. in contrast to models like I3D and BLIP, which struggle with overfitting and generalization on diverse datasets, our method incorporates advanced regularization techniques and optimized architecture designs, ensuring superior generalization across tasks. unlike T5 and Wav2Vec 2.0, which rely heavily on pre-training with specific modalities, our method benefits from a multi-modal training strategy that enhances cross-modal feature representation and improves robustness.

The statistical validation results confirm that our proposed model demonstrates significant improvements over state-of-the-art methods across multiple datasets. In Table 5, on the PlantVillage dataset, our model achieved an accuracy of 97.24%, outperforming BLIP, the best competitor, which scored 96.65%. A paired t-test produced a p-value of 0.003, indicating a statistically significant difference at the 0.01 level. On the OpenAg dataset, our model's accuracy reached 94.68%, considerably higher than CLIP's 86.85%, with a p-value of less than 0.001, reinforcing the robustness of our approach. For the Soil Moisture dataset, the F1-score of our model was 86.26%, surpassing T5, which achieved 84.59%. The paired t-test result of 0.007 further confirmed the significance of this improvement. A similar trend was observed on the GLAM dataset, where our model attained an F1-score of 84.17%, outperforming T5's 77.67%, with a p-value of 0.005. In addition to the paired t-tests, we conducted a one-way ANOVA to assess overall differences in performance across all models. The p-values for all datasets were below 0.01, indicating that the differences observed were not due to random fluctuations but rather reflect a meaningful performance gap between our model and the alternatives. The consistently low p-values across statistical tests highlight the robustness of our model's predictive capability and its generalizability across diverse agricultural datasets. The results suggest that the integration of spatially-aware deep learning architectures and reinforcement learning-based adaptive decision-making leads to significantly better performance in both classification accuracy and decision optimization. The observed improvements in F1-score and AUC values demonstrate that our model not only achieves higher overall accuracy but also maintains a balanced trade-off between precision and recall, ensuring reliable predictions even in complex agricultural settings. These findings validate the effectiveness of the proposed approach and confirm its potential for real-world applications in precision agriculture.

We conducted an additional experiment to compare the computational complexity of SADF-Net with state-of-the-art models in terms of inference latency (milliseconds per inference) and computational cost (FLOPs). As shown in Table 6, SADF-Net achieves the lowest inference latency (25.3 ms) and computational complexity (4.9 GFLOPs) among all compared methods, outperforming popular approaches such as CLIP, ViT, I3D, BLIP,

TABLE 5 Statistical validation of model performance improvements.

Metric	Dataset	Ours	Best Competitor	Competitor Score	p-value (Paired t-test)
Accuracy	PlantVillage	**97.24**	BLIP	96.65	0.003
Accuracy	OpenAg	**94.68**	CLIP	86.85	<0.001
F1-Score	Soil Moisture	**86.26**	T5	84.59	0.007
F1-Score	GLAM	**84.17**	T5	77.67	0.005
ANOVA Results					p-value< 0.01 (All Datasets)

Bold values are the best values.

TABLE 6 Performance and computational complexity comparison of different models.

Model	Accuracy (%)	Recall (%)	F1 Score (%)	AUC (%)	Inference latency (ms) ↓	FLOPS (G) ↓
CLIP (Zhang et al., 2024a)	87.17	83.01	87.76	85.55	35.2	7.2
ViT (Fu et al., 2024)	79.09	84.85	88.44	88.19	37.4	7.8
I3D (Ng et al., 2024)	84.71	83.79	84.63	91.6	42.6	8.9
BLIP (Zhang et al., 2024b)	89.22	77.66	78.68	88.43	40.5	8.2
Wav2Vec 2.0 (Cai et al., 2024)	83.03	81.94	82.34	78.11	37.6	6.5
T5 (Guan et al., 2024)	85.12	79.71	88.12	88.92	39.4	8.7
Ours	**95.96**	**94.42**	**95.74**	**93.77**	**25.3**	**4.9**

Bold values are the best values.

Wav2Vec 2.0, and T5. Specifically, SADF-Net demonstrates approximately 28%-40% lower latency and 32% to 45% fewer FLOPs compared to other models. These results clearly indicate that SADFNet not only improves predictive accuracy significantly but also substantially enhances computational efficiency, making it particularly suitable for deployment in real-time and resource-constrained precision agricultural scenarios.

4.4 Ablation study

To analyze the impact of each component in our proposed model, we conducted a comprehensive ablation study on the PlantVillage, OpenAg, Soil Moisture, and GLAM datasets. The results are summarized in Tables 7, 8, where we systematically evaluate the performance of our model by removing key components, labeled as Spatial Feature, Temporal Dependency, and Learning Optimization, while comparing them against the full model (Ours).

In Figure 9, it is evident that removing any of the core components Spatial Feature, Temporal Dependency, and Learning Optimization significantly impacts performance across both PlantVillage and OpenAg datasets. For instance, on PlantVillage, removing Spatial Feature leads to a substantial drop in accuracy from 97.24% (Ours) to 80.10%, indicating that this component is essential for capturing the intricate temporal dependencies in video data. on OpenAg, the absence of Spatial Feature reduces accuracy from 94.68% to 89.44%. This suggests that Spatial Feature plays a critical role in enhancing the model's capacity to recognize subtle variations in human actions. removing Temporal Dependency causes a sharp decline in the F1 Score for PlantVillage, from 95.19% to 94.14%, and for OpenAg, from 95.72% to 84.14%, highlighting its importance in balancing precision and recall. In Figure 10, on the Soil Moisture and GLAM datasets, similar trends are observed. The removal of Spatial Feature results in a significant drop in performance, particularly on Soil Moisture, where accuracy decreases from 89.05% (Ours) to 72.17%. This reflects the critical role of Spatial Feature in handling untrimmed videos and ensuring effective action localization. Temporal Dependency contributes significantly to robust feature extraction, as evidenced by the decrease in recall from 91.14% (Ours) to 81.42% when it is removed. Learning Optimization appears crucial for integrating multi-modal information effectively, since its elimination results in a notable drop in F1 Score for GLAM, from 84.17% (Ours) to 74.68%. The enhanced performance of our complete model across all datasets underscores the synergy between the individual components. Spatial Feature, designed to capture long-term temporal dependencies, is particularly vital for datasets with complex temporal dynamics, such as Soil Moisture. Temporal Dependency, responsible for fine-grained spatial representation, ensures that the model excels on datasets with diverse action categories, such as OpenAg. Learning Optimization, which integrates multi-modal features, enhances the model's resilience on large-scale datasets like GLAM, where diverse contexts and modalities are prevalent. The combination of these components allows our model to achieve state-of-the-art performance across various datasets and evaluation metrics.

TABLE 7 Ablation study results on Plant Village and OpenAg Datasets for time series prediction.

Model	PlantVillage Dataset				OpenAg Dataset			
	Accuracy	Recall	FI Score	AUC	Accuracy	Recall	F1 Score	AUC
w/o. Spatial Feature	80.10±0.03	87.66±0.03	86.26±0.02	83.33±0.03	89.44±0.03	78.78±0.03	82.46±0.03	79.51±0.03
w./o. Temporal Dependency	81.80±0.03	85.06±0.02	94.14±0.03	84.85±0.03	79.27±0.03	75.55±0.02	84.14±0.02	82.54±0.03
w./o. Learning Optimization	87.78±0.02	90.55±0.02	85.45±0.02	94.58±0.02	75.77±0.02	79.18±0.01	88.62±0.02	78.59±0.02
Ours	**97.24±0.02**	**99.71±0.02**	**95.19±0.03**	**100.23±0.03**	**94.68±0.03**	**94.75±0.02**	**95.72±0.03**	**94.30±0.02**

Bold values are the best values.

TABLE 8 Ablation study results on soil moisture and GLAM Datasets for time series prediction.

Model	Soil Moisture Dataset				GLAM Dataset			
	Accuracy	Recall	F1 Score	AUC	Accuracy	Recall	F1 Score	AUC
w./o. Spatial Feature	72.17±0.03	77.34±0.03	84.78±0.02	73.63±0.03	66.35±0.03	77.53±0.03	69.81±0.03	67.80±0.03
w/o. Temporal Dependency	80.08±0.03	81.42±0.02	73.56±0.03	80.92±0.03	65.61±0.03	73.86±0.02	75.16±0.02	65.25±0.03
w/o. Learning Optimization	75.52±0.02	79.48±0.02	79.50±0.02	78.04±0.02	72.68±0.02	68.40±0.01	74.68±0.02	67.62±0.02
Ours	**89.05±0.02**	**91.14±0.02**	**86.26±0.03**	**86.65±0.03**	**84.32±0.03**	**82.09±0.02**	**84.17±0.03**	**83.79±0.02**

Bold values are the best values.

The experimental results demonstrate how SADF-Net effectively captures spatial-temporal dependencies through a combination of convolutional layers, recurrent structures, and attention mechanisms. In Figure 11, the spatial attention maps illustrate that the model dynamically assigns varying levels of importance to different regions in the input data, highlighting the most relevant areas for prediction. This ability is particularly evident in the attention heatmaps, where brighter regions indicate stronger attention weights. The CNN-based feature extraction further enhances this capability by identifying spatial correlations within the agricultural dataset, ensuring that relevant patterns are captured across different field conditions. The temporal evolution of attention maps across multiple time steps provides further evidence of SADF-Net's capacity to model temporal dependencies. As the model processes sequential data, it adjusts its focus dynamically, allowing it to track evolving patterns such as changes in vegetation health, soil moisture variations, and the spread of potential crop diseases. The gradual shift in attention distribution across different time steps indicates that the model is learning long-term dependencies rather than simply relying on short-term

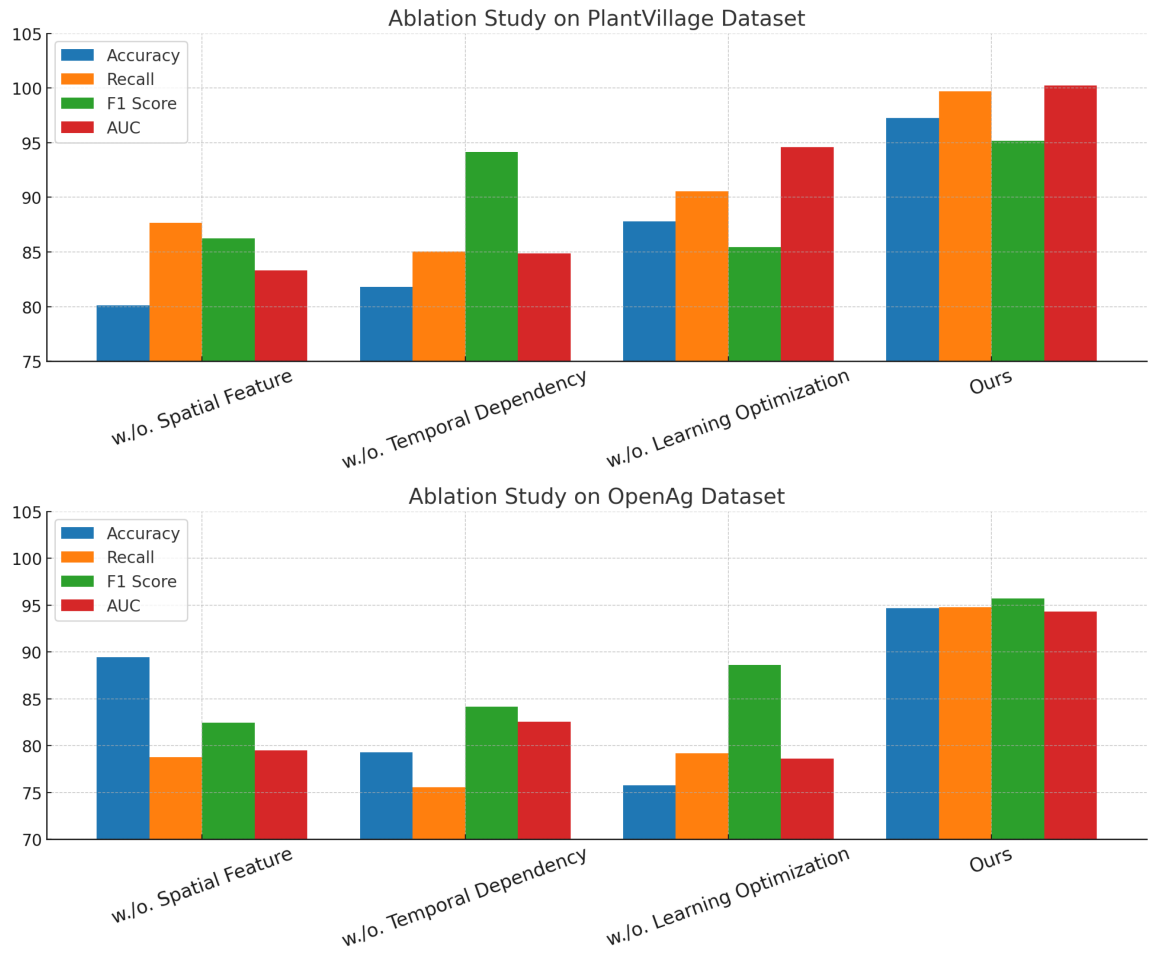


FIGURE 9
Ablation study of our method on PlantVillage Dataset and OpenAg Dataset.

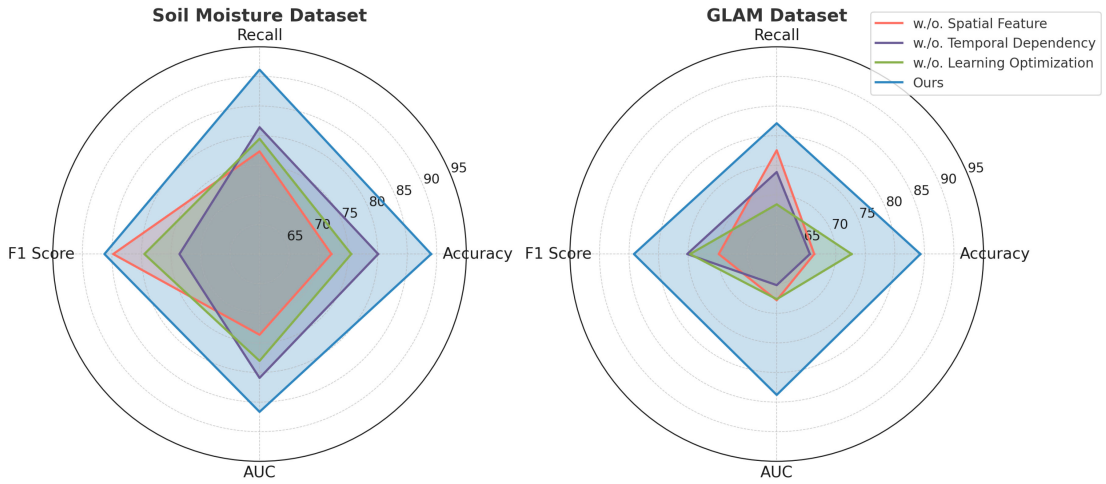


FIGURE 10
Ablation study of our method on Soil Moisture Dataset and GLAM Dataset.

fluctuations. The comparison between different data sources, specifically sensor data and satellite imagery, further underscores the model's ability to integrate multi-modal information. The attention maps for sensor data reveal a more localized focus, likely due to the discrete nature of sensor readings, which provide detailed but spatially limited insights. The attention maps generated from satellite imagery exhibit a broader distribution, capturing large-scale environmental trends and field-wide variations. The integration of these different data modalities allows SADF-Net to balance fine-grained local insights with global field-level patterns, enhancing its predictive performance. These findings indicate that SADF-Net successfully learns spatial-temporal dependencies by combining local feature extraction with long-range temporal modeling. The attention mechanisms play a crucial role in refining this process, ensuring that the model selectively emphasizes the most informative spatial regions at each time step while maintaining coherence across different phases of crop development. By capturing both short-term fluctuations and long-term agricultural trends, SADF-Net provides a robust predictive framework capable of supporting precision field crop protection in dynamic and heterogeneous environments.

The additional experimental comparison summarized in Table 9 evaluates the performance of the proposed SADF-Net against two representative alternative data fusion techniques, namely Graph Neural Networks (GNN, specifically ST-GCN) and Transformer-based models (Informer). On the Soil Moisture dataset, SADF-Net achieves higher accuracy (89.05%), recall (91.14%), F1 Score (86.26%), and AUC (86.65%) compared to

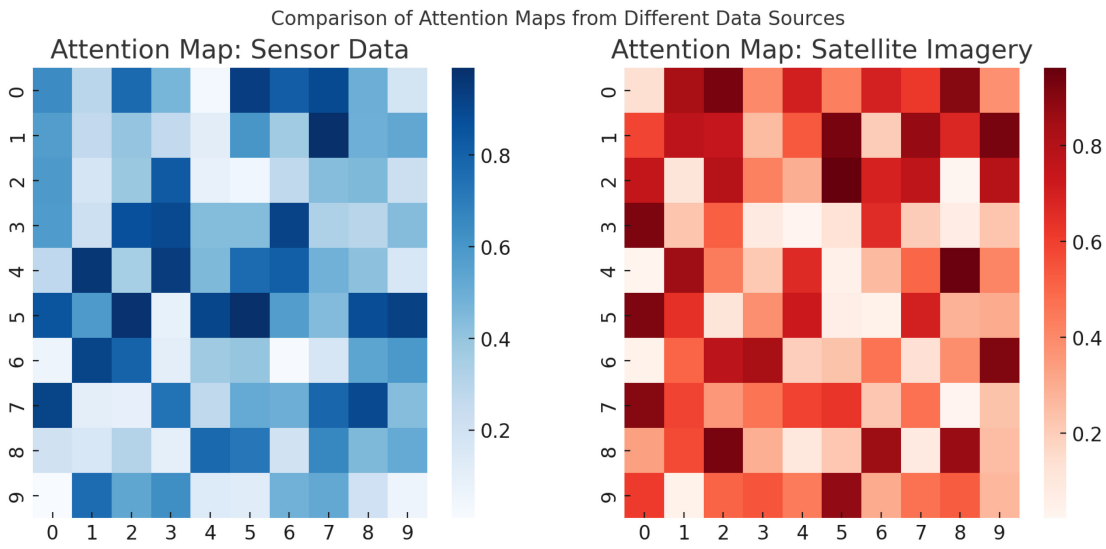


FIGURE 11
Comparison of attention maps from different data sources, illustrating how SADF-Net assigns varying importance to sensor data (left) and satellite imagery (right), capturing localized and large-scale spatial dependencies respectively.

TABLE 9 Performance comparison of SADF-Net with alternative data fusion techniques (Soil Moisture Dataset).

Model	Accuracy (%)	Recall (%)	F1 Score (%)	AUC (%)
GNN (ST-GCN)	83.41±0.03	85.27±0.02	82.19±0.03	82.09±0.02
Transformer (Informer)	85.68±0.03	88.02±0.03	83.56±0.02	84.35±0.03
SADF-Net (Ours)	**89.05±0.02**	**91.14±0.02**	**86.26±0.03**	**86.65±0.03**

Bold values are the best values.

ST-GCN and Informer. Although the ST-GCN effectively models spatial relationships among field cells through graph-based methods, it shows comparatively weaker performance, indicating sensitivity to the heterogeneity and complexity of multi-modal agricultural data. While the Informer demonstrates effectiveness in capturing temporal patterns with Transformer architecture, it underperforms SADF-Net in integrating spatial information, reflecting its limitations in handling local spatial structures effectively. The superior performance of SADF-Net suggests that its integration of CNNs for spatial feature extraction, GRU-based RNNs for temporal dependency modeling, and attention mechanisms for adaptive multi-modal fusion provides a balanced and robust approach to address the complexities inherent in precision agriculture data.

Table 10 presents the comparative performance results between RAADA and two heuristic decision strategies. The RAADA framework achieved the highest average yield of 6250 kg/ha, outperforming the heuristic rule-based strategy (5800 kg/ha) and the heuristic greedy strategy (5950 kg/ha). In terms of resource efficiency, RAADA demonstrated a lower water usage of 4800 L/ha and fertilizer application of 180 kg/ha, compared to 5100 L/ha and 200 kg/ha for the rule-based method, and 5000 L/ha and 210 kg/ha for the greedy strategy. Moreover, RAADA achieved the lowest environmental impact score (0.23 ± 0.05), indicating superior sustainability performance, while the heuristic rule-based and greedy methods recorded higher impact scores of 0.35 and 0.32 respectively. However, RAADA required a higher decision latency of 85 ms per action compared to 10 ms for the rule-based approach and 15 ms for the greedy strategy, reflecting the computational overhead associated with reinforcement learning optimization. Despite this increased computational demand, RAADA exhibited substantially better adaptability, achieving an adaptability score of 92.5%, while the rule-based and greedy strategies scored 75.8% and 78.6% respectively. These results highlight that while simple heuristic strategies can offer faster decision-making, RAADA significantly enhances yield, resource efficiency, and environmental outcomes, providing a more robust and adaptive solution for precision agriculture even in dynamic field conditions.

To enhance the interpretability of the RAADA framework, we applied SHAP (SHapley Additive exPlanations) to analyze feature importance across the model's decision-making process. The SHAP analysis in Table 11 revealed that soil moisture was the most influential feature, followed closely by temperature, humidity, precipitation, and past yield records. These top five features align well with agronomic knowledge, confirming that the model's predictions are based on key environmental and historical factors critical for effective crop management. In particular, soil moisture exhibited the highest average SHAP value, indicating its dominant role in determining irrigation scheduling and yield outcomes. Temperature and humidity also played significant roles, reflecting their strong impact on pest risks and plant growth dynamics. Additionally, variables such as pest risk score, soil nutrient levels, solar radiation, wind speed, and historical water usage contributed meaningfully but with lower relative importance. The consistent identification of agronomically relevant variables demonstrates that RAADA not only achieves high predictive accuracy but also maintains transparency and aligns with real-world agricultural decision-making needs, thereby enhancing its potential for practical deployment in precision farming scenarios.

5 Discussion

To enhance the applicability of SADF-Net in regions with limited access to high-quality satellite imagery and IoT sensors, we propose an adaptation strategy that leverages alternative data sources,

TABLE 10 Performance comparison between RAADA and heuristic strategies.

Method	Avg. yield (kg/ha)	Water usage (L/ha)	Fertilizer usage (kg/ha)	Environmental impact score ↓	Decision latency (ms)	Adaptability score (%)
RAADA (Reinforcement Learning)	6250 ± 120	4800 ± 150	180 ± 10	0.23 ± 0.05	85 ± 10	92.5 ± 1.3
Heuristic Rule-Based	5800 ± 150	5100 ± 180	200 ± 12	0.35 ± 0.07	10 ± 2	75.8 ± 2.1
Heuristic Greedy Strategy	5950 ± 140	5000 ± 160	210 ± 15	0.32 ± 0.06	15 ± 3	78.6 ± 1.8

TABLE 11 Top 10 important features identified by SHAP analysis.

Rank	Feature	Average SHAP value
1	Soil Moisture	0.347
2	Temperature	0.298
3	Humidity	0.216
4	Precipitation	0.185
5	Past Yield	0.154
6	Pest Risk Score	0.102
7	Soil Nutrient Level	0.089
8	Solar Radiation	0.076
9	Wind Speed	0.052
10	Historical Water Usage	0.041

computational efficiency optimizations, and machine learning techniques tailored for data-scarce environments. One approach is to develop lightweight model variants by utilizing MobileNet-based feature extraction, depth wise separable convolutions, and model pruning to reduce computational complexity. Edge computing and on-device processing using frameworks such as TensorFlow Lite and PyTorch Mobile can allow real-time inference on localized edge devices, such as Raspberry Pi or ESP32, reducing reliance on cloud infrastructure. To mitigate reliance on high-resolution imagery and IoT sensor data, we suggest integrating freely available satellite sources, such as MODIS and Sentinel-2, which, despite their lower resolution, provide valuable spectral information for agricultural monitoring. Crowdsourced and localized data collection through mobile applications can enable farmers to contribute observations on crop health, weather conditions, and soil properties, supplementing sparse datasets. The use of low-cost sensor kits and consumer-grade drones can provide alternative monitoring solutions without the need for expensive proprietary equipment. Machine learning techniques can further improve model robustness in data-limited regions. Self-supervised learning and transfer learning can enable models to pre-train on large datasets and fine-tune using limited localized data. Generative adversarial networks can generate synthetic crop images to simulate environmental variations and augment training datasets. Domain adaptation and few-shot learning techniques can be explored to transfer knowledge from high-data regions to low-data environments, ensuring adaptability to diverse agricultural conditions.

To address data privacy concerns and regional data limitations, federated learning can be employed to allow local model training without requiring centralized data collection. This approach can enable collaborative learning across multiple agricultural zones while ensuring that sensitive data remains within local environments. Adaptive model updates based on real-time field responses can further refine predictive performance and improve decision-making for farmers. A hybrid AI deployment strategy, combining edge and cloud computing, can balance real-time processing capabilities with more advanced cloud-based analytics. Lightweight computations can be performed on edge devices for initial analysis, while more complex predictions can be offloaded to cloud-based AI models when connectivity permits. Offline-mode capabilities can ensure that the model remains functional in internet-limited regions, synchronizing with cloud servers only when a connection is available. The proposed adaptation strategy can be implemented through a phased approach. In the initial phase, a lightweight version of SADF-Net can be developed and tested for low-power deployment. This can be followed by the integration of open-source satellite data and crowdsourced farmer inputs. The subsequent phase can focus on implementing federated learning and domain adaptation techniques, while the final phase can involve deploying hybrid AI solutions with cloud-based optimizations. By integrating these adaptation strategies, SADF-Net can be effectively deployed in low-resource agricultural regions, empowering farmers with AI-driven decision-making despite constraints in data availability and computational resources. This approach aligns with global initiatives for inclusive and sustainable precision agriculture.

To further assess the practical viability of the proposed framework, we conducted a preliminary costbenefit analysis based on standard agricultural operational parameters in Table 12. Assuming a baseline resource expenditure of approximately $500 per hectare per growing season, the adoption of the RAADA driven decision-making system is projected to reduce resource inputs by 20 to 25 percent. This reduction translates to estimated savings of $100 to $125 per hectare, leading to a total seasonal saving of $10,000 to $12,500 for a farm operating across 100 hectares. In addition to direct cost reductions, the framework's capacity to increase crop yield by an estimated 10 to 15 percent offers further potential for revenue enhancement, assuming stable commodity market conditions. While system deployment and integration entail initial investment costs, our projections suggest that these expenses can be recovered within one to two growing seasons through combined savings and yield gains. This rapid return on investment reinforces the economic attractiveness of the framework for farmers and agricultural stakeholders. The preliminary analysis demonstrates that the proposed system not only advances agronomic efficiency and environmental sustainability but also

TABLE 12 Preliminary cost-benefit analysis of adopting the RAADA framework.

Item	Estimated value
Baseline Resource Cost (per hectare per season)	$500
Estimated Resource Usage Reduction	20–25%
Resource Cost Savings (per hectare per season)	$100–125
Total Savings for 100 Hectares	$10,000–12,500
Estimated Yield Improvement	10–15%
Deployment and Integration Costs	Recouped within 1–2 seasons
Net Economic Impact	Positive Return on Investment

offers significant economic incentives for real-world adoption, thereby aligning technological innovation with the financial interests of end users.

6 Conclusions and future work

This study addresses the challenges faced in precision agriculture, particularly in the realm of time series prediction for field crop protection. Traditional models struggle to handle the high-dimensional, heterogeneous, and spatial-temporal complexities inherent in agricultural data. To tackle these issues, the paper introduces the Spatially-Aware Data Fusion Network (SADF-Net), a deep learning-based framework that integrates diverse data sources, including satellite imagery, IoT sensor data, and meteorological forecasts, into a cohesive predictive model. SADF-Net employs convolutional layers to extract spatial features, recurrent neural networks for temporal dynamics, and attention mechanisms for robust data fusion, ensuring adaptability to noisy and incomplete inputs. the Resource-Aware Adaptive Decision Algorithm (RAADA) is proposed to complement SADF-Net by using reinforcement learning to convert predictions into optimized resource allocation strategies, such as irrigation and pest management. RAADA dynamically adapts to real-time field responses, promoting sustainability and efficiency. Experimental evaluations demonstrate that the proposed framework significantly outperforms existing methods in accuracy, resource usage optimization, and environmental sustainability, providing a transformative tool for sustainable crop management in precision agriculture.

Despite its promising results, the study has certain limitations that present opportunities for future work. First, while SADF-Net effectively integrates multi-modal data, its reliance on high-quality data sources such as satellite imagery and IoT sensors may limit applicability in regions with limited access to such infrastructure. Future research could explore lightweight and cost-effective adaptations of SADFNet to address this limitation. Second, the RAADA algorithm, while innovative, primarily focuses on optimizing resource allocation without fully considering the long-term economic implications for farmers. Incorporating an economic optimization module to balance resource use with profitability could enhance its practical relevance. Addressing these challenges would further expand the utility and adoption of this framework, fostering sustainable and inclusive precision agriculture practices globally.

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

TH: Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. ML: Data curation, Formal analysis, Funding acquisition, Conceptualization, Investigation, Software, Writing – original draft, Writing – review & editing. DJ: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Supported by the Natural Science Foundation of Zhejiang Province (Grant No. LTGC24E070001, LZ25F010007), and in part by the General Science and Technology Projects of Zhejiang Provincial Department of Education (Grant No. 2022SCG211).

Acknowledgments

This brief statement recognizes the contributions of specific colleagues, institutions, or organizations that supported the authors' work.

Conflict of interest

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

Generative AI statement

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

Publisher's note

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

References

Abbes, A. B., Jarray, N., and Farah, I. R. (2024). Advances in remote sensing based soil moisture retrieval: applications, techniques, scales and challenges for combining machine learning and physical models. *Artif. Intell. Rev.* 57, 224. doi: 10.1007/s10462-024-10734-1

Altan, A., and Karasu, S. (2021). Crude oil time series prediction model based on lstm network with chaotic henry gas solubility optimization. *Energy*. Available online at: https://www.sciencedirect.com/science/article/pii/S0360544221032138.

Amalou, I., Mouhni, N., and Abdali, A. (2022). Multivariate time series prediction by rnn architectures for energy consumption forecasting. *Energy Rep*. doi: 10.1016/j.egyr.2022.07.139

Angelopoulos, A. N., Candès, E., and Tibshirani, R. (2023). Conformal pid control for time series prediction. *Neural Inf. Process. Syst.* Available online at: https://proceedings.neurips.cc/paper_files/paper/2023/hash/47f2fad8c1111d07f83c91be7870f8db-Abstract-Conference.html.

Cai, J., Song, Y., Wu, J., and Chen, X. (2024). Voice disorder classification using wav2vec 2.0 feature extraction. *J. Voice* Available online at: https://www.sciencedirect.com/science/article/pii/S0892199724002935.

Chandra, R., Goyal, S., and Gupta, R. (2021). Evaluation of deep learning models for multi-step ahead time series prediction. *IEEE Access*. doi: 10.1109/ACCESS.2021.3085085

Chen, Z., Fang, Y., Yin, J., Lv, S., Sheikh Muhammad, F., and Liu, L. (2024). A novel lightweight yolov8-pss model for obstacle detection on the path of unmanned agricultural vehicles. *Front. Plant Sci.* 15, 1509746. doi: 10.3389/fpls.2024.1509746

Dudukcu, H. V., Taskiran, M., Taskiran, Z. G. C., and Yıldırım, T. (2022). Temporal convolutional networks with rnn approach for chaotic time series prediction. *Appl. Soft Computing* Available online at: https://www.sciencedirect.com/science/article/pii/S1568494622009942.

Durairaj, D. M., and Mohan, B. G. K. (2022). A convolutional neural network based approach to financial time series prediction. *Neural computing Appl*. doi: 10.1007/s00521-022-07143-2

Fan, J., Zhang, K., Yipan, H., Zhu, Y., and Chen, B. (2021). Parallel spatio-temporal attention-based tcn for multivariate time series prediction. *Neural computing Appl*. doi: 10.1007/s00521-021-05958-z

Fu, X., Ma, Q., Yang, F., Zhang, C., Zhao, X., Chang, F., et al. (2024). Crop pest image recognition based on the improved vit method. *Inf. Process. Agric.* 11, 249–259. doi: 10.1016/j.inpa.2023.02.007

Guan, B., Zhu, X., and Yuan, S. (2024). A t5-based interpretable reading comprehension model with more accurate evidence training. *Inf. Process. Manage.* 61, 103584. doi: 10.1016/j.ipm.2023.103584

Hou, M., Xu, C., Li, Z., Liu, Y., Liu, W., Chen, E., et al. (2022). Multi-granularity residual learning with confidence estimation for time series prediction. *Web Conf*. doi: 10.1145/3487553

Hu, J., Wang, X., Zhang, Y., Zhang, D., Zhang, M., and nan Xue, J. (2020). Time series prediction method based on variant lstm recurrent neural network. *Neural Process. Lett*. doi: 10.1007/s11063-020-10319-3

Kang, H., Yang, S.-H., Huang, J., and Oh, J. (2020). Time series prediction of wastewater flow rate by bidirectional lstm deep learning. *Int. J. Control Automation Syst*. doi: 10.1007/s12555-019-0984-6

Karevan, Z., and Suykens, J. (2020). Transductive lstm for time-series prediction: An application to weather forecasting. *Neural Networks*. doi: 10.1016/j.neunet.2019.12.030

Kim, T., and King, B. R. (2020). Time series prediction using deep echo state networks. *Neural computing Appl*. doi: 10.1007/s00521-020-04948-x

Li, Y., Wu, K., and Liu, J. (2023). Self-paced arima for robust time series prediction. *Knowledge-Based Syst*. doi: 10.1016/j.knosys.2023.110489

Lindemann, B., Müller, T., Vietz, H., Jazdi, N., and Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. *Proc. CIRP*. doi: 10.1016/j.procir.2021.03.088

Lu, X., Zhang, J., Yang, R., Yang, Q., Chen, M., Xu, H., et al. (2024). Effective variance attention-enhanced diffusion model for crop field aerial image super resolution. *ISPRS J. Photogrammetry Remote Sens.* 218, 50–68. doi: 10.1016/j.isprsjprs.2024.08.017

Morid, M., Sheng, O. R., and Dunbar, J. A. (2021). Time series prediction using deep learning methods in healthcare. *ACM Trans. Manage. Inf. Syst.* doi: 10.1145/3531326

Moskolaï, W., Abdou, W., Dipanda, A., and Kolyang, (2021). Application of deep learning architectures for satellite image time series prediction: A review. *Remote Sens*. doi: 10.3390/rs13234822

Net, F., Folia, M., Casals, P., Bagdanov, A. D., and Gómez, L. (2025). Eufcc-340k: A faceted hierarchical dataset for metadata annotation in glam collections. *Multimedia Tools Appl.*, 1–24. doi: 10.1007/s11042-024-20561-9

Ng, D. H. L., Chia, T. R. T., Young, B. E., Sadarangani, S., Puah, S. H., Low, J. G. H., et al. (2024). Study protocol: infectious diseases consortium (i3d) for study on integrated and innovative approaches for management of respiratory infections: respiratory infections research and outcome study (respiro). *BMC Infect. Dis.* 24, 123. doi: 10.1186/s12879-023-08795-8

Ni, J., Dong, J., Zhang, J., Pang, F., Cao, W., and Zhu, Y. (2016). The spectral calibration method for a crop nitrogen sensor. *Sensor Rev.* 36, 48–56. doi: 10.1108/SR-04-2015-0051

Ni, J., Yao, L., Zhang, J., Cao, W., Zhu, Y., and Tai, X. (2017). Development of an unmanned aerial vehicle-borne crop-growth monitoring system. *Sensors* 17, 502. doi: 10.3390/s17030502

Ni, J., Zhang, J., Wu, R., Pang, F., and Zhu, Y. (2018). Development of an apparatus for crop-growth monitoring and diagnosis. *Sensors* 18, 3129. doi: 10.3390/s18093129

Ruan, L., Bai, Y., Li, S., He, S., and Xiao, L. (2021). Workload time series prediction in storage systems: a deep learning based approach. *Cluster Computing*. doi: 10.1007/s10586-020-03214-y

Shen, L., and Kwok, J. (2023). Non-autoregressive conditional diffusion models for time series prediction. *Int. Conf. Mach. Learn.* Available online at: https://proceedings.mlr.press/v202/shen23d.html

Wang, J., Jiang, W., Li, Z., and Lu, Y. (2021a). A new multi-scale sliding window lstm framework (mssw-lstm): A case study for gnss time-series prediction. *Remote Sens.* doi: 10.3390/rs13163328

Wang, J., Peng, Z., Wang, X., Li, C., and Wu, J. (2021b). Deep fuzzy cognitive maps for interpretable multivariate time series prediction. *IEEE Trans. fuzzy Syst*. doi: 10.1109/TFUZZ.2020.3005293

Wang, Z., Wang, R., Wang, M., Lai, T., and Zhang, M. (2024). "Self-supervised transformer-based pre-training method with general plant infection dataset," in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*. 189–202 (Springer). doi: 10.1007/978-981-97-8490-5_14

Wen, J., Yang, J., Jiang, B., Song, H., and Wang, H. (2021). Big data driven marine environment information forecasting: A time series prediction network. *IEEE Trans. fuzzy Syst*. doi: 10.1109/TFUZZ.91

Widiputra, H., Mailangkay, A., and Gautama, E. (2021). Multivariate cnn-lstm model for multiple parallel financial time-series prediction. *Complex*. doi: 10.1155/2021/9903518

Wu, D., Wang, X., Su, J., Tang, B., and Wu, S. (2020). A labeling method for financial time series prediction based on trends. *Entropy*. doi: 10.3390/e22101162

Xiao, Y., Yin, H., Zhang, Y., Qi, H., Zhang, Y., and Liu, Z. (2021). A dual-stage attention-based conv-lstm network for spatio-temporal correlation and multivariate time series prediction. *Int. J. Intelligent Syst*. doi: 10.1002/int.22370

Xu, M., Han, M., Chen, C. L. P., and Qiu, T. (2020). Recurrent broad learning systems for time series prediction. *IEEE Trans. Cybernetics*. doi: 10.1109/TCYB.6221036

Yan, B., Cui, Z., Deng, G., Li, G., Zheng, S., He, F., et al. (2024). Design and validation of a real-time cassava planter seed quality monitoring system based on optical fiber sensors and rotary encoders. *Front. Plant Sci.* 15, 1481909. doi: 10.3389/fpls.2024.1481909

Yang, M., and Wang, J. (2021). Adaptability of financial time series prediction based on bilstm. *Int. Conf. Inf. Technol. Quantitative Manage*. Available online at: https://www.sciencedirect.com/science/article/pii/S1877050922000035

Yin, L., Wang, L., Li, T., Lu, S., Tian, J., Yin, Z., et al. (2023). U-net-lstm: Time series-enhanced lake boundary prediction model. *Land*. doi: 10.3390/land12101859

Yu, C., Wang, F., Shao, Z., Sun, T., Wu, L., and Xu, Y. (2023). "Dsformer: A double sampling transformer for multivariate time series long-term prediction," in *International Conference on Information and Knowledge Management*.

Yu, J., Xu, X., Duan, J., Jiang, Y., Yuan, H., Liang, H., et al. (2025). Effect of operational parameters on droplet deposition characteristics using an unmanned aerial vehicle for banana canopy. *Front. Plant Sci.* 15, 1491397. doi: 10.3389/fpls.2024.1491397

Zhang, B., Zhang, P., Dong, X., Zang, Y., and Wang, J. (2024a). "Long-clip: Unlocking the long-text capability of clip," in *European Conference on Computer Vision*. 310–325 (Springer). doi: 10.1007/978-3-031-72983-6_18

Zhang, B., Zhu, X., Gao, L., Zhang, X., and Lei, Z. (2024b). "Blip-adapter: Bridging vision-language models with adapters for generalizable face anti-spoofing," in *2024 IEEE International Joint Conference on Biometrics (IJCB)*. 1–11 (IEEE). Available online at: https://ieeexplore.ieee.org/abstract/document/10744465/

Zheng, W., and Chen, G. (2021). An accurate gru-based power time-series prediction approach with selective state updating and stochastic optimization. *IEEE Trans. Cybernetics*. Available online at: https://ieeexplore.ieee.org/abstract/document/9600449/

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., et al. (2020). Informer: Beyond efficient transformer for long sequence time-series forecasting. *AAAI Conf. Artif. Intell.* Available at: http://ojs.aaai.org/index.php/AAAI/article/view/17325

frontiers | Frontiers in Plant Science

TYPE Systematic Review
PUBLISHED 02 September 2025
DOI 10.3389/fpls.2025.1579355

Check for updates

OPEN ACCESS

EDITED BY
Lei Shu,
Nanjing Agricultural University, China

REVIEWED BY
Orly Enrique Apolo-Apolo,
KU Leuven, Belgium
Zhanhao Shi,
Shandong Agriculture and Engineering University, China

*CORRESPONDENCE
Hongyan Zhu
hyzhu-zju@foxmail.com
Junfeng Gao
junfeng.gao@abdn.ac.uk

†These authors have contributed equally to this work and share first authorship

RECEIVED 19 February 2025
ACCEPTED 28 July 2025
PUBLISHED 02 September 2025

CITATION
Zhu H, Qin S, Su M, Lin C, Li A and Gao J (2025) Harnessing large vision and language models in agriculture: a review.
Front. Plant Sci. 16:1579355.
doi: 10.3389/fpls.2025.1579355

COPYRIGHT
© 2025 Zhu, Qin, Su, Lin, Li and Gao. 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.

Harnessing large vision and language models in agriculture: a review

Hongyan Zhu[1,2*†], Shuai Qin[1,2†], Min Su[1,2], Chengzhi Lin[1,2], Anjie Li[1,2] and Junfeng Gao[3*]

[1]Guangxi Key Laboratory of Brain-inspired Computing and Intelligent Chips, School of Electronic and Information Engineering, Guangxi Normal University, Guilin, China, [2]Key Laboratory of Integrated Circuits and Microsystems (Guangxi Normal University), Education Department of Guangxi Zhuang Autonomous Region, Guilin, China, [3]Department of Computer Science, University of Aberdeen, Aberdeen, United Kingdom

Introduction: Agriculture is a cornerstone of human society but faces significant challenges, including pests, diseases, and the need for increased production efficiency. Large models, encompassing large language models, large vision models, and multimodal large language models, have shown transformative potential in various domains. This review aims to explore the potential applications of these models in agriculture to address existing problems and improve production.

Methods: We conduct a systematic review of the development trajectories and key capabilities of large models. A bibliometric analysis of literature from Web of Science and arXiv is performed to quantify the current research focus and identify the gap between the potential and the application of large models in the agricultural sector.

Results: Our analysis confirms that agriculture is an emerging but currently underrepresented field for large model research. Nevertheless, we identify and categorize promising applications, including tailored models for agricultural question-answering, robotic automation, and advanced image analysis from remote sensing and spectral data. These applications demonstrate significant potential to solve complex, nuanced agricultural tasks.

Discussion: This review culminates in a pragmatic framework to guide the choice between large and traditional models, balancing data availability against deployment constraints. We also highlight critical challenges, including data acquisition, infrastructure barriers, and the significant ethical considerations for responsible deployment. We conclude that while tailored large models are poised to greatly enhance agricultural efficiency and yield, realizing this future requires a concerted effort to overcome the existing technical, infrastructural, and ethical hurdles.

KEYWORDS

large model, agriculture, natural language processing, computer vision, multimodal model

1 Introduction

The significance of agriculture in the global economy is increasing steadily, and there is growing awareness regarding its sustainability. Ahirwar et al. (2019) believe that it is necessary to increase global agricultural food production by a minimum of 70% to meet the needs of the increasing world population. Unfortunately, there are many factors in agriculture that make it difficult to steadily increase grain production, including 1): crop diseases caused by pathogens such as bacteria, fungi, and viruses. These diseases can spread rapidly, often leading to devastating effects on entire crops. For instance, bacterial blight in rice and late blight in potatoes can wipe out significant portions of harvests. The economic impact is staggering, as farmers face not only reduced yields but also increased costs associated with disease management; 2): poor seed quality can lead to weak plant growth, reduced yields, and greater susceptibility to both diseases and pests. Farmers who use low-quality seeds often experience crop failures, which not only jeopardizes their income but also contributes to broader food insecurity within communities. Transitioning to certified, high-quality seeds is essential for improving crop resilience and productivity; 3): many agricultural tasks remain inefficient and labor-intensive, hindering productivity. Traditional methods of weeding, planting, watering, and harvesting are often time-consuming and can lead to resource wastage. For example, manual weeding not only consumes labor but may also fail to effectively control weed populations, resulting in reduced crop yields. The adoption of mechanization and modern farming techniques, such as precision agriculture (PA), can significantly improve efficiency.

PA is an agricultural management approach that utilizes modern technology to enhance production efficiency and sustainability. It encompasses sensor technology, unmanned aerial/ground vehicles (UAVs/UGVs), remote sensing technology, automation equipment, big data, machine learning (ML), and deep learning (DL) (Tokekar et al., 2016; Khanal et al., 2020; Saleem et al., 2021). This enables farmers to reduce production costs and improve decision-making capabilities, providing significant economic and social benefits. For crop diseases, traditional detection methods like polymerase chain reactions based on unique deoxyribonucleic acid sequences of pathogens, enzyme-linked immunosorbent assays on the basis of pathogens proteins and hyperspectral imaging, are constrained by their operational complexity and the requirement for bulky instruments (Yao et al., 2024). For selecting high-quality seeds, quality assurance programs employ various ways to attest seed quality attributes, including germination and vigor tests (ElMasry et al., 2019). But these methods have limitations in terms of time overhead, subjectivity, and the destructive nature of assessing seed quality (Medeiros et al., 2020). For a general tasks in agriculture, the use of pesticides for weed control may have negative impacts on the environment, and Phytotoxicity reactions can lead to diminished crop quality and reduced yields (Visentin et al., 2023). And the traditional solutions to these tasks are also inefficient due to these manned implements are dreadfully slow. On the other hand, driven by growing health consciousness, the public has long been worried about the safety and quality of food, which is linked to agricultural products. Reducing food losses and improving food safety rely significantly on the continuous monitoring of crop quality, especially the inspection of diseases during crop growth stage (Karthikeyan et al., 2020).

DL technologies in PA can effectively address the limitations of traditional methods by leveraging their powerful data processing and pattern recognition capabilities. For instance, DL can analyze vast amounts of data from sensors, drones, and satellite imagery to accurately identify crop health, soil characteristics, and potential diseases and pests (Nasir et al., 2021; Bouguettaya et al., 2022). This application enables farmers to obtain real-time insights, allowing for more scientifically informed management decisions, optimized resource use, and increased crop yields. However, DL technologies also have their limitations, primarily due to the high demand for model training (Sun et al., 2017). DL models typically require large amounts of labelled data to train and often need to be retrained when faced with new agricultural environments or crop varieties (Thenmozhi and Reddy, 2019). This repetitive training process is not only time-consuming but also requires significant computational resources and expertise. The effectiveness of transfer learning lies in its ability to apply models trained in one domain to a related domain, thus reducing the need for new datasets (Bosilj et al., 2020; Paymode and Malode, 2022). However, the diversity and complexity of agricultural environments can limit the effectiveness of transfer learning (Raffel et al., 2020). For example, differences in soil conditions, climate variations, and crop growth characteristics across regions can result in models trained in one area performing poorly in another. Therefore, although DL holds tremendous potential in PA, its adaptability and generalizability must be carefully considered to ensure that models remain effective in the ever-evolving agricultural field.

Large models are fundamentally distinguished from conventional DL models by their vast parameter counts (often billions) and extensive pre-training on massive, diverse datasets. By being exposed to a rich array of information, these models can better understand and adapt to various contexts, making them highly versatile tools in fields such as natural language processing (NLP), computer vision (CV), and decision-making (Kung et al., 2023). Crucially, unlike traditional DL models, large models develop "emergent abilities"—such as few-shot/zero-shot learning, complex reasoning, and strong generative abilities—that are not simply scaled-up versions of prior performance (Bommasani et al., 2021; Zhao et al., 2023c). As an efficient analytical means, large model, has found extensive application in the agricultural sector (Stella et al., 2023; Yang et al., 2023c). They have demonstrated excellent performance in analyzing agricultural data, pest and disease management, PA, and more. However, they still face many problems such as difficulty in obtaining agricultural data (Lu and Young, 2020), low model training efficiency, distribution shift (Chawla et al., 2021), and plant blindness (Geitmann and Bidhendi, 2023). In response to the challenges faced by traditional agriculture, we committed to conducting a comprehensive analysis of large models. First, we systematically summarized the history of large models, their applications in other fields, and their significance for agriculture. Subsequently, we introduced many applications of

large models in agriculture. Furthermore, recognizing that large models were a relatively new technological approach, we outlined some solutions from ethical and responsibility perspectives. Finally, we summarized the current challenges and future directions of large models and drew conclusions on their effectiveness in the agricultural domain.

2 Feasibility analysis of large models in agriculture

Artificial intelligence (AI), whose main purpose is to establish systems that learn and think like human (Holzinger et al., 2019), just like human language and visual abilities. At present, research on large models is also focused on NLP and CV. Next, large language model (LLM), large vision model (LVM) and multimodal large language model (MLLM) will be introduced in detail.

2.1 Evolution and key milestones of large models

2.1.1 Development trajectories of large language models

LLM is a model based on NLP with a vast number of parameters (typically billions) trained on massive datasets of text and code, and we can divide the development of it into four stages (Figure 1):

1. Statistical Language Models (SLM): SLMs use traditional statistical methods (like n-grams) to learn word probabilities. Their effectiveness relies on the amount of data and estimation algorithms (Chelba et al., 2013). While SLMs are widely used in NLP, they have three main drawbacks: Scalability: Larger n requires more memory and parameters (n represents how many preceding words the model considers when predicting the next word); Information sharing: N-grams can't share semantic information across similar words; Data sparsity: Techniques like data smoothing can help, but neural networks handle this better.
2. Neural Language Models (NLM): NLMs utilize various neural networks and are more effective than SLMs (Bengio et al., 2000; Mikolov et al., 2010; Sundermeyer et al., 2012). They address data sparsity using feedforward and recurrent neural networks (RNNs), which learn features automatically. Key developments include: Feedforward neural networks (FFNNLM): Proposed by Bengio et al. in 2003, they learn distributed word representations (Bengio et al., 2000); RNN Language Model (RNNLM): Introduced by Mikolov et al., but struggles with long-term dependencies (Mikolov et al., 2010). Long short-term memory (LSTM) networks were later added to overcome this (Sundermeyer et al., 2012).
3. Pre-trained Language Models (PLM): PLMs are categorized into feature-based and fine-tuning methods: Feature-based:

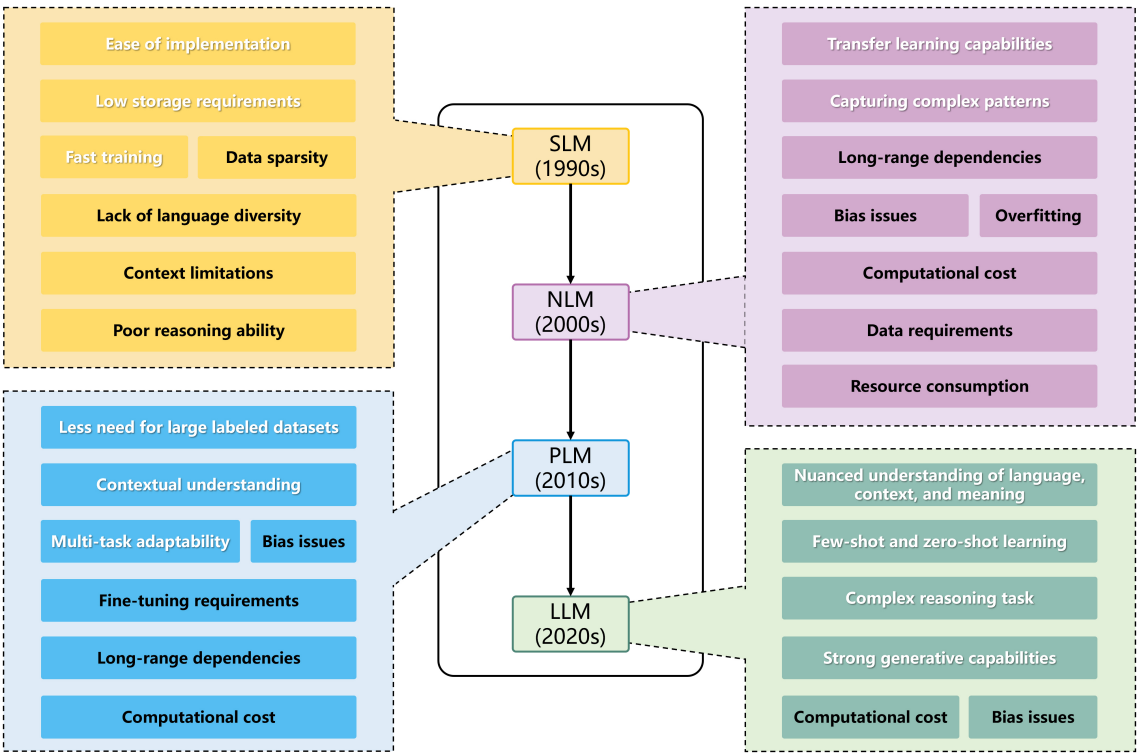


FIGURE 1
Development timeline of NLP models and their pros and cons. White characters represent advantages; black characters represent disadvantages.

Extracts features from large datasets (e.g., ElMo); Fine-tuning: Transfers entire model parameters to specific tasks, exemplified by BERT and GPT. Transformers, introduced by Google, employ a self-attention mechanism, facilitating better training and performance (Vaswani et al., 2017), GPT is fine-tuned from the Transformer. Due to the significant acceleration of model training by Transformer, it has gradually become the fundamental architecture for LLMs.

4. Large Language Models (LLM): LLMs have billions of parameters and exhibit unique capabilities, known as "emergent abilities". Research shows that larger models perform better and are more sample-efficient. For instance, GPT-3 can generate expected outputs from input sequences without additional training, a feat beyond smaller models like GPT-2.

The transition from SLM to LLM signifies a progressive increase in model complexity, data handling abilities, and adaptability to tasks. Each new generation improves upon its predecessor to overcome limitations, fostering advancement in natural language processing. As shown in Figure 1, compared to other models, LLMs have a comprehensive understanding of language and excel at complex reasoning. Their strong few-shot, zero-shot, and generative capabilities allow them to adapt to new tasks with minimal examples. However, high computational costs and bias issues prevent them from being perfect. The high computational cost remains an unresolved challenge in today's era of large data training. Bias issues can be mitigated through a series of review and regulatory measures, which will be detailed in section 4.

2.1.2 Key advancements and capabilities of large vision models

LVMs are a new generation of models associated with CV, characterized by their immense scale and broad pre-training. Initially, LVM might have denoted purely vision-based models trained solely on image data. However, inspired by multimodal learning in LLM, the concept has evolved to include large models trained on both images and text, enabling rich cross-modal associations. CV models began their development in the 20th century and have continued to evolve significantly to the present day (Figure 2). Fueled by the availability of massive image datasets, the development of powerful DL architectures, and significant progress in large-scale pre-training techniques, LVMs have become one of the major development trends in CV models in recent years.

The research on CV models initially focused on shallow image feature extraction algorithms, including scale-invariant feature transform, histogram of oriented gradient, and other methods, but had significant limitations. In 2012, AlexNet (Krizhevsky et al., 2017) achieved a breakthrough success in ImageNet large scale visual recognition challenge, sparking a wave of convolutional neural networks (CNN) for vision models. With the development of DL, deep residual networks including VGGNet (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), and ResNet (He et al., 2016) were successively proposed, which improved the performance of image classification, object detection, semantic segmentation, etc. The boom of the Internet also enabled large-scale image datasets to be used for training vision models. Faster R-CNN (Ren et al., 2016), YOLO (Redmon et al., 2016), Mask R-CNN (He et al., 2017) emerged one after another.

In recent years, Transformer has been successfully applied in the domain of LVM, leading to the emergence of models like Vision Transformer (ViT) (Dosovitskiy et al., 2020) and DALL-E (Ramesh et al., 2021) which have garnered significant public attention. Unlike the traditional DL models mentioned above, LVMs such as ViT leverage transformer architectures and are typically pre-trained on significantly larger and more diverse datasets (e.g., billions of images). This foundational pre-training enables them to develop a more generalized understanding of visual concepts and emergent capabilities, allowing for superior performance on a wide range of downstream tasks, often with only limited domain-specific data.

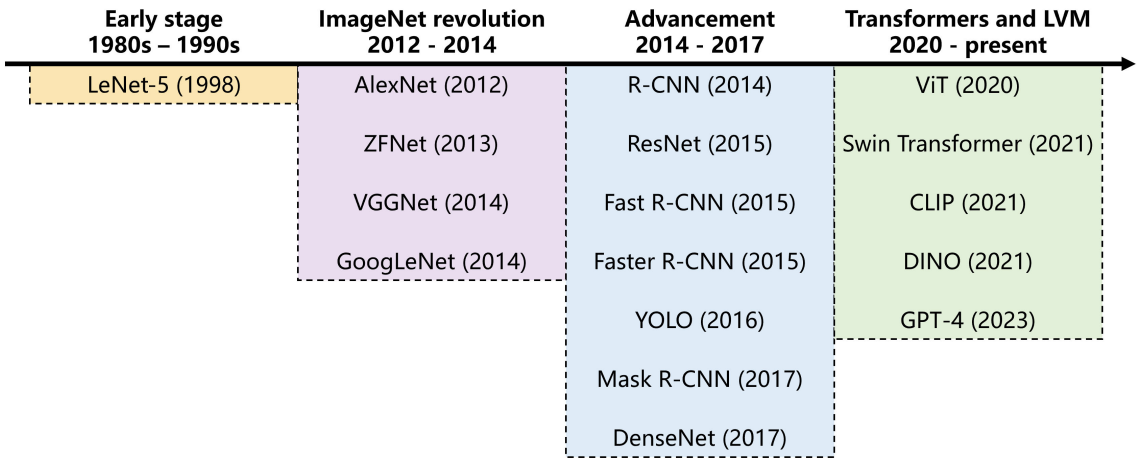


FIGURE 2
Development timeline of computer vision models.

Their ability to grasp complex visual patterns and adapt to new conditions makes them highly versatile.

As detailed in Table 1, large vision models (LVMs) and traditional vision models differ significantly in their core architecture, data requirements, and capabilities. The fundamental distinction lies in their approach to context: Transformer-based LVMs leverage global self-attention to capture broad visual context and long-range dependencies, a significant leap from the local receptive fields of traditional convolutional models (Aymen et al., 2024; Malla et al., 2024). While this architectural shift grants LVMs superior generalization, it also introduces challenges like high computational demands and data hunger. Notably, research is actively addressing these limitations. For instance, Shi et al. (2024) proposed "Scaling on Scales (S^2)," which enhances performance by increasing image scales rather than model size, providing new insights for the future development of vision models.

2.1.3 The emergence of multimodal large language models

In addition to the LLMs and LVMs introduced above, MLLMs are also a research focus in the domain of AI. While LLMs perform well in text-based tasks, their capabilities alone cannot effectively reason about information presented in non-textual formats. Although LVMs perform well in the field of CV and possess some NLP abilities, researchers are not content with large models solely trained on text and images. MLLMs (Wu et al., 2023a) integrate multiple data types, such as images, text, language, audio, and more. It not only possesses the advantages of LLMs and LVMs, but also address the limitations of LLMs and LVMs by integrating multiple modalities, enabling a more comprehensive understanding of various data. It can be said that the developments in MLLMs have set up new avenues for AI, which make binary machines to understand and then process various data types (Wu et al., 2023a). For agriculture, MLLM allows tasks to no longer be confined to just images or text; instead, it can leverage both, and even utilize multimodal inputs like audio and video, breaking the limitations of images and text.

TABLE 1 Comparison of large vision models and traditional vision models.

Feature/ Aspect	Large vision models	Traditional vision models
Core architecture	Transformer-based, global self-attention	Convolution-based, local receptive fields
Context & dependencies	Global context, excels at long-range dependencies	Local focus, struggles with long-range dependencies
Image handling	Processes image patches; more robust to variations	Uses sliding filters; sensitive to some variations
Data needs	Best with large-scale pre-training	Can work with smaller datasets, benefits from pre-training
Multimodal ability	Stronger, more inherent multimodal integration	Requires more specialized designs for multimodal
Parallelism	High (sequence processing)	Good for convolutions, poor for sequential tasks
Key advantages	Global understanding, long dependencies, scalability	Efficient local feature extraction
Key limitations	Can be data-hungry for pre-training	Limited global view, adversarial vulnerability

2.2 Current applications of large models in other domains

As shown in Table 2, many LLMs are designed to develop chatbots (BLOOM (Le Scao et al., 2023), PaLM2 (Anil et al., 2023), ERNIE 4.0) or complete NLP tasks, including text classification, machine translation, and sentiment analysis [OPT (Zhang et al., 2022b)]. Similarly, LVMs are primarily engineered to interpret and process visual information. They excel at core CV tasks such as image classification, object detection, segmentation, and image generation, often forming the foundation for systems needing to understand or interact with the visual world. Models like InternImage (Wang et al., 2023) and LLaVA (Liu et al., 2023b) represent efforts to enhance performance on complex visual analysis tasks, aiming to simulate and automate human visual processes.

Although LLM and LVM satisfies some functions and takes large models a big step towards artificial general intelligence (AGI), it is not enough to achieve the goal that machines can emulate human thinking and carry out a wide range of general tasks through transfer learning and diverse other modalities without achieving the multimodality of the model (Zhao et al., 2023b). Some large models have implemented multimodality, enabling them to analyze different types of information [GPT-4 (Bubeck et al., 2023), LLaMA, Gemini (Team et al., 2023), ImageBind (Girdhar et al., 2023)] and interact with users. It is worth mentioning that most of the newer large models are MLLMs, and many models that were originally LLMs or LVMs have gradually acquired multimodal capabilities after multiple updates.

However, many current models are generic models and their training datasets are too broad, they cannot provide a satisfactory answer to knowledge in certain professional fields. As Goertzel (2014) believed, for a system to be considered AGI, it is not necessary for it to have infinite generality, adaptability, or flexibility. Therefore, some researchers have optimized and adjusted existing large models and have released some large models specifically for a single field. BloombergGPT can be used in the financial field, showcasing remarkable performance on general LLM benchmarks and surpassing comparable models on financial tasks (Wu et al., 2023c). The meteorological model in panguLM developed by Huawei can provide predictions of variables such as gravity potential, humidity, wind speed, temperature, and pressure within 1 hour to 7 days. Embedding PaLM-E into robots can achieve multiple specific tasks, like visual question answering, sequential robotic manipulation planning, and captioning (Driess et al., 2023). OceanGPT is an expert in various marine science tasks (Bi et al., 2023). It exhibits not only a higher level of knowledge expertise for oceans science tasks but also acquires preliminary embodied intelligence capabilities in ocean engineering. PMC-

TABLE 2 The currently popular and representative large models.

Original version	Latest version	Release date (original)	Release date (latest)	Types (original → latest)	Information	References
OPT	/	May 2rd, 2022	/	LLM	OPT promotes transparency, reproducibility, and broader community engagement and innovation in NLP research. (open source)	(Zhang et al., 2022b)
BLOOM	BLOOMZ	July 12th, 2022	December 15th, 2022	LLM	A decoder only model based on Transformer architecture. (open source)	(Le Scao et al., 2023)
PMC-LLaMA	/	Aprill 27th, 2023	/	LLM	Inject medical knowledge into existing LLM using 4.8 million biomedical academic papers. (open source)	(Wu et al., 2024)
PaLM2	/	May 11st, 2023	/	LLM	PaLM2 was a neural network-based language model that was considered one of the most advanced language models available at the time of its release in May 2023.	(Anil et al., 2023)
BloombergGPT	/	March 30th, 2023	/	LLM	A LLM for the financial field.	(Wu et al., 2023c)
OceanGPT-Basic-7B	OceanGPT-Basic-14B/7B/2B	October 3rd, 2023	July 4th, 2024	LLM	OceanGPT is the first LLM in the ocean domain. (open source)	(Bi et al., 2023)
DeepSeek LLM	DeepSeek-R1	November 29th, 2023	January 20th, 2025	LLM	DeepSeeke-R1 excels in complex tasks such as mathematics, coding, and natural language reasoning	(Guo et al., 2025a)
Llama 2	Llama 4	July 20th, 2023	April 6th, 2025	LLM → MLLM	A series of large models released by Meta.	/
Qwen-7B	Qwen2.5-Omni-7B	August 3rd, 2023	March 27th, 2025	LLM → MLLM	A super large model launched by Alibaba Cloud. (open source)	(Bai et al., 2023)
Kimi Chat	Kimi k1.5	October 10th, 2023	January 20th, 2025	LLM → MLLM	Kimi k1.5 surpasses GPT-4o by 550% in mathematics, coding, and other capabilities under short-chain thinking mode.	/
Gemma	Gemma 3	February 21st, 2024	March 12th, 2025	LLM → MLLM	Gemma 3 is a MLLM released by Google. (open source)	(Team et al., 2024, 2025)
PaLM-E	/	March 6th, 2023	/	LVM	PaLM-E can integrate vision and language into robot control.	(Driess et al., 2023)
InternImage	InternImage-H	November 10th, 2022	October 4th, 2023	LVM	A LVM based on deformable convolution. (open source)	(Wang et al., 2023)
PanguCVLM 3.0	PanguCVLM 5.0	July 7th, 2023	June 21th, 2024	LVM	A LVM that simulates and automates human visual processes.	/
LLaVA	LLaVA-NeXT (Stronger)	April 17th, 2023	May 10th, 2024	LVM	LLaVA has the ability to align and fuse the visual information of images with the semantic information of text. (open source)	(Liu et al., 2023b)
mPLUG-Owl	mPLUG-Owl3	April 27th, 2023	August 20th, 2024	LVM → MLLM	mPLUG-Owl is developed by Alibaba DAMO Academy. (open source)	(Ye et al., 2023, 2024)
SPARK 1.0	SPARK 4.0 Turbo	May 6th, 2023	October 24th, 2024	LVM → MLLM	A new generation of cognitive intelligence model with Chinese as its core.	/
Claude 3	Claude 3.7 Max	March 4th, 2024	March 18th, 2025	MLLM	A MLLM that primarily focuses on code processing.	/
ERNIE 4.0	ERNIE 4.5	October 17th, 2023	March 16th, 2025	MLLM	ERNIE is a new generation of Baidu's large model for knowledge enhancement.	/
ImageBind	/	May 9th, 2023	/	MLLM	ImageBind is the first AI model that can bind information from six modes.	(Girdhar et al., 2023)
GPT-4	GPT-4.5	March 14th, 2023	February 28th, 2025	MLLM	GPT-4.5 significantly enhances its knowledge reserves and emotional intelligence by expanding unsupervised learning and reasoning capabilities.	(Bubeck et al., 2023)

(Continued)

TABLE 2 Continued

Original version	Latest version	Release date (original)	Release date (latest)	Types (original → latest)	Information	References
Skywork	Skywork 4.0	April 17th, 2023	January 6th, 2025	MLLM	Skywork is a series of large models developed by the Kunlun · Skywork team.	(Wei et al., 2023)
Gemini	Gemini 2.5	December 6th, 2023	March 26th, 2025	MLLM	Gemini is a MLLM launched by Google DeepMind.	(Team et al., 2023)
Sora	/	February 15th, 2024	/	MLLM	Sora can create realistic and imaginative scenes from text instructions.	(Peebles and Xie, 2023)
Hunyuan-t1	Hunyuan-t1 (official)	February 17th, 2025	March 21st, 2025	MLLM	Hunyuan-t1 is a deep-thinking model independently developed by Tencent.	/

The arrow '→' represents the change in the model's category, from its original type to the type of its latest version.

LLaMA represents the pioneering open-source medical specific language model that demonstrates exceptional performance on diverse medical benchmarks, outperforming ChatGPT and LLaMA-2 with significantly fewer parameters (Wu et al., 2024). The success of large models across diverse fields, as highlighted in this section, underscores their potential to generalize and tackle complex problems, suggesting their applicability to the intricate tasks within agriculture.

2.3 Assessing the attention to large models within agriculture

In the past few decades, the advancement of agricultural technology has significantly improved global agricultural production efficiency. According to the forecast released by the food and agriculture organization (FAO) of the United Nations, the global grain production in 2023 was 2.84 billion tons, nearly twice that of the early 20th century. Although global agricultural production efficiency is high, the world population is also constantly growing. Continuously improving agricultural production efficiency is the lifeline of economic development and the foundation for ensuring human food, clothing, and survival needs. Hence, how to make agricultural practices advance is a crucial issue. Next, we will use bibliometric methods in conjunction with practical analysis to explain why large models are important for agriculture.

2.3.1 Bibliometric analysis and data sources

Bibliometrics is a quantitative analysis method that integrates mathematics, statistics, and bibliology, based on mathematical statistics. It focuses on the external characteristics of scientific literature to conduct statistical and visual analyses of the literature (Wang et al., 2019). Keywords encapsulate important information about the research topic. They can intuitively reflect the themes and content of the study, reveal the connections between research contents, results, and characteristics in a particular field, and demonstrate the research dynamics and emerging trends within that area (Li and Zhao, 2015). Our analysis used two methods:

1. Searching for research literature related to large models using the Science Citation Index Expanded (SCI-E) and Social Sciences Citation Index (SSCI) from Web of Science (WoS) with keywords such as "large models", "large language models", "large vision models", or "foundation models", covering the period from 2019 to 2024.
2. Collecting 3,496 papers from the arXiv in the field of artificial intelligence from 2019 to 2024 and categorizing them by discipline based on keyword searches.

Our analysis of WoS aimed to identify the established trends and peer-reviewed research regarding large models, and specifically, the frequency of agriculture-related keywords within this body of published work. This provides a view of the validated research landscape. Complementarily, we included arXiv to capture the more recent and rapidly evolving trends in artificial intelligence research. arXiv, as a leading platform for pre-prints in AI, offers valuable insights into emerging research directions and the early exploration of applying large models across various disciplines, including potential initial interest in agriculture. Pre-prints often precede formal publication, providing a timelier snapshot of the research frontier.

By analysing both published literature (WoS) and pre-prints (arXiv), we aimed to gain insights from two different perspectives: the established, peer-reviewed research landscape and the more immediate, evolving research front. This allows us to observe both the current state of validated research and the potential emerging trends and initial explorations within the field.

2.3.2 Detailed analysis and design protocol

As described in section 2.3.1, two data sources were used for the specific analysis method: (1) WoS; (2) arXiv. Next, we will elaborate on the details of using these two bibliometric analysis methods.

For the Method 1, after entering the official website of WoS, search in "Web of science Core Collection" and select Science Citation Index Expanded (SCI-EXPANDED) and Social Sciences Citation Index (SSCI) in edition. Both SCI-EXPANDED and SSCI primarily index peer-reviewed journals with established reputations within their respective fields. This ensures a certain level of quality control and scholarly rigor in the literature being analysed. Then search for topics with the keywords "large models", "large language models", "large

vision models", or "foundation models", covering articles from 2019 to 2024, and export the authors, titles, sources and abstracts of these articles in plain text file. Finally, import these plain text files into VOSviewer to draw a network of keywords in the field of large models. For the Method 2, we use the keyword "artificial intelligence" to search for articles on arXiv, and crawl the relevant articles from 2019 to 2024, including title, author, abstract and other information, to build a csv file. Then search this file according to relevant keywords. For example, in the medical field, keywords such as medical, healthcare, hospital, etc. are used to filter out relevant articles and count the number. Finally, a graph of the proportion of articles in different fields under the AI domain is constructed.

In this way, we obtained a network map of keyword through the Method 1, and a graph of the proportion of different fields in the AI domain through the Method 2. The specific results and analysis will be explained in the next section.

2.3.3 Analysis results

A total of 1,789 papers were filtered using Method 1, and a network diagram of keyword occurrences was generated using VOSviewer. As shown in Figure 3, the term 'agriculture' appears infrequently in these large model papers, indicating that large models have not received widespread attention in the agricultural field.

The reasons why large models have not received attention in the agricultural sector are diverse. First, large models are a relatively new technology that has emerged in recent years, and many researchers and practitioners in agriculture may not fully understand their capabilities and potential applications in the field. Second, implementing large models often requires substantial computational resources and expertise, which may not be easily accessible in many agricultural environments. Third, agricultural tasks can be very specific and localized, leading people to prefer traditional methods over large models.

Moreover, Figure 4 illustrates a difference: the application and research of large models in agriculture are currently limited compared to other fields. This observation from our bibliometric analysis (Figures 3, 4) suggests that despite the evident potential of large models to address agricultural challenges discussed in the introduction, the field is still in the early stages of exploring and adopting this technology. Therefore, a detailed review of their potential applications, associated challenges, and responsible deployment is crucial to guide future research and accelerate their integration into agriculture.

3 Large models in agricultural applications

As mentioned in the introduction, agriculture faces multiple challenges, including pests and diseases, seed quality, and crop grading. Large models have demonstrated significant potential in addressing these issues, and some researchers have already developed models specifically tailored for the agricultural domain.

3.1 Emerging potential and existing applications of large models in agriculture

3.1.1 Potential and applications

Many large models have emerged, and although they are not yet truly applied in agriculture, their problem-solving capabilities

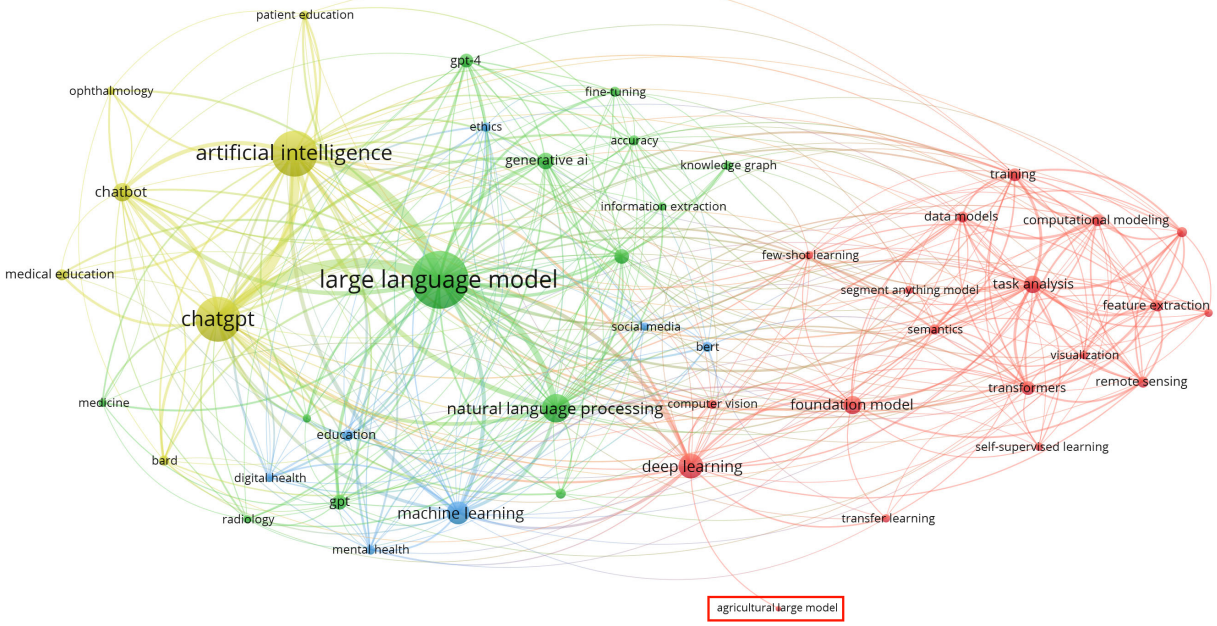


FIGURE 3
47 keywords co-occurrence network map.

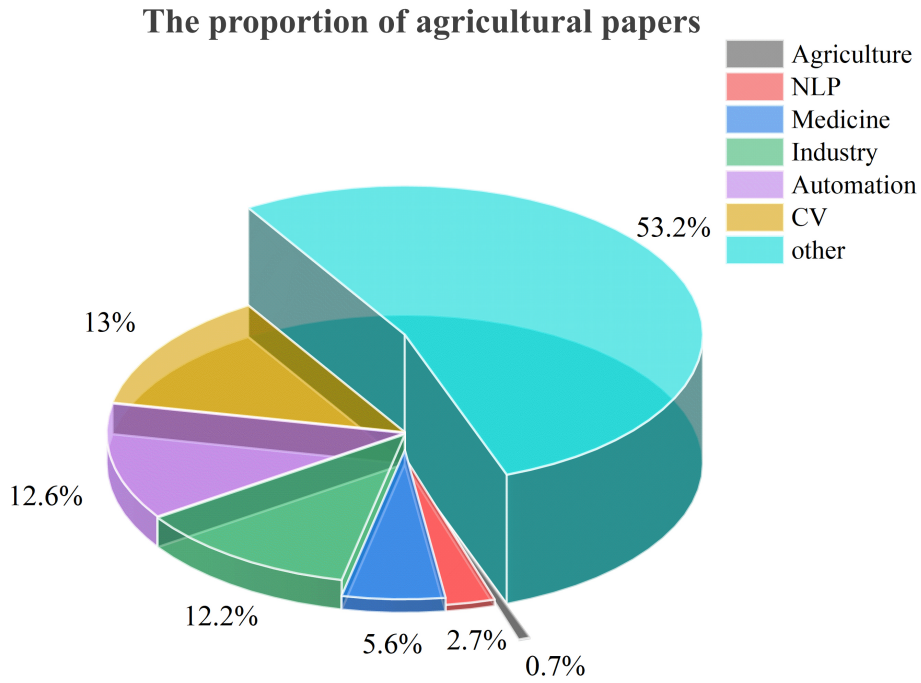


FIGURE 4
The proportion of arXiv papers on agriculture in AI.

TABLE 3 Large models with agricultural potential.

Type	Based	Method	Problem	Application prospect	References
LLM	ChatGPT	GPT-3.5-turbo	Agricultural information extraction	Rapid querying of agricultural information	(Peng et al., 2023)
		AugGPT	Text data augmentation	Few-shot learning for agricultural data	(Dai et al., 2023)
	/	Aurora	Atmospheric prediction	Predicting weather in agriculture	(Bodnar et al., 2024)
LVM	/	MAE, DINO, DINOv2	Plant phenotyping tasks	Monitoring crop health	(Chen et al., 2023a)
	PaLM, ViT	PaLM-E	Robot control	Agricultural intelligent machines or robots	(Driess et al., 2023)
MLLM	SAM	TAM	Object tracking and segmentation in videos	Monitor animals in agricultural farming	(Yang et al., 2023a)

indicate potential prospects in agricultural applications. As shown in Table 3, some large models are modifications of existing models, while others are entirely original. For example, given 50 original descriptions related to "wheat rust," AugGPT can generate 200+ expanded samples covering different growth stages and climatic conditions, thereby enhancing the robustness of disease identification models in complex environments. Aurora is a large model for weather forecasting (Bodnar et al., 2024), and if applied in agriculture, it could enable farmers to schedule activities such as planting, fertilizing, and harvesting based on accurate weather forecasts, as well as proactively mitigate losses from extreme weather events. In addition to ordinary large models, there are also some special existences. HuggingGPT is an AI agent framework designed to orchestrate multiple specialized models, including LLMs like ChatGPT (Shen et al., 2024). It acts as a 'model coordinator,' integrating and managing diverse AI components to enhance decision-making in complex scenarios such as agricultural planning. This capability offers possibilities for managing a series of complex agricultural tasks, from planting to harvesting.

Notably, there are already large models applied in agriculture (Table 4). For instance, TimeGPT demonstrates its capability as a smart agriculture tool (Deforce et al., 2024), being used for predicting soil moisture, which helps farmers determine whether the soil is suitable for certain crops. FMFruit showcases the

importance of large models in agricultural detection tasks (Li et al., 2024), providing new directions and foundations for the development of robotic harvesting systems. ITLMLP performs disease recognition on cucumbers with limited sample sizes, playing a significant role in agricultural automation and intelligence (Cao et al., 2023).

Tables 2, 3 demonstrate the feasibility and importance of large models in agriculture, where many agricultural tasks involve complex reasoning. For example, when presented with an image of a soybean field, agricultural scientists or farmers rely on large models to undertake several key steps. Firstly, the large model must identify any abnormal symptoms evident in the soybean leaves, such as leaf wrinkling. Subsequently, it must ascertain the name of the specific problem that troubles plants, such as soybean mosaic. Next, the model needs to determine the underlying cause of the disease, such as soybean mosaic virus. Finally, it must develop an appropriate treatment strategy. This multi-step, cross modal diagnostic and decision-making process is precisely the unique advantage that large models can demonstrate compared to traditional DL models with a single task.

Many question answering (QA) and dialogue systems are designed to address this type of reasoning problem (Rose Mary et al., 2021; Mostaco et al., 2018; Niranjan et al., 2019). For instance, a chatbot based on a RNN is specifically designed to handle questions related to soil testing, plant protection, and nutrient management (Rose Mary et al., 2021). Although, these QA and dialogue systems and chatbots can answer most inquiries without the need for human interaction and with excellent accuracy, they have limited capabilities for complex problems by reason of their small model size as well as of inadequate training data. Therefore, the agricultural domain requires large models to promote the development of QA and dialogue systems and chatbots. The traditional methods for detecting crop pests and diseases mainly rely on special methods such as serology and molecular biology-based technical means, in addition to artificial visual evaluation. Although these methods can accurately determine pests and diseases to a certain extent, they often require a lot of time and money. And some methods of sampling crops often lead to crop damage, which goes against the original intention of diagnosing diseases and pests to protect crops. Therefore, image processing and analysis is an important task for large models in the field of agriculture, and another important task is to embed LVMs into robots to solve some agricultural problems (Weeding, pruning branches, harvesting, etc.) and achieve automated agriculture.

3.1.2 The advantages of agriculture-specific large models

In the field of agriculture, agriculture-specific models can offer notable advantages over general large models, particularly by effectively integrating diverse, domain-relevant data modalities such as image, text, and crucial label information. This multimodal strategy, often employing techniques like combined contrastive learning methods within a unified feature space, allows these models to address the prevalent challenge of data scarcity in

TABLE 4 Agricultural large models.

Type	Name	Achievement	Significance	References
LLM	TimeGPT	Predicting soil moisture	Contributes to sustainable agricultural practices	(Deforce et al., 2024)
	ChatGPT	Designed a tomato-picking robot	Simplify the design process of agricultural robots	(Stella et al., 2023)
	FMFruit	Identifying multiple types of fruits	Research on robotic harvesting and fruit detection	(Li et al., 2024)
	AgriGPT	Multimodal agricultural knowledge Q&A	Promote precision agriculture practices	(Liu et al., 2023a)
	ShenNong	Development of specialized large models for multiple agricultural domains	Driving agricultural intelligence and comprehensive efficiency improvement	/
	ChatAgri	Cross-linguistic classification of agricultural texts	Provide decision support for precision agriculture	(Zhao et al., 2023a)
LVM	SpectralGPT	Process spectral remote sensing data	Greatly enhanced the processing capability of agricultural spectral data	(Hong et al., 2024)
	SAM	Chicken segmentation and tracking	Facilitates segmentation and tracking tasks in agriculture	(Yang et al., 2023c)
		Agricultural field boundary delineation	Beneficial for PA, crop monitoring, and yield estimation	(Tripathy et al., 2024)
MLLM	ITLMLP	Cucumber disease recognition	Agricultural disease recognition	(Cao et al., 2023)
	AIE-SEG	High-precision segmentation of agricultural imagery	Enables automated field monitoring and yield estimation	(Xu et al., 2023)

agriculture more effectively than models relying solely on single modalities or vast, generic pre-training datasets.

By explicitly learning and leveraging the semantic correlations between visual features (e.g., specific crop disease symptoms) and related textual descriptions or categorical labels, agriculture-specific models can develop more comprehensive, robust, and discriminative representations tailored to the nuances of the field. For example, ChatAgri excels in the specific task of agricultural visual diagnostics (Zhao et al., 2023a). A general MLLM might identify visual anomalies, and lack the specialized knowledge to accurately name the specific agricultural disease or pest, understand its lifecycle, or recommend appropriate, targeted treatments. Especially when faced with limited training samples, agriculture-specific large models may perform better compared to models with poor adaptability. Unlike general large models that often require vast datasets for pre-training and may not adapt well to fine-tuning on limited agricultural data, ITLMLP is designed to be effective with small sample sizes. It extracts richer and more discriminative features from scarce data, leading to significantly higher recognition accuracy (achieving 94.84% in their paper) compared to general large models to the same small dataset (Cao et al., 2023).

Furthermore, their focused training enables them to better identify and weigh agriculturally significant features, accurately discerning subtle but critical patterns for tasks like disease recognition while potentially mitigating the influence of irrelevant background elements, ultimately resulting in improved accuracy, reliability, and greater practical applicability within the complex agricultural environment.

3.2 Leveraging large language models for agricultural data processing, insights, and decision support

LLM can play many roles in the agricultural domain, such as processing and generating agricultural data, providing insights into agricultural production work, and supporting agricultural decision-making for farmers.

3.2.1 Large language models for processing and generating agricultural data

3.2.1.1 Information extraction

LLMs can extract structured information from unstructured agricultural text data. First, the text is divided into individual tokens and LLMs represent each token as a numerical vector called a word embedding. Then, LLMs analyse the surrounding context of each token to understand its meaning within the sentence or document, and identify and categorize named entities within the text, like names of individuals, locations, organizations, or specific agricultural terms. Finally, LLMs employ techniques like information extraction to identify and extract structured information from unstructured text (Involve identifying relationships between entities, extracting key facts, or populating knowledge graphs). LLMs extract information from data using a process known as NLP. Beyond mere extraction, modern LLM applications increasingly employ a paradigm known as retrieval-augmented generation (RAG). In this approach, the LLM first retrieves relevant, up-to-date information from external, domain-specific knowledge bases—such as recent agronomic research, real-time market prices, or local pest outbreak databases. This retrieved context then "augments" the model's input, enabling it to generate responses that are not only more accurate and timelier but also grounded in verifiable sources, thereby significantly mitigating the risks of data lag and factual inaccuracies in the agricultural domain (Gao et al., 2023).

3.2.1.2 Agricultural data generation

Generative AI models are a multimodal LLM, which is the MLLM mentioned above. An obstacle encountered when applying specialized CV algorithms to agricultural vision data is the insufficient availability of training data and labels (Qi et al., 2017; He et al., 2017). In addition, collecting data that encompasses the wide range of variations caused by season and weather changes is exceedingly challenging. Acquiring high-quality data requires a lot of time, and labelling them is even more costly (Zhou et al., 2017). To address these challenges, one approach is to fine-tune multimodal generative LLMs on the target agricultural data domain. This allows the models to generate massive training data and labels, thereby constructing an augmented training set that closely resembles the distribution of the original data (Dai et al., 2023). Besides, text-based generation models can generate images (Rombach et al., 2022) and videos (Ho et al., 2022) of specific scenes based on text descriptions, thereby supplementing training datasets that may lack certain visual content. This helps in expanding the training data and improving the performance of downstream models.

3.2.2 Large language models provide insights

LLMs possess the capability to analyse textual data and uncover trends in agricultural practices, market conditions, consumer preferences, and policy developments. Through analysis of agricultural text data from sources such as news articles, reports, and social media, these models can offer valuable insights into market dynamics and pricing trends (Yang et al., 2024). This provides support for farmers to understand domains outside of agriculture. Many researchers believe that the integration of LLMs into different stages of designment and development for agricultural applications is also experiencing a noticeable rise (Stella et al., 2023; Lu et al., 2023). In (Stella et al., 2023) study, Stella et al. incorporated LLM into the design phase of robotic systems. They specifically focused on designing an optimized robotic gripper for tomato picking and outlined the step-by-step process. In the initial ideation phase, they leveraged LLMs like ChatGPT (Bubeck et al., 2023) to gain insights into the possible challenges and opportunities associated with the task. Building upon this knowledge, they identified the most promising and captivating pathways, engaging in ongoing discussions with the LLM to refine and narrow down the design possibilities. Throughout this process, the human collaborator harnesses the expansive knowledge of the LLM to tap into insights transcend their individual expertise. In the following stage of the design process, which emphasizes technical aspects, the

broad directions derived from the collaboration need to be transformed into a real, completely functional robot. Although LLMs do not provide comprehensive technical support, they can offer their own insights on whether the technology is feasible, helping researchers reduce the risk of failure.

Presently, LLMs lack the ability to generate comprehensive CAD models, evaluate code, or independently fabricate robots. Nevertheless, advancements in LLM research suggest that these algorithms can offer significant assistance in executing software (Chen et al., 2021), mathematical reasoning (Das et al., 2024), and even in the generation of shapes (Ramesh et al., 2022). Lu et al. specifically focused on the utilization of LLMs for organizing unstructured metadata, facilitating the conversion of metadata between different formats, and discovering potential errors in the data collection process (Lu et al., 2023). They also envisioned the next generation of LLMs as remarkably potent tools for data visualization (Bubeck et al., 2023), and anticipated that these advanced models will provide invaluable support to researchers, enabling them to extract meaningful insights from extensive volumes of phenotypic data.

Although LLMs provide insights can indirectly help farmers solve a small number of agricultural tasks, it's important to note that their insights should be used in conjunction with human judgment and domain expertise. That is to say, the insights provided by LLMs cannot be separated from human experience.

3.2.3 Large language models empower decision-making for farmers

According to a recent study, ChatGPT demonstrates the ability to comprehend natural language requests, extract valuable textual and visual information, select appropriate language and vision tasks, and effectively communicate the results to humans (Shen et al., 2024). Shen et al. proposed a system named HuggingGPT to solve AI tasks. HuggingGPT is a collaborative AI task resolution framework built on LLMs. This system connects LLM with AI models through language interface, and these AI models are derived from HuggingFace. This coordinating capability positions LLMs as the core of modern AI Agents. As the core of decision-making, LLM can be applied to agriculture to help solve the tasks proposed by farmers (Shen et al., 2024).

An AI Agent is an autonomous system that perceives its environment, reasons, plans, and acts to achieve specific goals. As shown in Figure 5, the LLM acts as the agent's "brain", performing crucial functions. When receiving a task request, LLM first divides the total task into subtasks and selects the appropriate AI model based on the needs of each subtask. For example, converting farmers' audio into text requires the use of an audio to text model [Amazon transcribe, Whisper (Radford et al., 2023)]; It is also necessary to recognize the sent image and integrate the text obtained from the audio conversion in the previous step to obtain a text-response (vit-gpt2); Considering that some farmers may have had limited access to formal education, it is necessary to further convert text-response into audio and ultimately obtain the audio-response [Fastspeech (Ren et al., 2019, 2020)]. Although LLM does not play a role in solving problems throughout the entire system, as a "conductor", it can coordinate various AI models to complete subtasks, thereby gradually solving complex tasks and playing a core role in decision-making support.

3.3 The role of large vision model in image processing, analysis, and agricultural automation

While LLMs excel in processing textual and knowledge-based information, many agricultural tasks fundamentally rely on visual data. Using a LVM to judge crop related information can not only greatly improve the time required for judgment, but also indirectly reduce the damage caused to crops. Moreover, after crops are invaded by pests and diseases, their color, texture, spectral characteristics will undergo certain changes, all of which are related to CV.

3.3.1 Image processing and analysis

At present, there are four types of methods for obtaining crop image information: 1) ordinary channels, taking photos to obtain images; 2) obtaining remote sensing images through agricultural machinery near the ground; 3) obtaining remote sensing images through aircraft monitoring platforms (Yuan et al., 2022); 4) obtaining remote sensing images through satellites (Zhang et al., 2019). Remote sensing can provide large-scale land use and land cover information. By analysing satellite images or high-altitude images, various surface information can be identified, such as surface conditions, soil moisture, vegetation coverage, and crop growth status (Khanal et al., 2020). Classifying and segmenting from limited examples obtained from remote sensing is a significant challenge. Regarding this, Wu et al. (2023b) put forward GenCo (a generator-based two-stage approach) for few-shot classification and segmentation on remote sensing and earth observation data. Their approach presents an alternative solution for addressing the labelling challenges encountered in the domains of remote sensing and agriculture. Spectral data can provide rich insights into the composition of observed objects and materials, especially in remote sensing applications. The challenges faced in processing spectral data in agriculture include: 1) effectively processing and utilizing vast amounts of remote sensing spectral big data derived from various sources; 2) deriving significant knowledge representations from intricate spatial-spectral mixed information; 3) addressing the spectral degradation in the modelling of neighbouring spectral relevance. Hong et al.'s SpectralGPT empowers intelligent processing of spectral remote sensing big data, and this LVM has also demonstrated its excellent spectral reconstruction capabilities in agriculture (Hong et al., 2024). Due to multispectral imaging (MSI) and hyperspectral imaging (HSI) make it possible to monitor crop health in the field. The integration of remotely sensed multisource data, such as HSI and LiDAR (Light detection and ranging), enables the monitoring of changes occurring in different parts of a plant (Omia et al., 2023). By using a large visual model to analyse these spectral data, the obtained crop health information can help farmers quickly and

accurately identify diseases and treat them, reducing the loss of crop yield.

Studies suggest that the use of LVMs for image recognition and predictive analysis of crop information is often more effective than traditional ML algorithms. When farmers need to obtain crop information, four types of image acquisition methods can be used to obtain crop image information (Figure 6). Then, the image information is processed through image recognition (Divided into four tasks: image classification, object detection, semantic segmentation, instance segmentation), and the identified results need to be further predictive analytics to obtain crop information that farmers can understand.

In addition to obtaining information by analysing the phenotypic characteristics of crops, Feng et al. (2022) developed a traditional DL model called organelle segmentation network (OrgSegNet). OrgSegNet is capable of accurately capturing the actual sizes of chloroplasts, mitochondria, nuclei, and vacuoles within plant cell, further inspecting plant phenotypic at the subcellular level. They have tested two applications: 1) A thermo-sensitive rice albino leaf mutant was cultivated at cold temperature conditions. In the transmission electron microscope images (TEMs), the albinotic leaves lacked typical chloroplasts, and OrgSegNet failed to identify any chloroplast structures; 2) Young leaf chlorosis 1 (Ylc1). Young leaves of the ylc1 mutant showed lower levels of chlorophyll and lutein compared to corresponding wild type, and its TEM analysis further revealed a noticeable loose arrangement of the thylakoid lamellar structures. It can be imagined that if a large model is used to replace DL algorithms, the recognition of subcellular cells may perform better, and the recognition results can be further predictive analytics to obtain information that non plant experts can also understand.

3.3.2 Automation and robotics

Enhancing the intelligence of agricultural robots is a crucial application area for large models. Conventional agricultural robot systems, typically composed of perception, decision-making, and actuation modules, often struggle with complex visual perception and intelligent, real-time decision-making, especially in unstructured and dynamic farm environments (Yang et al., 2023b; Hamuda et al., 2016). Integrating large models is a promising approach to overcome these limitations and significantly enhance the intellectual features of agricultural robots.

Current LVMs can be used in drones to monitor crops and obtain information on their growth, disease, yield, and other factors (Ganeshkumar et al., 2023; Chin et al., 2023; Pazhanivelan et al., 2023). In addition to the above functions, ground machines that used LVMs can also be used for harvesting and classifying crops, as well as detecting pests up close. In (Yang et al., 2023c), a LVM, segment anything model (SAM) (Kirillov et al., 2023), uses infrared thermal images for chicken segmentation tasks in a zero-shot means. SAM can be used in agriculture to segment immature fruits on a fruit tree and quickly achieve yield prediction. Yang et al. (2023a) subsequently proposed the Track Anything Model (TAM) by combining SAM and video. Unfortunately, TAM places more emphasis on maintaining short-term memory rather than long-term memory. Nevertheless, based on its capabilities, TAM still has great potential in the agricultural field. If its long-term memory ability can be improved, it can monitor early changes in crop diseases and provide early warning to farmers. Embedding LVMs such as SAM and TAM into robots can not only achieve automation in agriculture, but these LVMs themselves can help achieve automation in agricultural robot design.

Beyond perception, large models are also revolutionizing the design process of agricultural robots. As mentioned previously, Stella et al. (2023) demonstrated using LLMs like ChatGPT to assist in designing an optimized robotic gripper for tomato picking. With the latest multimodal versions like GPT-4.5, designers can now input not only text descriptions but also sketches to partially automate the robot design process. This integration of LVMs for perception and LLMs for both control logic and design automation marks a significant step towards fully autonomous agricultural systems.

3.4 Integration of multimodal models

LVMs provide powerful capabilities for visual analysis and robotic perception. However, the most complex agricultural challenges often require integrating information from multiple sources. MLLM recently has emerged as a prominent research hotspot (Figure 7), which uses powerful LLMs as a core to tackle multimodal tasks (Yin et al., 2023). In recent years, many researchers have utilized and merged diverse types of data inputs, such as text, images, audio, video (Zhang et al., 2023a), sensor data (Driess et al., 2023), depth information, point cloud (Chen et al., 2024), and more.

The agricultural community has started exploring the realm of multimodal learning in agricultural applications. By incorporating multimodal learning techniques, the agricultural community seeks to unlock new opportunities for optimizing various agricultural processes and achieving improved outcomes. As an example, Bender et al. have released an open-source multimodal dataset specifically curated for agricultural robotics (Bender et al., 2020). This dataset was collected from cauliflower and broccoli fields and aims to foster research endeavors in robotics and ML within the agricultural domain. It encompasses a diverse range of data types, including stereo color, thermal, hyperspectral imagery, as well as essential environmental information such as weather conditions and soil conditions. The availability of this comprehensive dataset uses as a precious resource for advancing the development of innovative solutions in agricultural robotics and ML. Cao et al. (2023) proposed a novel approach for cucumber disease recognition using a MLLM that incorporates image-text-label information. Their methodology effectively integrated label information from many domains by employing image-text multimodal contrastive learning and image self-supervised contrastive learning. The approach facilitated the measurement of sample distances within the common image-text-label space. The results of the experiment demonstrated the effectiveness of this innovative approach, achieving a recognition accuracy rate of 94.84% on a moderately sized multimodal cucumber disease dataset.

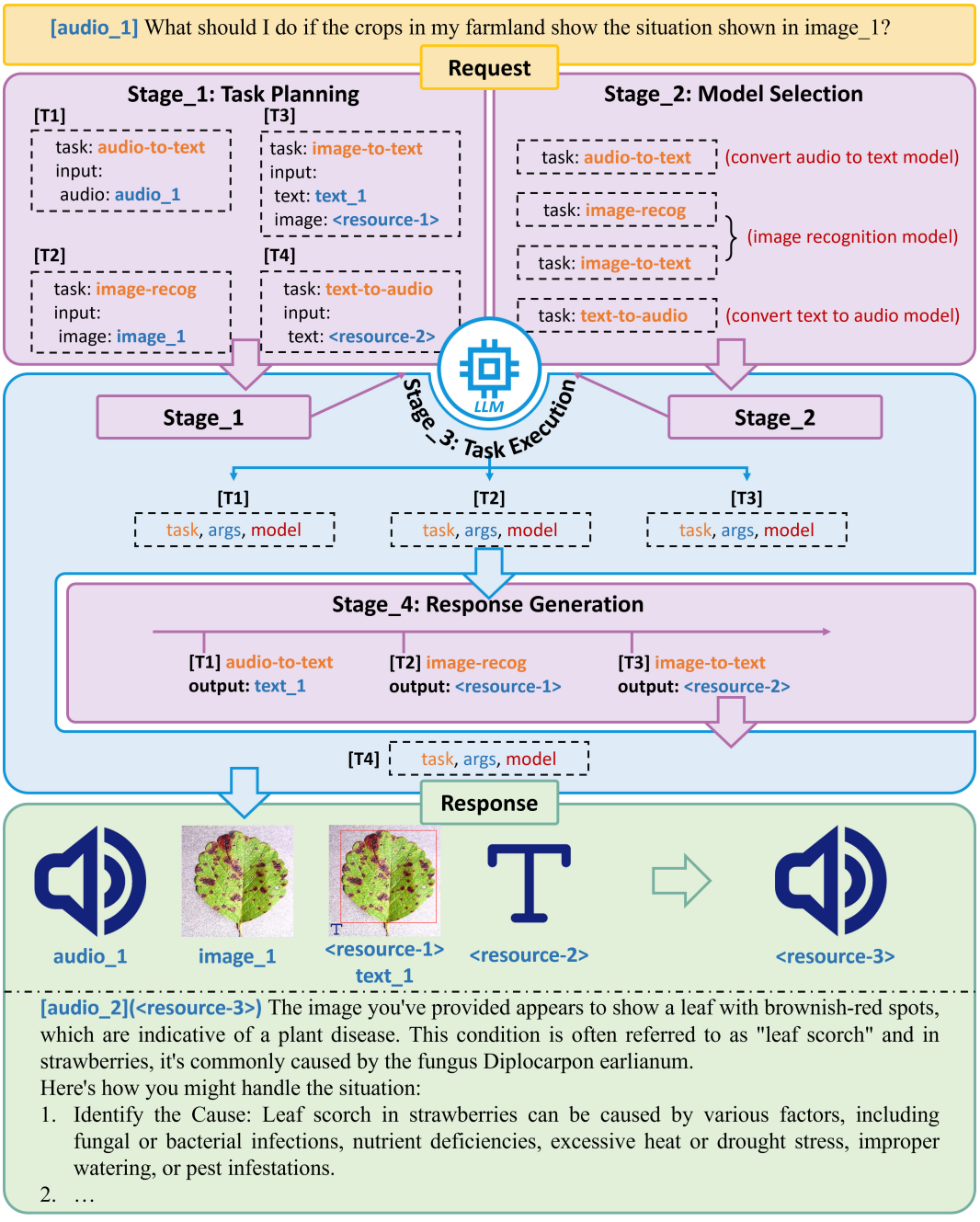


FIGURE 5
An LLM-based AI Agent architecture for agricultural decision-making support.

Nevertheless, it is important to highlight that existing models primarily rely on text-image data and are mostly limited to QA tasks. There is a noticeable lack of applications in the realm of agricultural robotics that incorporate inputs like images, text, voice (Human instructions), and depth information (From LiDAR or laser sensors). These agricultural robots, commonly deployed for tasks such as fruit picking or crop monitoring (Tao and Zhou, 2017), present a significant opportunity for the integration of multimodal data sources to enhance their capabilities. In short, large models lacking a high degree of multimodality perform fewer tasks and lack good applicability.

3.5 The choice between large models and traditional models

The decision to implement either a large model or a traditional model in agriculture is not straightforward. It involves considering a

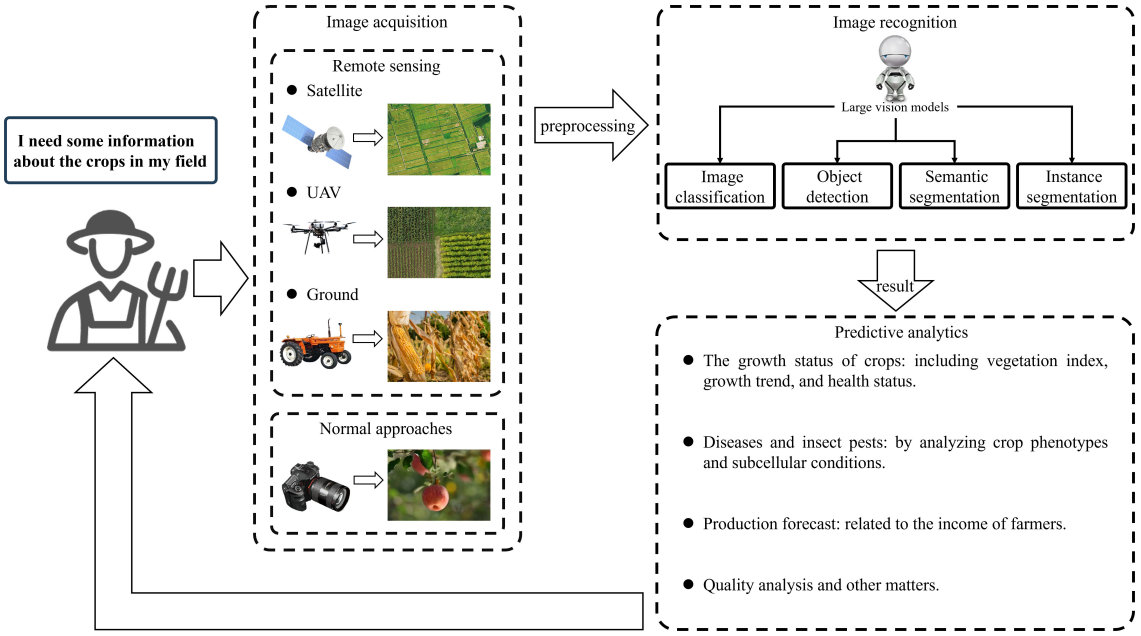


FIGURE 6
Farmers can obtain crop information through the process of image acquisition, image recognition, predictive analytics.

multitude of factors, such as the volume and quality of available data, the required model generalizability, and the practical limits on computational power and inference speed. However, by analyzing the studies of Deforce et al. (2024); Zhao et al. (2023a), and Cao et al. (2023), we found that these considerations can be effectively categorized under two primary factors: Data and deployment conditions. Similar to how large models are divided into LLM, LVM, and MLLM, traditional models can be classified according to the specific agricultural task, falling into the categories of NLP, CV, and multimodal. For instance, models like AGRONER and PSO-LSTM are designed to handle NLP tasks (Veena et al., 2023; Zheng and Li, 2023), AG-YOLO and CMTNet address CV tasks (Lin et al., 2024; Guo et al., 2025b), while ITK-Net and Multi-ResNet34 are tailored for multimodal applications (Zhou et al., 2021; Zhang et al., 2022a). Before selecting a model, it is best to first determine which category the agricultural task belongs to.

When approaching an agricultural task, a critical step is to assess the sufficiency of available data. If a substantial volume of high-quality, task-specific data is available, a traditional model becomes a good option. Conversely, in scenarios marked by data insufficiency, leveraging a large model is often the more suitable choice. Figure 8 presents a comparative analysis of traditional models versus large models based on data conditions and deployment constraints. PSO-LSTM can be retrained on abundant data, and it can deliver superior performance for a particular agricultural task, thus positioning this model as a "specialist". TimeGPT, on the other hand, functions as a "generalist", capable of handling diverse, non-specific agricultural tasks using only minimal fine-tuning or a zero-shot approach in data-scarce situations, thereby avoiding the need for complete model retraining for each new task (Deforce et al., 2024). The pre-embedded knowledge within large models can effectively compensate for the lack of domain-specific data.

On the other hand, deployment conditions are also a crucial factor in model selection. While devices with high computational capacity can deploy both large models and traditional models, the significant computational and time costs associated with large models make them unsuitable for edge devices and systems requiring real-time response. For an agricultural task that requires the model to be deployed on an edge device with real-time detection needs, ITK-Net is the pragmatic and superior choice due to its efficiency and low resource requirements (Zhou et al., 2021). While the ITLMLP model proposed by Cao et al. (2023) also targets crop disease recognition, it is suited for deployment only on devices that can handle high computational costs. As a large model, ITLMLP's deployment conditions are considerably more stringent than those of ITK-Net, the traditional model. However, this does not imply that ITLMLP is without its merits. Its value lies not in real-time field deployment, but in its powerful offline analysis capabilities. By batch-processing vast agricultural data stored on cloud servers, it can perform in-depth retrospective diagnostics and trend analysis. Leveraging its powerful feature extraction and generalization capabilities, acquired from pre-training on large-scale data, ITLMLP can conduct reclassification of historical disease data, compile statistics on disease occurrence frequencies across different periods, and uncover potential correlations between image features and specific environmental descriptions. By the way, by optimizing the model architecture, using efficient inference algorithms, and utilizing hardware acceleration techniques, the real-time performance of LVMs can be improved to a certain extent (Chen et al., 2023b). In addition, we have also

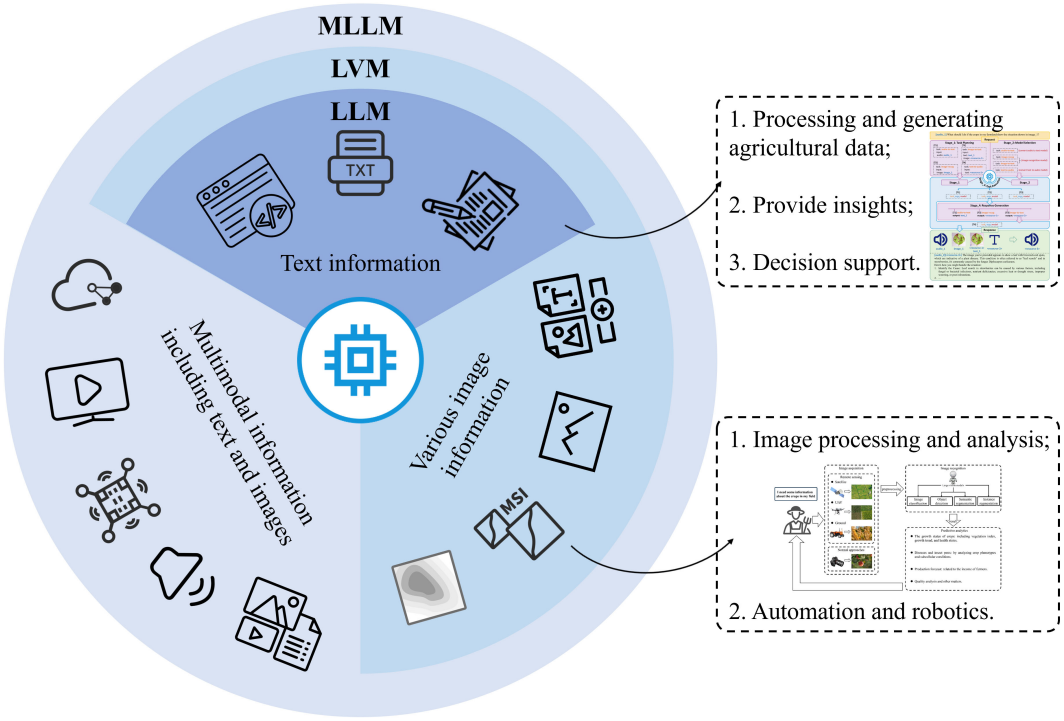


FIGURE 7
Multimodal information fusion analysis driven by MLLM.

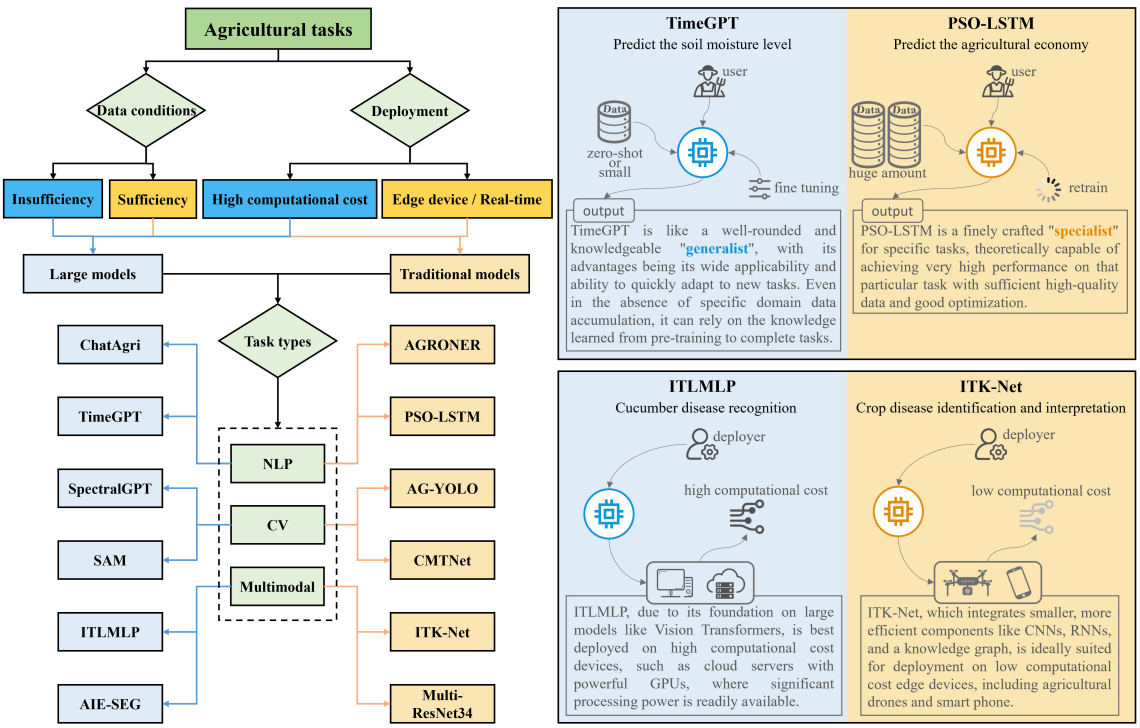


FIGURE 8
Comparison of large models and traditional models for agricultural tasks.

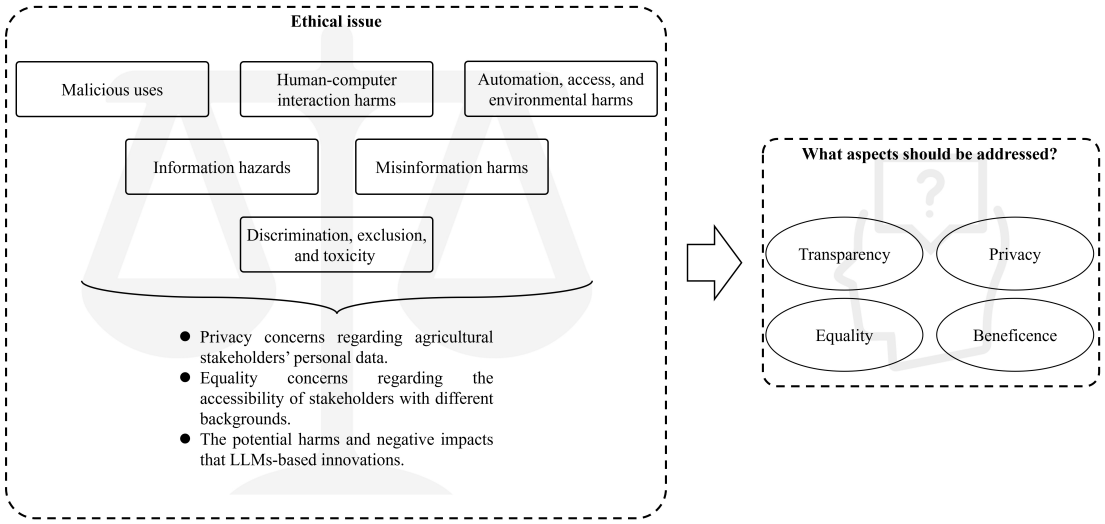


FIGURE 9
The ethical issues faced by large models.

discovered that ITLMLP could process a large dataset to generate highly accurate annotated labels, which can then be used to train smaller, more efficient models like ITK-Net. This creates a synergistic ecosystem where the power of large models enables the effectiveness of traditional models on the edge.

The choice between a large model and a traditional model for agricultural tasks is not a matter of one being definitively superior to the other, but rather a strategic decision based on a careful evaluation of trade-offs. Large models, with their powerful generalization capabilities, offer a robust solution for data-scarce environments, while traditional models excel in data-rich scenarios where their specialized nature can be fully leveraged. Similarly, the high computational cost of large models makes them suitable for offline, server-based analysis, whereas the efficiency of traditional models is indispensable for real-time, on-device deployment.

4 Ethical issues and responsible use of large vision and language models in agriculture

As large models demonstrate their powerful potential and are increasingly applied to agricultural tasks (referencing section 3), it is crucial to critically examine the ethical and societal implications of their deployment. However, there are often ethical and responsibility issues in the development and deployment process of AI today. The digital gap between those who have the resources to develop and utilize large models and those who cannot afford to do so creates an inequality in accessing large models, resulting in an unfair distribution of risks and benefits (Harfouche et al., 2024). Not only that, this divide can be exacerbated by the presence of AI biases (Dara et al., 2022; Ryan, 2023). Accordingly, to ensure ethical issues and responsible use of large models, this chapter starts from the ethical and responsibility issues in the agricultural large models and explore corresponding measures.

4.1 Ethical considerations in the deployment of large models in agriculture

Predicting and solving ethics problems of large models in agriculture is a critical scientific and societal challenge. Although large models point the way for the future of smart agriculture, due to their characteristic of being influenced by close association, large models often learn some bad knowledge in addition to useful knowledge. Ethical issues have always been an indispensable topic of discussion in the process of technological progress (Such as the ethical issues discussed by Holmes et al. in the field of education regarding educational AI (Holmes et al., 2022)), and we also need to pay attention to ethics issues when using large models in the agricultural direction. As mentioned below, many relevant institutions and personnel have put forward their own ideas on ethical issues related to large models.

Weidinger et al. (2021) put forward six types of ethical risks (Figure 9): 1) Malicious uses, 2) Human-computer interaction harms, 3) Automation, access, and environmental harms, 4) Information hazards, 5) Misinformation harms, and 6) Discrimination, exclusion, and toxicity. Understanding these issues can help us responsibly use large models in the agricultural field.

1. Malicious uses. Prior to the release of GPT-4, OpenAI hired a team of 50 experts and scholars to conduct a six-month adversarial test on GPT-4. Andrew White, a professor of chemical engineering at the University of Rochester who was invited to participate in the test, stated that early versions of GPT-4 could assist in the manufacture of

chemical weapons and even provide a convenient manufacturing location. From the perspective of the agricultural sector, if this issue is not properly addressed, some may use large models to learn ways to destroy other people's farmland for the sake of profit, thereby allowing themselves to have a larger market. Over time, this will lead to vicious competition in the market.

2. Human-computer interaction harms. The potential harms of human-computer interaction arise when users excessively trust a large model or mistakenly treat it as human.
3. Automation, access, and environmental harms. The large model can give rise to automation, access, and environmental harms due to its potential environmental or downstream economic impacts.
4. Information hazards. Due to the involvement of information from different countries, religions, and ethnicities, model outputs leaking or inferring sensitive information often led to political violence.
5. Misinformation harms. A study discussed the potential risks of using poorly performing large models. The original intention of this study was to provide a natural language generation model in MOOC to respond to students and improve their participation rate (Li and Xing, 2021). Even so, due to the poor performance of the model, the corresponding negative results further reduced the enthusiasm of students. If a poorly performing large model is used in the agricultural field, it may mislead farmers in their judgment (Such as analysing incorrect disease types), not only causing further damage to crops in the farmland, but also making farmers increasingly distrust the large model. For this phenomenon, Angelone et al. proposed that warning labels can be applied to the content generated by the large model (Angelone et al., 2022), but this also involves the trust issue of the large model in its own generated results.
6. Discrimination, exclusion, and toxicity. Two researches have indicated that potential discrimination, exclusion, and toxicity issues may occur if adopting a model that is accurate but unfair (Sha et al., 2021; Merine and Purkayastha, 2022).

Despite Weidinger et al.'s viewpoint can provide us with a fundamental understanding of the risks associated with large models, manners of systematic ethical supervision of large models' research and innovation (R&I) are especially restricted. Coincidentally, the European Commission has officially approved comprehensive "ethics guidelines for trustworthy AI" specifically designed for R&I. These guidelines require that principal investigators recognize and tackle the ethical matters raised by their proposed research. Principal investigators are also required to adhere to ethical principles and relevant legislation in their work. In a similar vein, Stanford University's Ethics and Society Review necessitates researchers to distinguish potential societal hazards associated with their research and incorporate mitigation measures into their research design (Bernstein et al., 2021).

Furthermore, projects with large models have a vast amount of data and often raise ethics issues. For instance, while raw plant science data itself may not inherently fall within the scope of the European Union General Data Protection Regulation (GDPR) as personal data, it can become subject to GDPR regulations when linked to identifiable individuals or specific farm locations tied to individuals, creating complex challenges concerning data ownership and privacy protection (Harfouche et al., 2024). Thus, relevant guidelines must consider code of conduct for data sharing, privacy protection, and the overall governance of datasets.

4.2 Responsible use in agriculture

With the expanding development and utilization of large models, there is a growing recognition of the need for agile and effective regulatory oversight. To address this issue, it may be necessary to use AI technology to assist in overseeing the development and deployment of large models. Regarding this aspect, the AI Act, which has been jointly agreed upon by the European Parliament and the Council of Europe, represents the first comprehensive set of harmonized rules on a global scale. It promotes responsible large model designment and development by regulating large model across various applications and contexts based on a risk-based framework. Within the framework, careful consideration must be given to the level of risk involved and how to evaluate different large models as risk-free or low-risk.

To evaluate the risk level of a large model, we focus on four aspects: transparency, privacy, equality, and beneficence. On the other hand, in addition to developing and adhere to a strong regulatory framework that guides the development, deployment, and use of large models, regulatory methods also need to be considered. Consider the potential societal impact, potential harms, and long-term implications of the technology. Firstly, due to the wide applicability of large models, we cannot make a one size fits all approach. Regulation must adapt to specific issues in different domains. The United States' food and drug administration (FDA) has tailored potential regulatory methods for AI and ML technologies used in medical devices, categorizing them into three major categories based on risk levels: Class I (Low risk), Class II (Moderate risk), and Class III (High risk). Large models in agriculture can also be regulated according to the FDA's approach, dividing them into several types of models ranging from low risk to high risk. For example, genetically modified crops may have environmental impacts, food safety issues, and ecosystem damage, so large models targeting genetically modified crops should be included in high-risk types. For large models of ordinary crops, they can be classified as low-risk types. And the regulatory methods proposed by relevant departments should be made public to ensure transparency of information. Regulators can promote fairness in the deployment of agricultural large models by enforcing the use of diverse and representative data sources, which

helps mitigate potential biases present in the training data (Meskó and Topol, 2023).

From the perspective of beneficence and privacy, privacy issues related to large models have received little attention or investigation in reviewed research (Yan et al., 2024). Specifically, if the training set used to train a large model contains some personal privacy information that has not been authorized by the information owner. The disregard for privacy concerns is especially worrisome considering that LLM-based innovations involve stakeholders' natural languages, which may contain personal and sensitive information pertaining to their private lives and identities (Brown et al., 2022). If users unintentionally learn about this information while using a large model, it may cause harm to the beneficence of the information owner. Developers of large models should ensure they gain explicit consent from individuals before collecting and utilizing these personal data. Clearly communicate the purpose and scope of data usage, and offer individuals with the choice to choose out or request data deletion. Besides, limit the amount of personal and sensitive data collected and stored. Follow the principle of data minimization, ensuring that only necessary data is collected and retained. Anonymize or aggregate data whenever possible to protect individual privacy.

In general, governance approaches that promote responsible utilization of large models and focus on the outcomes rather than the technology itself will enhance research efforts and drive more innovation. By combining governance and ethics, we can harness the powerful synergy to expedite the implementation of large models in agriculture and other domains, fostering innovation at a larger scale.

5 Challenges and future directions

Although large models can play a powerful role in the field of agriculture, they still face challenges in many aspects.

5.1 Technical and practical challenges

5.1.1 Difficulty in obtaining agricultural data

A primary and recurring obstacle highlighted throughout this review is the acquisition of suitable agricultural data. While large models' data generation capabilities can partially alleviate this, as discussed in section 3.2.1, several fundamental difficulties persist:

1. **Cost and quality**: Acquiring comprehensive, high-quality, and accurately labeled real-world data is a time-consuming, labor-intensive, and costly process, especially for supervised learning approaches (Li et al., 2023a; Lu and Young, 2020).
2. **Privacy and trust**: As mentioned in section 4.2, the private nature of farmland data raises significant privacy and trust concerns among farmers, often leading to a reluctance to share information crucial for model training.
3. **Temporal complexity**: Agricultural data is inherently temporal. The need to capture entire crop growth cycles, which are influenced by daily, seasonal, and annual variations, adds another layer of complexity to data collection efforts (Li et al., 2023b).

5.1.2 Low training efficiency

Directly related to the need for massive datasets is the challenge of low training efficiency and high computational cost. As systematically compared against traditional models in section 3.5 (Figure 8), training large agricultural models is a resource-intensive endeavor. Their massive parameter counts demand significant computational power and lengthy training times, often measured in thousands of GPU hours (Li et al., 2023b). This stands in stark contrast to the efficiency of traditional models like YOLO and Faster R-CNN, whose lower computational requirements make them a more practical and cost-effective solution for many specific, real-time agricultural tasks (Badgujar et al., 2024). This efficiency gap explains the continued prevalence of traditional models despite the emergence of more powerful large-scale architectures.

5.1.3 Distribution shift

The problem of distribution shift is a major challenge when using large models in agriculture. When the data encountered by the model during deployment is obviously different from the data used in its training phase, a distribution shift will occur. The environmental conditions for collecting data may vary greatly in different regions and climates. These changes may include differences in crop types, soil conditions, weather patterns, and agricultural practices, all of which can lead to significant changes in data distribution (Wiles et al., 2021). The distribution shift will result in the trained large model not having strong applicability and may not achieve good results in some agricultural tasks. For example, it has been proven that applying large models directly to leaf segmentation tasks in a zero-shot means led to unsatisfactory performance, which can be attributed to possible distribution shifts (Chawla et al., 2021).

5.1.4 The lag of data

After the trained large model is put into use, the data used for training has a certain timeliness for a long period of time. But after a long time, some data lags in time, and the results obtained by using a large model may deviate from the current facts (Figure 10).

5.1.5 Query formulation impacts model output

The results obtained from large models can vary significantly depending on how the query is formulated. Like Figure 11, when multiple images are spliced together for questioning, GPT-4 provides ambiguous answers; When only asking for one image, GPT-4 provides a clearer answer.

To clear these obstacles, future research and development work needs to pay attention to model optimization techniques such as model compression and efficient network structure design, reducing model size without affecting performance (Zhong et al., 2023). It is also

necessary to provide update and maintenance functions for the model to ensure its timeliness. Developers need to write relevant usage instructions to help users get started quickly. Notably, emerging frameworks like RAG offer a direct solution to the data lag and accuracy challenges by connecting LLMs to real-time, external knowledge bases. Similarly, developing more sophisticated AI Agents capable of autonomous planning and tool use will be crucial for creating robust and adaptable agricultural systems.

5.2 Infrastructure and cost barriers

Applying large models to rural areas faces significant barriers related to poor connectivity and high implementation costs. These limitations disproportionately affect small-scale farmers and regions with underdeveloped infrastructure, exacerbating existing inequalities in agricultural productivity and technological access (Da Silveira et al., 2023).

Dibbern et al. (2024) found that farmers often abandon digital tools due to unreliable broadband or mobile connectivity, even when initial investments are made. Technologies like IoT, cloud-based analytics, and real-time monitoring systems remain underutilized in areas lacking stable network access. This has brought some warnings for the application of large models in rural areas. In addition, the high cost of agricultural machinery using large models—render them inaccessible to resource-limited farmers. For example, autonomous machinery and AI platforms often require upfront investments exceeding $10,000 USD, a prohibitive sum for smallholders (Bolfe et al., 2020).

To overcome poor connectivity, investing in and expanding rural broadband and mobile infrastructure is crucial, potentially through government subsidies, public-private partnerships, and the exploration of alternative network solutions like satellite internet or mesh networks tailored to agricultural regions. To mitigate high implementation costs, promoting the development of affordable, modular agricultural machinery and large model platforms designed specifically for smallholder farmers is essential. In short, bridging the digital divide and promoting inclusive technological progress requires joint efforts among technology developers, agricultural researchers, policy makers, and local farmer organizations.

5.3 Future trends in the integration of agricultural and food sectors and large models

In the future, there will undoubtedly be agricultural large models with better performance and higher applicability. And the large models in agriculture should not be limited to text and image inputs. We believe that future multimodal agricultural models can support multimodal information such as videos (Analyzing crops in videos) and audio (Tapping watermelons, and judging maturity through the sound emitted). On the other hand, agriculture is closely related to food, and the development of large models in agriculture is likely to promote the development of large models in the food domain. Trust is indispensable for agriculture and food system technologies given food's universality and importance to people (Tzachor et al., 2022). Researchers need to navigate

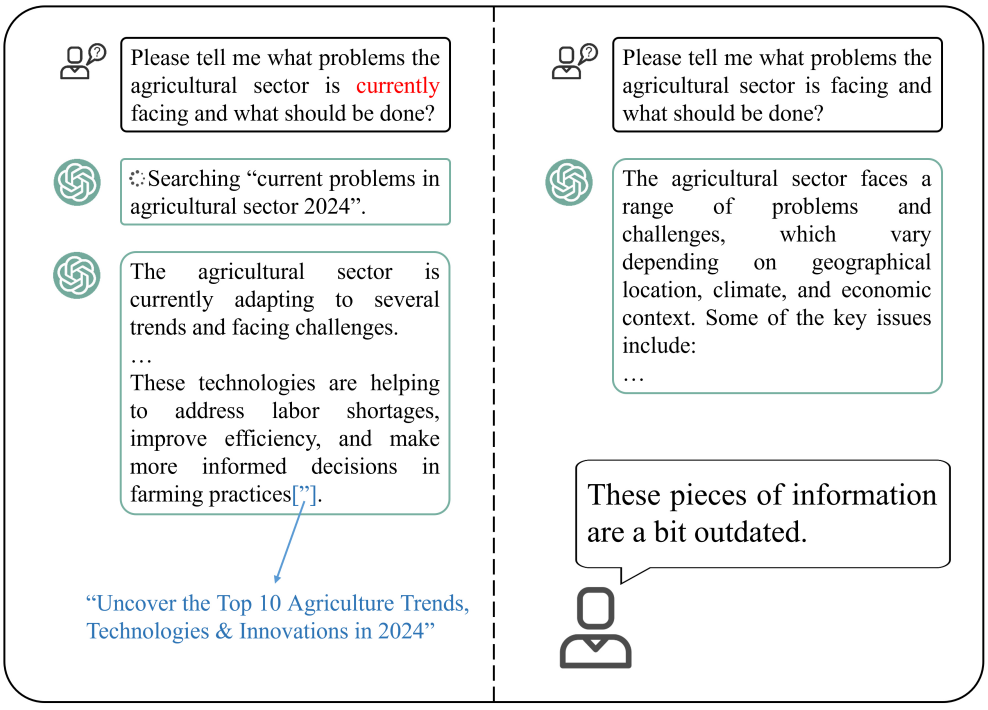


FIGURE 10
The lag of data.

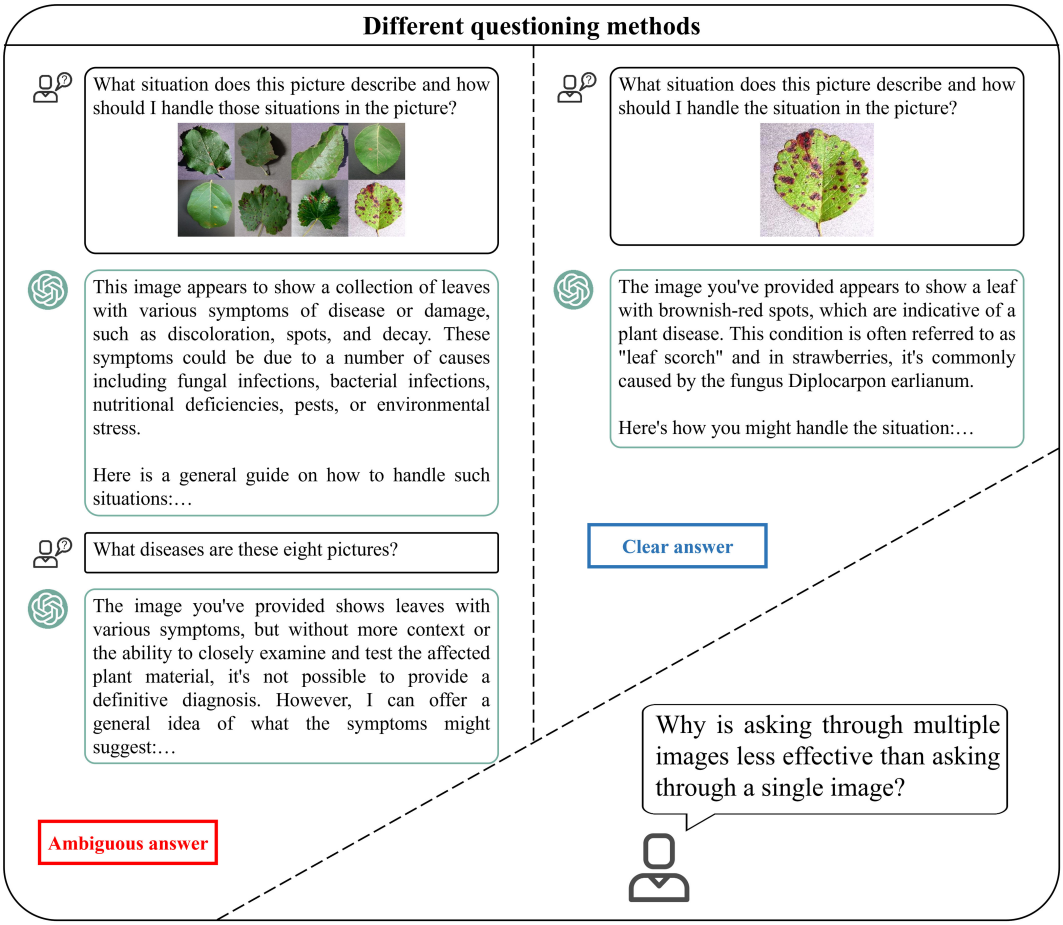


FIGURE 11
Different questioning methods can lead to different results.

complicated social, political, economic, and environmental landscapes to develop appropriate large models in the food industry. In the future food industry, researchers will strive to establish trust with governmental agencies and funders, as well as with food system partners, to provide food and products that the public trusts (Alexander et al., 2024).

Overall, although the agricultural large model still faces many challenges at present, we believe that through the joint efforts of relevant researchers in the future, these challenges can be properly addressed. And due to the close relationship between the food and agricultural domains, with the gradual development of agricultural large models, food large models will also receive further research, thereby achieving mutual positive feedback between the development of large models in these two fields.

6 Conclusion

In summary, this study investigated the application status of large models in the agricultural field. Our analysis establishes that these models offer unprecedented advantages through their capacity for complex reasoning, multimodal information processing, and the execution of nuanced tasks ranging from pest identification to robotic automation. We further determined that the efficacy of these powerful tools is significantly amplified when they are tailored to the agricultural domain, a crucial strategy for overcoming the pervasive challenge of limited labeled data. Furthermore, this review provided a pragmatic framework for choosing between large and traditional models, emphasizing that the decision hinges on a careful trade-off between data availability and deployment constraints. While large models excel as “generalists” in data-scarce or offline analytical scenarios, efficient traditional models remain indispensable as “specialists” for real-time, on-device tasks.

However, this vast potential is tempered by critical, interconnected challenges that must be addressed. A primary hurdle is the acquisition and utilization of suitable agricultural data; issues of data scarcity, high collection costs, inherent data diversity (across crops, regions, conditions), privacy concerns associated with farmland data, and the need for time-series information create significant obstacles. Furthermore, the high computational resources required for training and deploying large models, coupled with the often-limited internet connectivity and financial resources in rural areas, creates a significant digital divide, potentially excluding smallholder farmers. Technical issues such as model susceptibility to

distribution shifts between training and deployment environments, the problem of data lag impacting real-time relevance, and sensitivity to query formulation also impact the reliability and practical applicability of current models. Finally, overarching ethical considerations, including potential biases in data or algorithms, ensuring data privacy, promoting equitable access to technology, and preventing misuse, are paramount and demand careful consideration and robust governance frameworks.

Although our study is comprehensive, there are inherent limitations to studying a rapidly developing field. To move forward, future research must directly confront the limitations and challenges identified. Developing novel techniques to mitigate data scarcity—such as advanced data augmentation and self-supervised learning tailored for agriculture—is a critical priority. Expanding multimodal capabilities to robustly incorporate inputs like video, audio, and diverse sensor data will unlock new frontiers in precision farming. Crucially, research must move beyond theoretical ethics to the practical implementation of governance structures for AI in agriculture. Furthermore, a significant opportunity lies in exploring the synergistic relationship between agricultural large models and the broader food system, addressing challenges from farm to fork.

Large models stand poised to be transformative technologies for agriculture. While significant challenges remain, the potential benefits for productivity, sustainability, and food security are immense. Addressing the technical hurdles, bridging the digital divide, and navigating the ethical landscape through collaborative, responsible innovation will be key to realizing this potential. We hope this article serves as a valuable resource and a cornerstone, stimulating further research and guiding the development of future agricultural large models that are not only powerful but also practical, efficient, and beneficial for all stakeholders in the global food system.

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

HZ: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review & editing. SQ: Conceptualization, Formal Analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – original draft, Writing – review & editing. MS: Funding acquisition, Writing – review & editing. CL: Software, Writing – original draft. AL: Software, Writing – original draft. JG: Conceptualization, Investigation, Project administration, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. Our research was supported by the Natural Science Foundation of Guangxi (No. 2024 GXNSFBA010381), the National Natural Science Foundation of China (No. 62361006), Guangxi Young Elite Scientist Sponsorship Program (GXYESS2025081), and the grant (No. NCOC-24-02) from Key Laboratory of Nonlinear Circuit and Optical Communications (Guangxi Normal University), Education Department of Guangxi Zhuang Autonomous Region.

Conflict of interest

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

Generative AI statement

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1579355/full#supplementary-material

References

Ahirwar, S., Swarnkar, R., Bhukya, S., and Namwade, G. (2019). Application of drone in agriculture. *Int. J. Curr. Microbiol. Appl. Sci.* 8, 2500–2505. doi: 10.20546/ijcmas.2019.801.264

Alexander, C. S., Yarborough, M., and Smith, A. (2024). Who is responsible for 'responsible AI'?: Navigating challenges to build trust in AI agriculture and food system technology. *Precis. Agric.* 25, 146–185. doi: 10.1007/s11119-023-10063-3

Angelone, A. M., Galassi, A., and Vittorini, P. (2022). "Improved automated classification of sentences in data science exercises," in *In Methodologies and Intelligent Systems for Technology Enhanced Learning, 11th International Conference*, Vol. 11. 12–21. doi: 10.1007/978-3-030-86618-1_2

Anil, R., Dai, A. M., Firat, O., Johnson, M., Lepikhin, D., Passos, A., et al. (2023). Palm 2 technical report. doi: 10.48550/arXiv.2305.10403

Aymen, F., Monir, H., and Pester, A. (2024). "Large Vision Models: How Transformer-based Models excelled over Traditional Deep Learning Architectures in Video Processing," in *2024 5th International Conference on Artificial Intelligence, Robotics and Control (AIRC)*. 50–54. doi: 10.1109/AIRC61399.2024.10672087

Badgujar, C. M., Poulose, A., and Gan, H. (2024). Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review. *Comput. Electron. Agric.* 223, 109090. doi: 10.1016/j.compag.2024.109090

Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., et al. (2023). Qwen-vl: A frontier large vision-language model with versatile abilities. doi: 10.48550/arXiv.2308.12966

Bender, A., Whelan, B., and Sukkarieh, S. (2020). A high-resolution, multimodal data set for agricultural robotics: A Ladybird's-eye view of Brassica. *J. Field Robot.* 37, 73–96. doi: 10.1002/rob.21877

Bengio, Y., Ducharme, R., and Vincent, P. (2000). A neural probabilistic language model. *Adv. Neural Inf. Process. Syst.* 13.

Bernstein, M. S., Levi, M., Magnus, D., Rajala, B. A., Satz, D., and Waeiss, Q. (2021). Ethics and society review: Ethics reflection as a precondition to research funding. *Proc. Natl. Acad. Sci.* 118, e2117261118. doi: 10.1073/pnas.2117261118

Bi, Z., Zhang, N., Xue, Y., Ou, Y., Ji, D., Zheng, G., et al. (2023). Oceangpt: A large language model for ocean science tasks. doi: 10.48550/arXiv.2310.02031

Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Allen, A., Brandstetter, J., et al. (2024). Aurora: A foundation model of the atmosphere. doi: 10.48550/arXiv.2405.13063

Bolfe, ÉL, Jorge, L. A. C., Sanches, I. D., Luchiari Júnior, A., da Costa, C. C., Victoria, D. C., et al. (2020). Precision and digital agriculture: adoption of technologies and perception of Brazilian farmers. *Agriculture* 10, 653. doi: 10.3390/agriculture10120653

Bommasani, R., Hudson, D. A., Adeli, E., Adeli, E., Altman, R., Arora, S., et al. (2021). On the opportunities and risks of foundation models. doi: 10.48550/arXiv.2108.07258

Bosilj, P., Aptoula, E., Duckett, T., and Cielniak, G. (2020). Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture. *J. Field Robot.* 37, 7–19. doi: 10.1002/rob.21869

Bouguettaya, A., Zarzour, H., Kechida, A., and Taberkit, A. M. (2022). Deep learning techniques to classify agricultural crops through UAV imagery: A review. *Neural Computing Appl.* 34, 9511–9536. doi: 10.1007/s00521-022-07104-9

Brown, H., Lee, K., Mireshghallah, F., Shokri, R., and Tramèr, F. (2022). "What does it mean for a language model to preserve privacy?," in *In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 2280–2292. doi: 10.1145/3531146.3534642

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., et al. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. doi: 10.48550/arXiv.2303.12712

Cao, Y., Chen, L., Yuan, Y., and Sun, G. (2023). Cucumber disease recognition with small samples using image-text-label-based multi-modal language model. *Comput. Electron. Agric.* 211, 107993. doi: 10.1016/j.compag.2023.107993

Chawla, S., Singh, N., and Drori, I. (2021). "Quantifying and alleviating distribution shifts in foundation models on review classification," in *NeurIPS 2021 Workshop on Distribution Shifts: Connecting Methods and Applications*. https://openreview.net/forum?id=OG78-TuPcvL.

Chelba, C., Mikolov, T., Schuster, M., Ge, Q., Brants, T., Koehn, P., et al. (2013). One billion word benchmark for measuring progress in statistical language modeling. doi: 10.48550/arXiv.1312.3005

Chen, C., Du, Y., Fang, Z., Wang, Z., Luo, F., Li, P., et al. (2024). Model composition for multimodal large language models. doi: 10.48550/arXiv.2402.12750

Chen, F., Giuffrida, M. V., and Tsaftaris, S. A. (2023a). "Adapting vision foundation models for plant phenotyping," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 604–613.

Chen, L., Zaharia, M., and Zou, J. (2023b). Frugalgpt: How to use large language models while reducing cost and improving performance. doi: 10.48550/arXiv.2305.05176

Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P.D.O., Kaplan, J., et al. (2021). Evaluating large language models trained on code. doi: 10.48550/arXiv.2107.03374

Chin, R., Catal, C., and Kassahun, A. (2023). Plant disease detection using drones in precision agriculture. *Precis. Agric.* 24, 1663–1682. doi: 10.1007/s11119-023-10014-y

Dai, H., Liu, Z., Liao, W., Huang, X., Cao, Y., Wu, Z., et al. (2023). Auggpt: Leveraging chatgpt for text data augmentation. doi: 10.48550/arXiv.2302.13007

Dara, R., Hazrati Fard, S. M., and Kaur, J. (2022). Recommendations for ethical and responsible use of artificial intelligence in digital agriculture. *Front. Artif. Intell.* 5. doi: 10.3389/frai.2022.884192

Das, D., Banerjee, D., Aditya, S., and Kulkarni, A. (2024). MATHSENSEI: a tool-augmented large language model for mathematical reasoning. doi: 10.48550/arXiv.2402.17231

Da Silveira, F., Da Silva, S. L. C., MaChado, F. M., Barbedo, J. G.A., and Amaral, F. G. (2023). Farmers' perception of the barriers that hinder the implementation of agriculture 4.0. *Agric. Syst.* 208, 103656. doi: 10.1016/j.agsy.2023.103656

Deforce, B., Baesens, B., and Asensio, E. S. (2024). Time-series foundation models for forecasting soil moisture levels in smart agriculture. *arXiv preprint* arXiv:2405.18913. doi: 10.48550/arXiv.2405.18913

Dibbern, T., Romani, L. A. S., and Massruhá, S. M. F. S. (2024). Main drivers and barriers to the adoption of Digital Agriculture technologies. *Smart Agric. Technol.* 8, 100459. doi: 10.1016/j.atech.2024.100459

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. doi: 10.48550/arXiv.2010.11929

Driess, D., Xia, F., Sajjadi, M. S., Lynch, C., Chowdhery, A., Wahid, A., et al. (2023). Palm-e: An embodied multimodal language model. doi: 10.48550/arXiv.2303.03378

ElMasry, G., Mandour, N., Al-Rejaie, S., Belin, E., and Rousseau, D. (2019). Recent applications of multispectral imaging in seed phenotyping and quality monitoring—An overview. *Sensors* 19, 1090. doi: 10.3390/s19051090

Feng, X., Yu, Z., Fang, H., Jiang, H., Yang, G., Chen, L., et al. (2022). Plantorgan hunter: a deep learning-based framework for quantitative profiling plant subcellular morphology. doi: 10.21203/rs.3.rs-1811819/v1

Ganeshkumar, C., David, A., Sankar, J. G., and Saginala, M. (2023). "Application of drone Technology in Agriculture: A predictive forecasting of Pest and disease incidence," in *Applying drone technologies and robotics for agricultural sustainability*, IGI Global, 50–81. doi: 10.4018/978-1-6684-6413-7.ch004

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., et al. (2023). Retrieval-augmented generation for large language models: A survey. doi: 10.48550/arXiv.2312.10997

Geitmann, A., and Bidhendi, A. J. (2023). Plant blindness and diversity in AI language models. *Trends Plant Sci.* 28, 1095–1097. doi: 10.1016/j.tplants.2023.06.016

Girdhar, R., El-Nouby, A., Liu, Z., Singh, M., Alwala, K. V., Joulin, A., et al. (2023). "Imagebind: One embedding space to bind them all," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15180–15190. doi: 10.1109/CVPR52729.2023.01457

Goertzel, B. (2014). Artificial general intelligence: concept, state of the art, and future prospects. *J. Artif. Gen. Intell.* 5, 1–48. doi: 10.2478/jagi-2014-0001

Guo, X., Feng, Q., and Guo, F. (2025b). CMTNet: a hybrid CNN-transformer network for UAV-based hyperspectral crop classification in precision agriculture. *Sci. Rep.* 15, 12383. doi: 10.1038/s41598-025-97052-w

Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., et al. (2025a). Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. doi: 10.48550/arXiv.2501.12948

Hamuda, E., Glavin, M., and Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field. *Comput. Electron. Agric.* 125, 184–199. doi: 10.1016/j.compag.2016.04.024

Harfouche, A. L., Petousi, V., and Jung, W. (2024). AI ethics on the road to responsible AI plant science and societal welfare. *Trends Plant Sci.* 29, 104–107. doi: 10.1016/j.tplants.2023.12.016

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*. 2961–2969. doi: 10.48550/arXiv.1703.06870

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778. doi: 10.1109/CVPR.2016.90

Ho, J., Chan, W., Saharia, C., Whang, J., Gao, R., Gritsenko, A., et al. (2022). Imagen video: High definition video generation with diffusion models. doi: 10.48550/arXiv.2210.02303

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., et al. (2022). Ethics of AI in education: Towards a community-wide framework. *Int. J. Artif. Intell. Educ.* 32 (3), 504-526. doi: 10.1007/s40593-021-00239-1

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., and Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip. Rev.: Data Min. Knowledge Discov.* 9, e1312. doi: 10.1002/widm.1312

Hong, D., Zhang, B., Li, X., Li, Y., Li, C., Yao, J., et al. (2024). SpectralGPT: Spectral remote sensing foundation model. *IEEE Trans. Pattern Anal. Mach. Intelligence*. 46 (8), 5227–5244. doi: 10.1109/TPAMI.2024.3362475

Karthikeyan, L., Chawla, I., and Mishra, A. K. (2020). A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. *J. Hydrol.* 586, 124905. doi: 10.1016/j.jhydrol.2020.124905

Khanal, S., Kc, K., Fulton, J. P., Shearer, S., and Ozkan, E. (2020). Remote sensing in agriculture—accomplishments, limitations, and opportunities. *Remote Sens.* 12, 3783. doi: 10.3390/rs12223783

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., et al. (2023). "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4015–4026. doi: 10.48550/arXiv.2304.02643

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM* 60 (6), 84-90. doi: 10.1145/3065386

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., et al. (2023). Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PloS Digital Health* 2, e0000198. doi: 10.1371/journal.pdig.0000198

Le Scao, T., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., et al. (2023). Bloom: A 176b-parameter open-access multilingual language model. doi: 10.48550/arXiv.2211.05100

Li, J., Chen, D., Qi, X., Li, Z., Huang, Y., Morris, D., et al. (2023a). Label-efficient learning in agriculture: A comprehensive review. *Comput. Electron. Agric.* 215, 108412. doi: 10.1016/j.compag.2023.108412

Li, J., Lammers, K., Yin, X., Yin, X., He, L., Sheng, J., et al. (2024). MetaFruit meets foundation models: leveraging a comprehensive multi-fruit dataset for advancing agricultural foundation models. doi: 10.48550/arXiv.2407.04711

Li, C., and Xing, W. (2021). Natural language generation using deep learning to support MOOC learners. *Int. J. Artif. Intell. Educ.* 31, 186–214. doi: 10.1007/s40593-020-00235-x

Li, J., Xu, M., Xiang, L., Chen, D., Zhuang, W., Yin, X., et al. (2023b). Large language models and foundation models in smart agriculture: Basics, opportunities, and challenges. doi: 10.48550/arXiv.2308.06668

Li, W., and Zhao, Y. (2015). Bibliometric analysis of global environmental assessment research in a 20-year period. *Environ. Impact Assess. Rev.* 50, 158–166. doi: 10.1016/j.eiar.2014.09.012

Lin, Y., Huang, Z., Liang, Y., Liu, Y., and Jiang, W. (2024). Ag-yolo: A rapid citrus fruit detection algorithm with global context fusion. *Agriculture* 14, 114. doi: 10.3390/agriculture14010114

Liu, H., Li, C., Wu, Q., and Lee, Y. J. (2023b). Visual instruction tuning. *Adv. Neural Inf. Process. Syst.* 36, 34892–34916.

Liu, B., Zhao, R., Liu, J., and Wang, Q. (2023a). *AgriGPTs. GitHub*. Available online at: https://github.com/AgriGPTS/AgriGPTs (Accessed April 15, 2025).

Lu, G., Li, S., Mai, G., Sun, J., Zhu, D., Chai, L., et al. (2023). Agi for agriculture. doi: 10.48550/arXiv.2304.06136

Lu, Y., and Young, S. (2020). A survey of public datasets for computer vision tasks in precision agriculture. *Comput. Electron. Agric.* 178, 105760. doi: 10.1016/j.compag.2020.105760

Malla, A., Omwenga, M. M., and Bera, P. K. (2024). "Exploring image similarity through generative language models: A comparative study of GPT-4 with word embeddings and traditional approaches," in *2024 IEEE International Conference on Electro Information Technology (eIT)*. 275–279. doi: 10.1109/eIT60633.2024.10609905

Medeiros, A. D.D., Silva, L. J.D., Ribeiro, J. P.O., Ferreira, K. C., Rosas, J. T.F., Santos, A. A., et al. (2020). Machine learning for seed quality classification: An advanced approach using merger data from FT-NIR spectroscopy and X-ray imaging. *Sensors* 20, 4319. doi: 10.3390/s20154319

Merine, R., and Purkayastha, S. (2022). "Risks and benefits of AI-generated text summarization for expert level content in graduate health informatics," in *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*. 567–574. doi: 10.1109/ICHI54592.2022.00113

Meskó, B., and Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digital Med.* 6, 120. doi: 10.1038/s41746-023-00873-0

Mikolov, T., Karafiát, M., Burget, L., Cernocký, J., and Khudanpur, S. (2010). Recurrent neural network based language model. *Interspeech* 2, 1045–1048. doi: 10.21437/Interspeech.2010-343

Mostaco, G. M., De Souza, I. R. C., Campos, L. B., Cugnasca, C. E., et al. (2018). "AgronomoBot: a smart answering Chatbot applied to agricultural sensor networks," in *14th international conference on precision agriculture*, Vol. 24. 1–13.

Nasir, I. M., Bibi, A., Shah, J. H., Khan, M. A., Sharif, M., Iqbal, K., et al. (2021). Deep learning-based classification of fruit diseases: An application for precision agriculture. *Comput. Mater. Contin* 66, 1949–1962. doi: 10.32604/cmc.2020.012945

Niranjan, P. Y., Rajpurohit, V. S., and Malgi, R. (2019). "A survey on chat-bot system for agriculture domain," in *2019 1st International Conference on Advances in Information Technology (ICAIT)*. 99–103. doi: 10.1109/ICAIT47043.2019.8987429

Omia, E., Bae, H., Park, E., Kim, M. S., Baek, I., Kabenge, I., et al. (2023). Remote sensing in field crop monitoring: A comprehensive review of sensor systems, data analyses and recent advances. *Remote Sens.* 15, 354. doi: 10.3390/rs15020354

Paymode, A. S., and Malode, V. B. (2022). Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG. *Artif. Intell. Agric.* 6, 23–33. doi: 10.1016/j.aiia.2021.12.002

Pazhanivelan, S., Kumaraperumal, R., Shanmugapriya, P., Sudarmanian, N. S., Sivamurugan, A. P., Satheesh, S., et al. (2023). Quantification of biophysical parameters and economic yield in cotton and rice using drone technology. *Agriculture* 13, 1668. doi: 10.3390/agriculture13091668

Peebles, W., and Xie, S. (2023). "Scalable diffusion models with transformers," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4195–4205. doi: 10.48550/arXiv.2212.09748

Peng, R., Liu, K., Yang, P., Yuan, Z., and Li, S. (2023). Embedding-based retrieval with llm for effective agriculture information extracting from unstructured data. doi: 10.48550/arXiv.2308.03107

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017). "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 652–660. doi: 10.1109/CVPR.2017.16

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., and Sutskever, I. (2023). "Robust speech recognition via large-scale weak supervision," in *International Conference on Machine Learning*. 28492–28518.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* 21, 1–67.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., and Chen, M. (2022). Hierarchical text-conditional image generation with clip latents 1, 3. doi: 10.48550/arXiv.2204.06125

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., et al. (2021). "Zero-shot text-to-image generation," in *International conference on machine learning*. 8821–8831. doi: 10.48550/arXiv.2102.12092

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 779–788. doi: 10.1109/CVPR.2016.91

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

Ren, Y., Hu, C., Tan, X., Qin, T., Zhao, S., Zhao, Z., et al. (2020). Fastspeech 2: Fast and high-quality end-to-end text to speech. doi: 10.48550/arXiv.2006.04558

Ren, Y., Ruan, Y., Tan, X., Qin, T., Zhao, S., Zhao, Z., et al. (2019). Fastspeech: Fast, robust and controllable text to speech. *Adv. Neural Inf. Process. Syst.* 32. doi: 10.48550/arXiv.1905.09263

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695. doi: 10.48550/arXiv.2112.10752

Rose Mary, C. A., Raji Sukumar, A., and Hemalatha, N. (2021). Text based smart answering system in agriculture using RNN. *agriRxiv* 2021), 20210310498. doi: 10.31220/agriRxiv.2021.00071

Ryan, M. (2023). The social and ethical impacts of artificial intelligence in agriculture: mapping the agricultural AI literature. *AI Soc.* 38, 2473–2485. doi: 10.1007/s00146-021-01377-9

Saleem, M. H., Potgieter, J., and Arif, K. M. (2021). Automation in agriculture by machine and deep learning techniques: A review of recent developments. *Precis. Agric.* 22, 2053–2091. doi: 10.1007/s11119-021-09806-x

Sha, L., Rakovic, M., Whitelock-Wainwright, A., Carroll, D., Yew, V. M., Gasevic, D., et al. (2021). "Assessing algorithmic fairness in automatic classifiers of educational forum posts," in *Artificial Intelligence in Education: 22nd International Conference, AIED 2021*, Utrecht, The Netherlands, June 14–18, 2021, Proceedings, Part I 22. 381–394. doi: 10.1007/978-3-030-78292-4_31

Shen, Y., Song, K., Tan, X., Li, D., Lu, W., and Zhuang, Y. (2024). Hugginggpt: Solving ai tasks with chatgpt and its friends in hugging face. *Adv. Neural Inf. Process. Syst.* 36. doi: 10.48550/arXiv.2303.17580

Shi, B., Wu, Z., Mao, M., Wang, X., and Darrell, T. (2024). "When do we not need larger vision models?," in *European Conference on Computer Vision*. 444–462 (Cham: Springer Nature Switzerland). doi: 10.1007/978-3-031-73242-3_25

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. doi: 10.48550/arXiv.1409.1556

Stella, F., Della Santina, C., and Hughes, J. (2023). How can LLMs transform the robotic design process? *Nat. Mach. Intell.* 5, 561–564. doi: 10.1038/s42256-023-00669-7

Sun, C., Shrivastava, A., Singh, S., and Gupta, A. (2017). "Revisiting unreasonable effectiveness of data in deep learning era," in *Proceedings of the IEEE international conference on computer vision*. 843–852.

Sundermeyer, M., Schlüter, R., and Ney, H. (2012). Lstm neural networks for language modeling. *Interspeech* 2012, 194–197. doi: 10.21437/Interspeech.2012-65

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9. doi: 10.1109/CVPR.2015.7298594

Tao, Y., and Zhou, J. (2017). Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking. *Comput. Electron. Agric.* 142, 388–396. doi: 10.1016/j.compag.2017.09.019

Team, G., Anil, R., Borgeaud, S., Alayrac, J. B., Yu, J., Soricut, R., et al. (2023). Gemini: a family of highly capable multimodal models. doi: 10.48550/arXiv.2312.11805

Team, G., Kamath, A., Ferret, J., Pathak, S., Vieillard, N., Merhej, R., et al. (2025). Gemma 3 technical report. doi: 10.48550/arXiv.2503.19786

Team, G., Mesnard, T., Hardin, C., Dadashi, R., Bhupatiraju, S., Pathak, S., et al. (2024). Gemma: Open models based on gemini research and technology. doi: 10.48550/arXiv.2403.08295

Thenmozhi, K., and Reddy, U. S. (2019). Crop pest classification based on deep convolutional neural network and transfer learning. *Comput. Electron. Agric.* 164, 104906. doi: 10.1016/j.compag.2019.104906

Tokekar, P., Vander Hook, J., Mulla, D., and Isler, V. (2016). Sensor planning for a symbiotic UAV and UGV system for precision agriculture. *IEEE Trans. Robotics* 32, 1498–1511. doi: 10.1109/TRO.2016.2603528

Tripathy, P., Baylis, K., Wu, K., Watson, J., and Jiang, R. (2024). Investigating the segment anything foundation model for mapping smallholder agriculture field boundaries without training labels. doi: 10.48550/arXiv.2407.01846

Tzachor, A., Devare, M., King, B., Avin, S., and Ó hÉigeartaigh, S. (2022). Responsible artificial intelligence in agriculture requires systemic understanding of risks and externalities. *Nat. Mach. Intell.* 4, 104–109. doi: 10.1038/s42256-022-00440-4

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* 30. doi: 10.48550/arXiv.1706.03762

Veena, G., Kanjirangat, V., and Gupta, D. (2023). AGRONER: An unsupervised agriculture named entity recognition using weighted distributional semantic model. *Expert Syst. Appl.* 229, 120440. doi: 10.1016/j.eswa.2023.120440

Visentin, F., Cremasco, S., Sozzi, M., Signorini, L., Signorini, M., Marinello, F., et al. (2023). A mixed-autonomous robotic platform for intra-row and inter-row weed removal for precision agriculture. *Comput. Electron. Agric.* 214, 108270. doi: 10.1016/j.compag.2023.108270

Wang, W., Dai, J., Chen, Z., Huang, Z., Li, Z., Zhu, X., et al. (2023). "Internimage: Exploring large-scale vision foundation models with deformable convolutions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14408–14419. doi: 10.1109/CVPR52729.2023.01385

Wang, J., Ren, S., Zhang, Z., et al. (2019). Research progress on unmanned aerial vehicle for ecological remote sensing monitoring based on bibliometric assessment. *Trop. Geogr.* 39 (4), 616–624. doi: 10.13284/j.cnki.rddl.003157

Wei, T., Zhao, L., Zhang, L., Zhu, B., Wang, L., Yang, H., et al. (2023). Skywork: A more open bilingual foundation model. doi: 10.48550/arXiv.2310.19341

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., et al. (2021). Ethical and social risks of harm from language models. doi: 10.48550/arXiv.2112.04359

Wiles, O., Gowal, S., Stimberg, F., Alvise-Rebuffi, S., Ktena, I., Dvijotham, K., et al. (2021). A fine-grained analysis on distribution shift. doi: 10.48550/arXiv.2110.11328

Wu, J., Gan, W., Chen, Z., Wan, S., and Yu, P. S. (2023a). "Multimodal large language models: A survey," in *2023 IEEE International Conference on Big Data (BigData)*. doi: 10.1109/BigData59044.2023.10386743

Wu, J., Hovakimyan, N., and Hobbs, J. (2023b). Genco: An auxiliary generator from contrastive learning for enhanced few-shot learning in remote sensing. doi: 10.48550/arXiv.2307.14612

Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., et al. (2023c). Bloomberggpt: A large language model for finance. doi: 10.48550/arXiv.2303.17564

Wu, C., Lin, W., Zhang, X., Zhang, Y., Xie, W., and Wang, Y. (2024). PMC-LLaMA: toward building open-source language models for medicine. *Journal of the American Medical Informatics Association* 31 (9), 1833–1843. doi: 10.1093/jamia/ocae045

Xu, H., Man, Y., Yang, M., Wu, J., Zhang, Q., and Wang, J. (2023). Analytical insight of earth: a cloud-platform of intelligent computing for geospatial big data. doi: 10.48550/arXiv.2312.16385

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., et al. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *Br. J. Educ. Technol.* 55, 90–112. doi: 10.1111/bjet.13370

Yang, X., Dai, H., Wu, Z., Bist, R., Subedi, S., Sun, J., et al. (2023c). Sam for poultry science. doi: 10.48550/arXiv.2305.10254

Yang, Q., Du, X., Wang, Z., Meng, Z., Ma, Z., and Zhang, Q. (2023b). A review of core agricultural robot technologies for crop productions. *Comput. Electron. Agric.* 206, 107701. doi: 10.1016/j.compag.2023.107701

Yang, J., Gao, M., Li, Z., Gao, S., Wang, F., and Zheng, F. (2023a). Track anything: Segment anything meets videos. doi: 10.48550/arXiv.2304.11968

Yang, G., Li, Y., He, Y., Zhou, Z., Ye, L., Fang, H., et al. (2024). Multimodal large language model for wheat breeding: a new exploration of smart breeding. doi: 10.48550/arXiv.2411.15203

Yao, S., Zhang, C., Ping, J., and Ying, Y. (2024). Recent advances in hydrogel microneedle-based biofluid extraction and detection in food and agriculture. *Biosens. Bioelectronics* 250, 116066. doi: 10.1016/j.bios.2024.116066

Ye, J., Xu, H., Liu, H., Hu, A., Yan, M., Qian, Q., et al. (2024). mplug-owl3: Towards long image-sequence understanding in multi-modal large language models. doi: 10.48550/arXiv.2408.04840

Ye, Q., Xu, H., Xu, G., Ye, J., Yan, M., Zhou, Y., et al. (2023). mplug-owl: Modularization empowers large language models with multimodality. doi: 10.48550/arXiv.2304.14178

Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., et al. (2023). A survey on multimodal large language models. doi: 10.48550/arXiv.2306.13549

Yuan, Y., Chen, L., Wu, H., and Li, L. (2022). Advanced agricultural disease image recognition technologies: A review. *Inf. Process. Agric.* 9, 48–59. doi: 10.1016/j.inpa.2021.01.003

Zhang, J., Huang, Y., Pu, R., Gonzalez-Moreno, P., Yuan, L., Wu, K., et al. (2019). Monitoring plant diseases and pests through remote sensing technology: A review. *Comput. Electron. Agric.* 165, 104943. doi: 10.1016/j.compag.2019.104943

Zhang, H., Li, X., and Bing, L. (2023a). Video-llama: An instruction-tuned audio-visual language model for video understanding. doi: 10.48550/arXiv.2306.02858

Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., et al. (2022b). Opt: Open pre-trained transformer language models. doi: 10.48550/arXiv.2205.01068

Zhang, N., Wu, H., Zhu, H., Deng, Y., and Han, X. (2022a). Tomato disease classification and identification method based on multimodal fusion deep learning. *Agriculture* 12, 2014. doi: 10.3390/agriculture12122014

Zhao, B., Jin, W., Del Ser, J., and Yang, G. (2023a). ChatAgri: Exploring potentials of ChatGPT on cross-linguistic agricultural text classification. *Neurocomputing* 557, 126708. doi: 10.1016/j.neucom.2023.126708

Zhao, L., Zhang, L., Wu, Z., Chen, Y., Dai, H., Yu, X., et al. (2023b). When brain-inspired ai meets agi. *Meta-Radiology* 1, 100005. doi: 10.1016/j.metrad.2023.100005

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., et al. (2023c). A survey of large language models. doi: 10.48550/arXiv.2303.18223

Zheng, C., and Li, H. (2023). The prediction of collective Economic development based on the PSO-LSTM model in smart agriculture. *PeerJ. Comput. Sci.* 9, e1304. doi: 10.7717/peerj-cs.1304

Zhong, J., Liu, Z., and Chen, X. (2023). Transformer-based models and hardware acceleration analysis in autonomous driving: A survey. doi: 10.48550/arXiv.2304.10891

Zhou, J., Li, J., Wang, C., Wu, H., Zhao, C., and Teng, G. (2021). Crop disease identification and interpretation method based on multimodal deep learning. *Comput. Electron. Agric.* 189, 106408. doi: 10.1016/j.compag.2021.106408

Zhou, L., Pan, S., Wang, J., and Vasilakos, A. V. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing* 237, 350–361. doi: 10.1016/j.neucom.2017.01.026

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 September 2025
DOI 10.3389/fpls.2025.1549896

Check for updates

OPEN ACCESS

EDITED BY
Jingcheng Zhang,
Hangzhou Dianzi University, China

REVIEWED BY
Klára Kosová,
Crop Research Institute (CRI), Czechia
Yingying Dong,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Zhaoyu Zhai
zhaoyu.zhai@njau.edu.cn

RECEIVED 22 December 2024
ACCEPTED 27 August 2025
PUBLISHED 15 September 2025

CITATION
Zhu S, Li H, Zou S, Xu H and Zhai Z (2025) FHB-Net: a severity level evaluation model for wheat *Fusarium* head blight based on image-level annotated aerial RGB images. *Front. Plant Sci.* 16:1549896. doi: 10.3389/fpls.2025.1549896

COPYRIGHT
© 2025 Zhu, Li, Zou, Xu and Zhai. 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.

FHB-Net: a severity level evaluation model for wheat *Fusarium* head blight based on image-level annotated aerial RGB images

Shuxin Zhu[1], Huayong Li[2], Shun Zou[1], Huanliang Xu[1] and Zhaoyu Zhai[1]*

[1]College of Smart Agriculture (College of Artificial Intelligence), Nanjing Agricultural University, Nanjing, China, [2]Institute of Germplasm Resources and Biotechnology, Jiangsu Academy of Agricultural Sciences, Nanjing, China

A leading concern for global wheat production, *Fusarium head blight* (FHB) can cause yield losses of up to 50% during severe epidemics. The cultivation of FHB-resistant wheat varieties is widely acknowledged as a highly effective and economical approach to disease management. The disease resistance breeding task depends on accurately evaluating the severity level of FHB. However, existing approaches may fail to distinguish among healthy and slightly infected wheats due to insufficient fine-grained feature learning, resulting in unreliable predictions. To tackle these challenges, this paper proposed the FHBNet model for evaluating the severity level of FHB under an end-to-end manner by simply using image-level annotated RGB images. In total, 6035 RGB aerial images taken from the wheat field were used to construct the dataset and each image was labelled by the light, moderate, or severe category. In FHBNet, we first utilized the multi-scale criss-cross attention (MSCCA) block to capture the global contextual relationships from each pixel, thereby modelling the spatial context of wheat ears. Furthermore, in order to accurately locate small lesions in wheat ears, we applied the bi-level routing attention (BRA) module, which suppressed the most irrelevant key-value pairs and only retained a small portion of interested regions. The experimental results demonstrated that FHBNet achieved an accuracy of 79.49% on the test se5t, surpassing the mainstream neural networks like MobileViT, MobileNet, EfficientNet, RepLkNet, ViT, and ConvNext. Moreover, visualization heatmaps revealed that FHBNet can accurately locate the FHB lesions under complex conditions, e.g., varying severity levels and illuminations. This study validated the feasibility of rapid and nondestructive FHB severity level evaluation with only image-level annotated aerial RGB images as an input, and the research result of this study can potentially accelerate the disease resistance breeding task by providing high-throughput and accurate phenotype analysis.

KEYWORDS

***Fusarium* head blight, severity evaluation, unmanned aerial vehicle, deep learning, attention mechanism**

1 Introduction

Globally, wheat stands as a pivotal crop in both human and animal sustenance, contributing significantly to food security (Gruber, 2017). Its protein composition boasts ample amino acids to fulfill daily human dietary requirements (Chen, 2020). *Fusarium* Head Blight (FHB), primarily induced by *Fusarium graminearum*, poses a substantial threat to wheat, ranking among the most deleterious fungal ailments (Sari et al., 2020). The pathogen infiltrates the spikelet during flowering, causing premature desiccation and discoloration, thereby precipitating a notable decline in the wheat yield. Simultaneously, the toxins will also compromise the immune responses of both humans and animals (Femenias et al., 2020; Gilbert and Tekauz, 2000). Breeding the FHB-resistant cultivars stands as an effective strategy in combating this disease. However, the breeding process requires to evaluate the disease resistance of hundreds of wheat varieties. Exisitng methods of evaluating the disease severity level predominantly rely on visual observations by manually recording the proportion of infected wheat ears to the total ears in the field. This approach is labor-intensive and time-consuming. Consequently, there arises an urgent need for efficient approaches to assess the FHB severity levels.

In recent years, the analysis of plant phenotype has been driven by diverse neural network architectures (Mochida et al., 2019; Toda and Okura, 2019; Singh et al., 2018). Deep learning and computer vision have revolutionized the field of plant phenotyping by providing advanced tools for analyzing and interpreting various traits. These technologies have leveraged the power of artificial intelligence, which mimics humans to recognize patterns and make decisions. By learning from a sufficient amount of image data, neural networks can identify subtle phenotypic traits that are difficult for human observers to detect. This capability has proven particularly useful in precision agriculture, where timely and accurate assessment of plant characteristics can lead to better crop management and disease control strategies.

Various publications have reported the applicability of neural networks to wheat disease detection. Hamila et al. (2024) developed a 3D convolutional neural network to detect wheat FHB symptoms and accurately estimate the number of healthy and infected spikes. Roble et al. (2023) used classification algorithms to directly estimate the FHB scores without prior segmentation or detection of diseased wheat spikes. However, this study achieved a low accuracy of approximately 52% in discriminating the severity levels of FHB. Wang et al. (2023) developed a deep learning-based multi-model fusion system for real-time accurate diagnosis of wheat FHB, achieving high precision in wheat spike and lesion segmentation. The effectiveness of incorporating HSV color features as weighting factors in the wheat spike grading model was verified. Gao et al. (2022) collected images of individually isolated wheat ears and applied a pre-trained network for FHB prediction. Zhang et al. (2022a) developed a detection method for detecting the wheat FHB by integrating spectral and image features from raw unmanned aerial vehicle hyperspectral imaging data. Weng et al. (2021) proposed a method using HSI and deep learning networks to identify FHB severity levels by selecting specific wavelengths and applying a residual attention convolutional neural network, enabling high-precision classification of healthy and infected wheat kernels. Mi et al. (2020) developed C-DenseNet by utilizing the convolutional block attention mechanism (CBAM) and DenseBlock to extract traits of wheat stripe rust at varying severity levels. The attention heatmap confirmed that the proposed model was able to extract fine-grained features, thus accurately distinguishing disease levels with similar symptoms. Zhang et al. (2019) established a pulse coupled neural network to segment wheat ears infected with FHB. Nevertheless, this investigation solely addressed one spike within the image, rendering it impracticality for high-throughput detection in complex field conditions.

To boost the plant disease detection and classification accuracy, integrating the attention mechanism to deep learning models is one of the favorable strategies. Wei et al. (2025) introduced the dual branch channel attention and efficient spatial attention modules to ShuffleNetV2 to enhance feature learning of few-shot samples. The integration of the attention mechanisms boosted the classification accuracy by around 3%. Srinivasan et al. (2025) proposed DBA-ViNet for fruit disease detection and classification. Specifically, DBA-ViNet utilized a dual branch attention to capture global contextual information and fine-grained local lesion details simultaneously. Zang et al. (2025) developed a segmentation algorithm, RSE-Swin Unet, for wheat powdery mildew. The SENet attention mechanism was introduced to the baseline model Swin-UNet for capturing global and local features. The experimental result indicated that SENet was beneficial to identify irregular and small lesions. Bhujel et al. (2022) examined the effect of multiple attention mechanisms, including CBAM, Squeeze-and-Excitation (SE), Self-Attention (SA), and dual attention on the tomato leaf disease classification task. Recent works also reported more advanced attention mechanisms, such as high-frequency attention (Tian et al., 2025) context-guided attention (Yan and Li, 2025), multi-dimensional attention (Yang et al., 2025), cross-branch channel attention (Li et al., 2025), coordinate attention (Gu and Liu, 2025), etc. All these works confirmed the potential of integrating the attention mechanisms to deep learning models to achieve higher accuracy of detecting and classifying the plant disease.

Despite significant improvements in wheat FHB detection, most exisiting methods either collected images of isolated wheat spikes under constrained conditions or used algorithms to detect the percentage of spike and lesion pixels. However, these methods were not targeting to handling the real-world application scenarios and had weak generalization due to the scale differences of wheat spikes. Furthermore, the FHB severity evaluation task can be influenced by multiple factors. First, the quality of the data source is concerning. RGB images may be collected from different shooting angles under various illumination conditions. For instance, a healthy wheat spike taken under strong illumination usually has the similar visual characteristics as the one infected by FHB, both of which would appear color changes. Second, the FHB infected wheat spikes are usually distributed within the field plot unevenly and may be occluded by leaves and healthy wheat spikes. Therefore, aggregating the contextual information to enhance the

feature learning is of great necessity. Although existing approaches have integrated attention mechanisms to tackle this issue, these attention mechanisms still experience weak generalization for FHB detection, particularly, may fail to adapt to the complex scenes in the wheat field. Third, at the early stage of FHB, the small spots on the spikes can be considered as small objects, which brings great challenges for the detection tasks.

To address the above-mentioned challenges in the FHB severity level evaluation task, we proposed an end-to-end deep learning model, namely FHBNet, which received an image-level annotated RGB image as an input and provided the severity level value of FHB as an output. The novel contributions of this study were summarized as follows.

- To enhance the contextual information of wheat ears, we developed a multi-scale criss-cross attention (MSCCA) block. It first replaced the standard 3×3 convolution kernels with ones that had a larger receptive field, enabling to effectively capture the long-term dependencies. Additionally, we introduced a multi-scale convolution module to adapt to variations in the orientation and scale distribution of wheat spikes, thus enhancing the spatial feature extraction. Concurrently, the MSCCA block performed a thorough semantic analysis of each pixel along vertical and horizontal directions, allowing each pixel to capture dependencies from its neighboring pixels, which greatly enhanced the pixel-level representational ability.
- As wheat ears were considered as small objects, we adopted the bi-level routing attention (BRA) block to further aid FHBNet to focus on the key regions within the feature map. Irrelevant key-value pairs were filtered out at a coarse region level. The BRA block was helpful to distinguish the FHB lesions among the complex background and locate the infected wheat ears more precisely.
- We conducted extensive comparative evaluations with several mainstream neural networks, including MobileViT, MobileNet, EfficientNet, RepLkNet, ViT, and ConvNext. The experimental results demonstrated the effectiveness of FHBNet in all evaluation indicators. Additionally, we validated the robustness of FHBNet under various settings (e.g., severity levels and illuminations). Through the heatmap visualization, it was evident that FHBNet can effectively address the detection challenges and achieve promising performance.

2 Materials and methods

2.1 Data acquisition and preprocess methods

The experiment site was located in Huai'an, Jiangsu Province, China (119°01′20.74″E, 33°36′59.45″N). This region has a North subtropical humid monsoon climate with an average temperature at 14.5°C and an average rainfall at 960 mm. The wheat variety chose in this study was Huaimai 33 as the experiment material, provided by the College of Plant Protection, Nanjing Agricultural University. The experiment field was divided into 125 plots and each plot has 5.0 m both in length and width (see Figure 1A). All plots received the same agricultural management (e.g., irrigation, fertilization, weeding, etc.) except the disease control measure. FHB naturally occurred, rather than artificial inoculation. We applied different fungicide treatments to achieve each FHB disease severity level. For the control group, no disease control measure was adopted, resulting in the severe disease level. For the moderate and light disease levels, we applied the 200 g/hm^2 and 400 g/hm^2 of Phenamacril suspension concentrate, respectively.

The field data was acquired on 24 May, 2024, when the wheat was at the filling stage. We adopted a DJI Mavic 3 Pro unmanned aerial vehicle (UAV) for acquiring the RGB imagery. This UAV was equipped with a Hasselblad camera (4/3CMOS, effective pixels 20 MP), which fulfilled the requirement of data acquisition. The flight height of the UAV was set at 20 m and the flight speed was set at 1.5 m/s. After obtaining the raw video data, we manually filtered and cropped the video frames. To increase the size of the dataset, we also utilized an FHB dataset from the research contributed by Roble et al. (2023), resulting in a total of 6035 images. Each image was assigned by a category from light, moderate, and severe (see Figure 1B). Table 1 listed the classification criteria and the number of samples included in all categories.

During training and validation, the images were resized to 224×224. To enhance the stability of the model and accelerate convergence, we normalized the data samples by adjusting features of different dimensions to similar ranges and pixel values to a specific interval. The dataset was divided to a training dataset and a test dataset by the ratio of 8:2.

2.2 FHBNet

We proposed a novel model, namely FHBNet, that can accurately identify infected wheat spikes and determine the severity level of FHB. The overall architecture was depicted in Figure 2. The original image first entered the Stem layer, where a series of convolutional layers were stacked. After passing through the first 3×3 convolutional layer with a sampling rate of 2, we used a depth-wise separable 3×3 convolutional layer to capture the low-level features. Next, a 1×1 convolutional layer and a depth-wise separable 3×3 convolutional layer was employed to perform the downsampling operations. The image then progressed through the Stage layer, which was the core of FHBNet. It was responsible to locate the wheat spikes and extract the FHB features. The Stage layer included the MSCCA and BRA blocks, which will be elaborated in subsequent sections. The number of blocks in each stage was presented in Supplementary Table S1. It was worth mentioning that Stage 3 contained 18 MSCCA blocks, more than any other stage. This was due to the reason that in deep network architectures, deeper layers were often responsible for extracting more abstract features. Located in the deeper layers of FHBNet, Stage 3 dealt with more complex feature information. The MSCCA block, with its

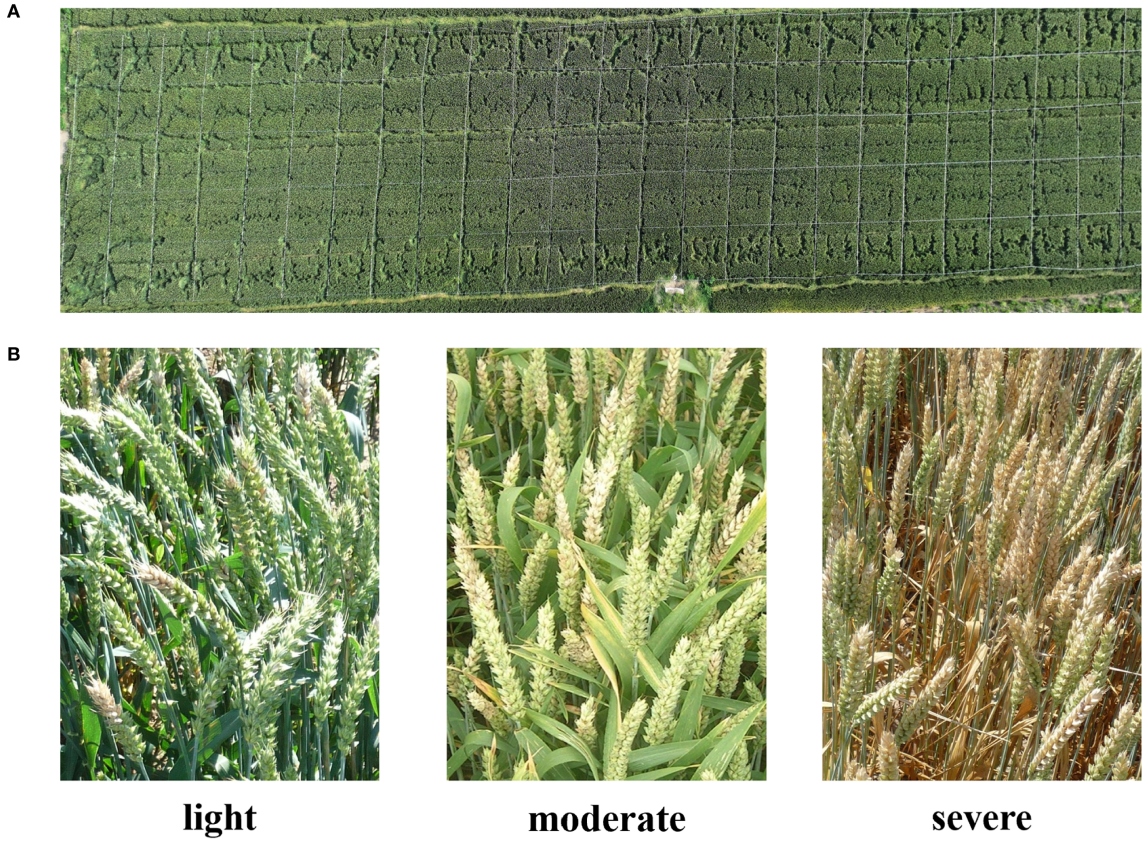


FIGURE 1
Data acquisition and labelling: **(A)** the plot setting. **(B)** example images of each category.

multi-scale and attention mechanism, effectively enhanced the model's ability to abstract and represent features, enabling FHBNet to better understand and process the high-level semantic information. Last, the Transition layer served as a bridge between each stage, acting as an independent downsampling layer that doubled the channel dimensions and halved the image size.

2.3 MSCCA block

To better capture the information at various scales and model spatial contextual semantic relationships, we designed the MSCCA block and embedded it in all the Stage layers. The detailed design of the MSCCA block was shown in Figure 3A. In general, the MSCCA block consisted of several sub modules, mainly including the feature pyramid split attention (FPSA) and criss-cross attention (CCA) modules. The former was deployed to extract the multi-scale features, while the latter was helpful to enhance the extraction of the semantic information. Besides, we heavily employed 1×1 convolution operators in the MSCCA block to increase the model depth and introduce the non-linearity. Meanwhile, the 1×1 convolution operator can facilitate the cross-channel communication. In both FPSA and CCA modules, the residual connection was considered to tackle the network degradation issue (i.e., the gradient vanishing issue) (He et al., 2016). Additionally, we utilized large-kernel depth-wise separable convolutions to further broaden the receptive field. It was important to note that the size of the large-kernel convolutions varied in the MSCCA blocks across each stage. Specifically, the dimensions of the large kernels in Stages 1-4 were 31×31, 29×29, 27×27, and 13×13, respectively (Ding et al., 2022b).

Inspired by the design of EPSANet (Zhang et al., 2022b), we introduced the FPSA module in the MSCCA block, as shown in Figure 3B. Let H, W, and C denoted the height, width, and channels of the feature map, respectively. We first divided the feature map generated by the large kernel convolution into four branches along the channel dimension, represented by $[Z_0, Z_1, Z_2, Z_3]$. The convolutional kernels with different sizes were assigned to generate feature maps at various scales. To handle input tensors of different kernel sizes without increasing computational costs, the group convolution method was adopted. This means that the filters in the specific input channels only interact within these channels, greatly reducing the total number of trainable parameters. The multi-scale feature map was generated by Equation 1 as follows.

$$f_i = \mathrm{Conv}(Z_i, (K_i \times K_i, G_i)) \tag{1}$$

where $\mathrm{Conv}(\cdot)$ denoted the group convolution operation. Z_i denoted the i^{th} feature map branch. K_i denoted the size of the convolutional kernel in the i^{th} branch, and G_i denoted the group size in the i^{th} branch. K_0, K_1, K_2, and K_3 are 3×3, 5×5, 7×7, and 9×9, respectively. G_0, G_1, G_2, and G_3 are 1, 4, 8, and 16, respectively. f_i denoted the feature maps.

TABLE 1 The severity annotation scale and dataset distribution.

FHB severity	Infested spike area/%	The number of samples
Light	0-5	1754
Moderate	5-14	2012
Severe	14-100	2269

The feature maps generated from each branch were then concatenated along the channel dimension, formulated by Equation 2.

$$f = \text{Concat}(f_0, f_1, f_2, f_3) \tag{2}$$

where Concat denoted the concatenation of feature maps in a serial manner.

Subsequently, the attention weight vectors with four scales were obtained by extracting the channel attention weight information (Hu et al., 2018). This operation enabled the FPSA module to capture adequate contextual information in the feature map. In particular, each channel attention vector was concatenated and recalibrated through the Softmax function. The recalibrated weights and the corresponding feature maps were undergoing element-wise multiplication to produce a feature map with attention weights, enhancing its representational capability across multiple scales. The relevant formulas are defined by Equations 3, 4.

$$S_i = \text{SEWeight}(f_i) \tag{3}$$

$$F = f \otimes \text{Softmax}(\text{Concat}(S_0, S_1, S_2, S_3)) \tag{4}$$

where S_i denoted the attention weight vectors for each scale, $\otimes$ denoted element-wise multiplication. F denoted the feature map obtained after processing by the FPSA module.

After passing through the FPSA module, the feature map was then processed by the semantic feature enhancement module (i.e., the CCA module). Inspired by feedforward networks (FFNs) widely used in transformers (Kong et al., 2022; Liu et al., 2021) and MLPs (Ding et al., 2022a; Tolstikhin et al., 2021), we employed a CNN-style block composed of residual connections, batch normalization (BN), two 1×1 layers, with the addition of criss-cross attention at the end (Huang et al., 2019). Compared to the standard FFN, which used layer normalization before the fully-connected layers, BN had an advantage in that it can be fused into convolutions for efficient inference. BN normalizes the output of a layer for a given mini-batch by centering them around a mean of zero and scaling them to a standard deviation of one. BN effectively address the issue of internal covariate shift by ensuring that the inputs to each layer have a consistent distribution, regardless of the variations in the previous layer's outputs.

The CCA module gathered the contextual information from each pixel along the vertical and horizontal paths. As shown in Figure 3C, the CCA module initially applied two convolutional

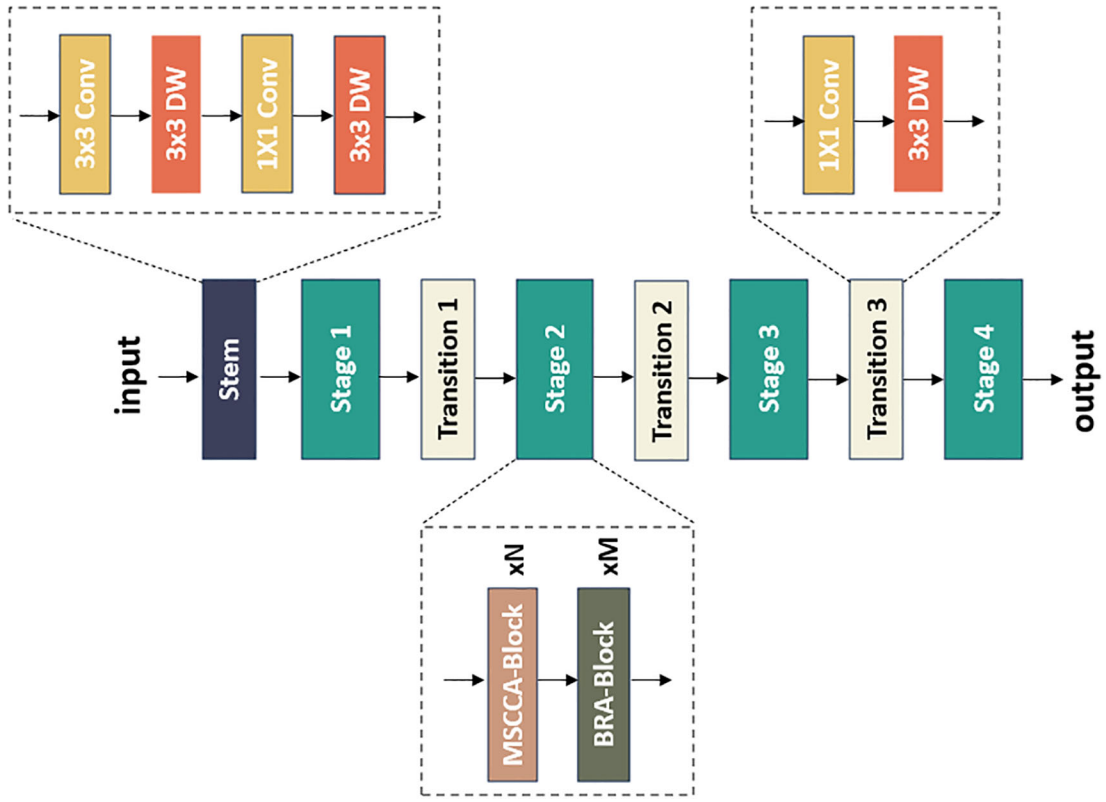


FIGURE 2
The architecture of the proposed FHBNet.

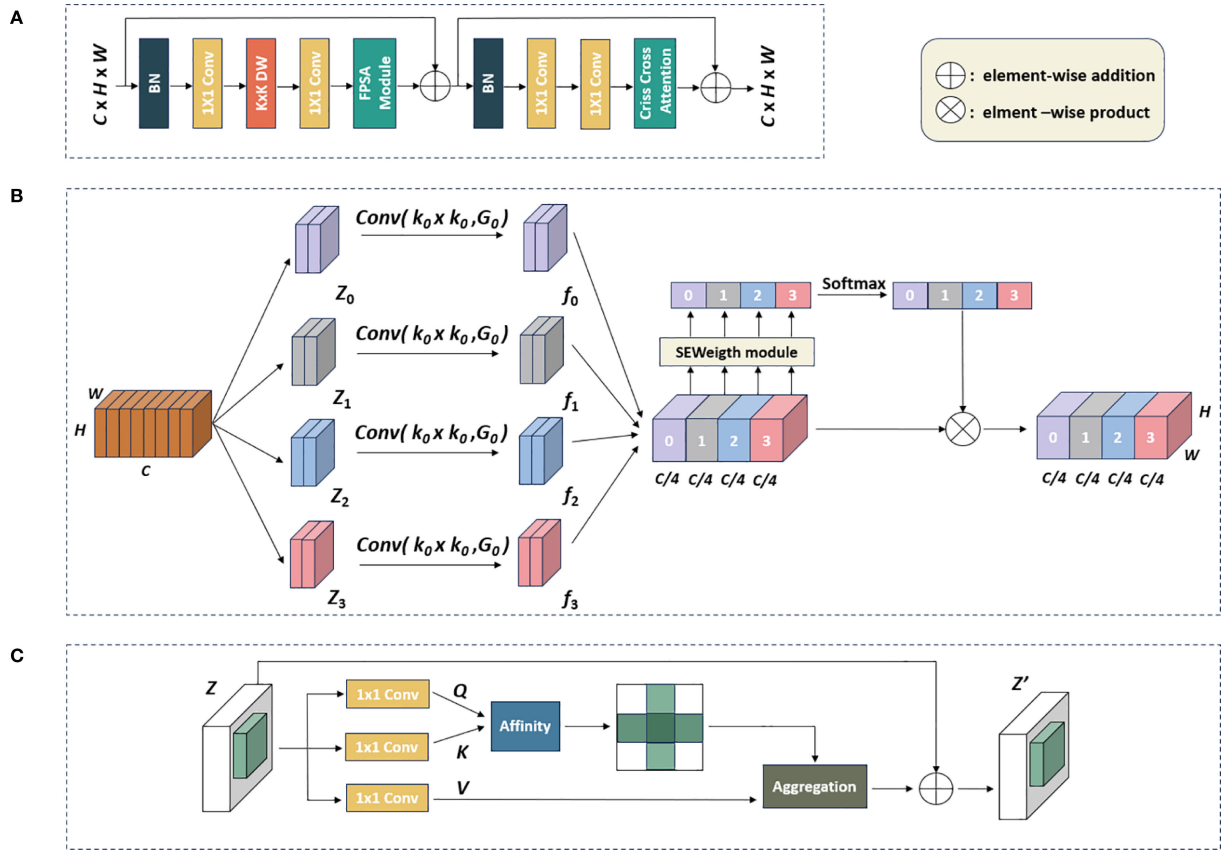


FIGURE 3
The workflow of the MSCCA block: **(A)** Structure of the MSCCA block; **(B)** the FPSA module; **(C)** the criss-cross attention module.

layers with 1×1 filters on $Z \in R^{C\times H\times W}$, producing two feature maps, Q and K, where $\{K, Q\} \in R^{C'\times H\times W}$ and C' was the number of channels. At each position u in Q, a vector $Q_u \in R^{C'}$ was obtained. Feature vectors from K corresponding to the same column or row at position u were also extracted, forming a set $\Omega_u \in R^{(H+W-1)\times C'}$. The attention map $A \in R^{(H+W-1)\times H\times W}$ was computed by applying a Softmax layer (that converted a vector of raw scores into a vector of probabilities) over the Affinity matrix (that quantified the similarity between two data points), which was calculated by Equation 5:

$$d_{i,u} = Q_u \Omega_{i,u}^{\mathrm{T}} \tag{5}$$

where, $d_{i,u} \in D$ represented the correlation between feature Q_u and $\Omega_{i,u}$, where $\Omega_{i,u} \in R^{C'}$ was the i^{th} element of Ω_u, and $i = [1, \ldots, |\Omega_u|]$, with $D \in R^{(H+W-1)\times H\times W}$.

A third convolutional layer applied to Z generated $V \in R^{C\times H\times W}$ for feature adaptability. At each position u in V, a feature vector $V_u \in R^{C}$ and a set $\Phi_u \in R^{(H+W-1)\times H\times W}$ were obtained. The set Φ_u comprised feature vectors from V that were in the same column or row as position u.

The contextual information was aggregated by Equation 6:

$$Z_u^{'} = \sum_{i=0}^{H+W-1} A_{i,u} \Phi_{i,u} + Z_u \tag{6}$$

where, $Z_u^{\}$ denoted the feature vector at position u in $Z' \in R^{C\times H\times W}$, $\Phi_{i,u}$ denoted the i^{th} element of Φ_u, where $i = [0, \ldots, H + W - 1]$, and $A_{i,u}$ denoted the scalar value at channel i and position u in the attention map A.

2.4 BRA block

Considering the complexity of backgrounds in the wheat field, and wheat spikes were small-scale targets, it was challenging to accurately locate the position of diseased wheat spikes. Additionally, the ability of most existing models to suppress background information was insufficient, which challenged the focus of the model on essential input features and increased its vulnerability to irrelevant background noise. To tackle this issue and enhance the model's capability to detect small targets like wheat spikes and concentrate on the pathological features of FHB on these spikes, we incorporated a module, known as BiFormer, into the Stage section of FHBNet (Zhu et al., 2023). BiFormer is a recent computer vision algorithm integrated the attention mechanism from the Transformer architecture. It proposed a dynamic sparse attention mechanism by determining a small, relevant set of key pixels for each query pixel to attend to, rather than having every pixel attend to every other pixel.

The core of constructing the BRA block was the bi-level routing attention. As depicted in Figure 4A, the BRA block comprised a 3×3 depth-wise separable convolution, two layers of layer normalization (LN), a bi-level routing attention, three residual connections, and a two-layer MLP with an expansion ratio of 3. Here, the depth-wise convolution (denoted by DW in Figure 4A) was utilized to decrease the number of parameters and the computational complexity of FHBNet, while LN aided in accelerating the training process and enhancing the generalization of FHBNet. The MLP further processed and adjusted attention weights to intensify the focus of FHBNet on high-level features. Additionally, a residual connection between the original input and the final output was introduced, in addition to the original three residual connections, to further improve the model's generalization capability. The computational process for the BRA block was outlined in (Equations 7–10).

$$\hat{z}^{l-1} = DW(z^{l-1}) + z^{l-1} \tag{7}$$

$$\hat{z}^{l} = BRA(LN(\hat{z}^{l-1})) + \hat{z}^{l-1} \tag{8}$$

$$z^{l} = MLP(LN(\hat{z}^{l})) + \hat{z}^{l} \tag{9}$$

$$z^{l} = z^{l} + \hat{z}^{0} \tag{10}$$

where $\hat{z}^0$ represented the initial input, $\hat{z}^{l-1}$, $\hat{z}^l$, and z^l represented the outputs of the depth-wise convolution, the BRA module, and the MLP module, respectively.

The most critical component of the BRA block was the bi-level routing attention, a novel dynamic sparse attention mechanism. BRA first utilized query adaptivity to filter the input feature map, removing the least relevant key-value pairs in the coarse-grained areas, thus effectively identifying highly relevant key-value pairs for attention computation later. This not only greatly reduced the computational and storage resource consumption, but also enhanced the perception of the input content. The BRA block cares about only a small subset of relevant key-value pairs under a query adaptive manner, thus avoiding distraction from irrelevant ones.

As depicted in Figure 4B, the input feature map $X \in R^{H\times W\times C}$ was initially partitioned into $S \times S$ sub-regions, also known as patches, each containing HW/S^2 feature vectors. Subsequently, X underwent a change in shape to obtain $X^r \in R^{S^2\times(HW/S^2)\times C}$. X^r was then subjected to linear transformation to obtain three feature vectors Q, K, and V. The formulas for calculating Q, K, V can be expressed by Equations 11–13, respectively.

$$Q = X^r W^q \tag{11}$$

$$K = X^r W^k \tag{12}$$

$$V = X^r W^v \tag{13}$$

Subsequently, a graph was constructed to obtain attention relationships between regions, determining the relevant regions. The specific implementation process was as follows. For each region, Q and V were processed by region averaging to obtain region-level Q^r and $K^r \in R^{S^2\times C}$. Then, the dot product of Q^r and K^r was computed to obtain the adjacency matrix $A^r \in R^{S^2\times S^2}$, which measured the correlation between regions. The formula of calculating A^r were expressed by Equation 14.

$$A^r = Q^r(K^r)^T,\ A^r \in R^{S^2\times S^2} \tag{14}$$

Next, at the coarse-grained level, the least relevant tokens in A^r were pruned, retaining the top k most relevant regions in A^r to obtain the routing index matrix $I^r \in N^{S^2\times k}$. The formula of calculating I^r were expressed by Equation 15.

$$I^r = topkIndex(A^r) \tag{15}$$

Subsequently, token-to-token attention was used at a fine-grained level. For a query in region i, this attention focused only on the k routing regions indexed by $I^r(i,1), I^r(i,2), \ldots, I^r(i,k)$ in the routing index matrix I^r, and collected all the K and V tensors from these k regions to obtain K^g and V^g. The formula of calculating K^g and V^g were expressed by Equation 16.

$$K^g = Gather(K, I^r), V^g = Gather(V, I^r) \tag{16}$$

Finally, attention processing was applied to the collected K^g and V^g, and a local context enhancement term $LCE\ (V)$ was added to obtain the output tensor O. The formula of calculating O were expressed by Equation 17.

$$O = Attention(Q, K^g, V^g) + LCE(V) \tag{17}$$

The function $LCE(\cdot)$ was parameterized using a depth-wise convolution.

2.5 Experimental environment and training details

All experiments in this study were conducted under the same hardware and software configurations to ensure fairness. The experiments were performed on a server running Ubuntu 20.04, equipped with an Intel® Core™ i7 central processing unit (CPU) and an NVIDIA GeForce RTX 3090 with a GPU memory size of 24 GB. Additionally, we utilized CUDA 11.3, cuDNN 8.9.0, and PyTorch 1.10.0 as the deep learning framework.

We utilized the Adam optimizer (Kingma and Ba, 2014) for training, with the momentum set at 0.9. Additionally, in 250 epochs, we set the batch size to 36 and employed a cosine annealing schedule (Loshchilov and Hutter, 2016) to adjust the learning rate, starting from 1×10^{-3} and gradually decaying to 1×10^{-6}. This scheduling method reduced the learning rate in a cosine curve fashion, effectively preventing the network from settling into local optima by periodically refreshing the learning rate, thus promoting better convergence and stability.

2.6 Evaluation metrics

In the experiment, we used four evaluation metrics: accuracy, precision, recall, and F1 score. The proportion of accurately classified instances, both positive and negative, out of the total

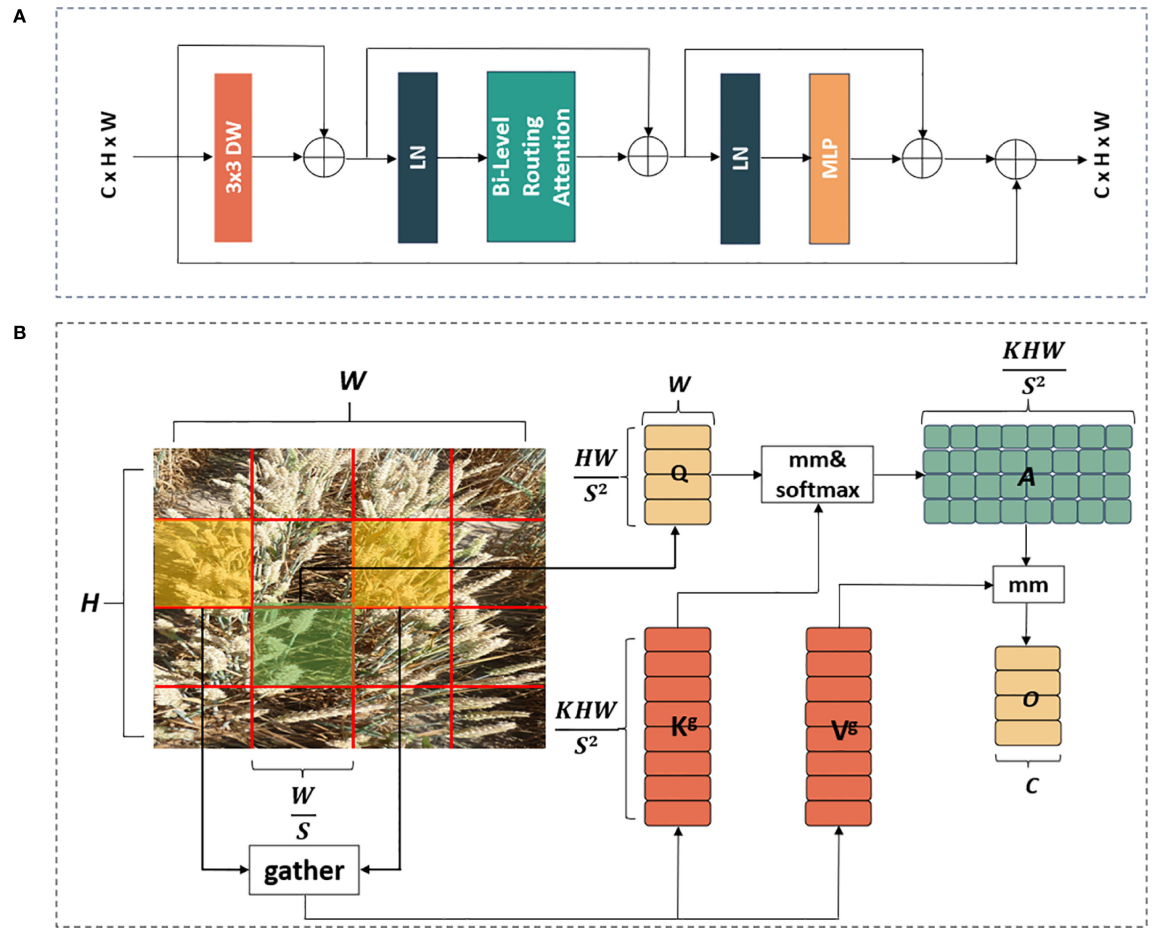


FIGURE 4
The workflow of BRA block: **(A)** overall structure. **(B)** bi-level routing attention module. In B, mm denoted the matrix multiplication operation.

number of instances was referred as *Accuracy*. It reflected the overall correctness of the classifier. *Precision* was defined as the proportion of accurately classified positive instances among all instances classified as positive. *Recall*, on the other hand, was the proportion of accurately identified positive instances among all true positive instances. The $F1-score$ combined *Precision* and *Recall* into a single metric by taking the harmonic mean of these two metrics. These definitions were given by (Equations 18–21). *Accuracy*, *Precision*, *Recall*, and $F1-score$ were denoted by *Acc*, *P*, *R* and *F*1, respectively, in the following evaluation results.

In these Equations, True Positives (TP) and True Negatives (TN) referred to cases where the true labels and false labels matched the predicted results, respectively. False Positives (FP) and False Negatives (FN) referred to cases where false labels were incorrectly predicted as true labels and true labels were incorrectly predicted as false labels, respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

$$Precision = \frac{TP}{TP + FP} \tag{20}$$

$$F1 - score = \frac{Precision \times Recall}{Precision + Recall} \tag{21}$$

3 Results

3.1 Ablation experiment

We first conducted the ablation experiments on FHBNet to validate the effectiveness of the individual modules.

The results in Table 2 demonstrated that each module had a positive impact on the prediction performance. The MSCCA block utilized large-kernel multi-scale convolutions and criss-cross attention to effectively adapt to changes in the orientation and scale of wheat ears in the field. It also thoroughly extracted the spatial contextual relationships, enhancing per-pixel representational capability, which increased accuracy by 4.06%. The attention mechanism in the BRA block focused on the

TABLE 2 Ablation experiment of FHBNet.

Blocks		Metric			
MSCCA	BRA	*Acc*/ %	*R*/ %	*F1*/ %	*P*/ %
×	×	74.07	74.07	74.03	74.21
✓	×	78.13	78.13	78.62	78.81
×	✓	75.42	75.42	75.63	76.17
✓	✓	79.49	79.49	79.54	79.94

critical information within the feature maps and effectively captured small, diseased wheat ears, raising accuracy by 1.35%. The improved model saw an overall accuracy increase of 5.42%, with all detection metrics showing substantial improvements. The FHBNet model achieved the optimal performance across all metrics, indicating that each component significantly contributed to the overall performance of the model.

3.2 The feature extraction process at each stage of FHBNet

To more clearly demonstrate the feature extraction process of FHBNet for diseased wheat spikes, we utilized GradCAM (Selvaraju et al., 2017) to visualize the feature maps at different stages, thereby providing a clear display of FHBNet learning curves.

As shown in Figure 5, we presented the test result using two example images. As the network deepened, FHBNet progressively extracted the features of wheat spikes, illustrating its effectiveness in feature extraction. This also highlighted the potential of FHBNet in detecting wheat spikes and identifying FHB. This formed the basis for the model to further discern pathological features on the spikes. According to the second test image, it was also evident that the model focused more on the diseased spikes.

3.3 Performance of FHBNet under different disturbances

To verify the generalization ability of the FHBNet model, we visualized network feature maps under different noise disturbances. In natural conditions, overexposure of RGB images due to excessive sunlight is common. Therefore, we selected two examples of overexposed images for visualization. As shown in Figure 6A, even in overexposed conditions, the FHBNet model can still accurately locate infected wheat spikes. Meanwhile, when selecting resistant varieties, the shooting angles are often in non-ideal states, such as spikes concentrated at the edges or in the center, or spikes oriented towards various angles. Therefore, we selected two images: one with wheat samples concentrated at the edge and another with spikes fallen towards the photographer, corresponding respectively to the upper and lower parts of Figure 6B. The visualization results demonstrated that, under different angles and orientations of spikes, FHBNet can still precisely locate FHB lesions and effectively perform FHB detection. We observed that in cases of moderate infection, healthy, and diseased spikes often intermingle, which posed challenges for assessing the severity of FHB. To address this, we selected two images with a moderate severity of FHB to test whether our model can accurately locate FHB lesions under such disturbances. As shown in Figure 6C, examples on the left were the original images, with the diseased parts of the spikes framed in red. On the right were the visualizations, where the framed parts of the spikes were the FHB lesions, distinctly redder compared to the healthy spikes. The result demonstrated that the FHBNet model focused on the pathological features of FHB on the spikes and can accurately locate FHB lesions amid interspersed healthy and diseased spikes, thereby enabling it to assess the severity of FHB based on these findings.

3.4 Comparison with other methods

To further analyze the performance of FHBNet, we conducted comparative experiments on the same testing environment and dataset with several current mainstream and advanced neural network models, including ConvNext (Liu et al., 2022), ViT (Dosovitskiy et al., 2020), RepLKNet (Ding et al., 2022a), MobileNet (Howard et al., 2017), MobileViT (Mehta and Rastegari, 2022), and EfficientNet (Tan and Le, 2019). The basic information of compared models can be found in the note of Supplementary Figure S1 in the supplementary file. To ensure a fair comparison, the hyperparameter configurations of these models were the same as FHBNet with Adam being the optimizer. The training Epoch was set to 250 and the batch size was 36. The cosine annealing schedule was adopted to promote better convergence and stability. Table 3 displayed the test result, and we also plotted the loss and accuracy curves of the seven models for a clear comparison of accuracy (see Supplementary Figure S1).

The results indicated that FHBNet achieved excellent performance, surpassing the second-best performing model by 3.7% in accuracy. We also observed that all models exhibited the lowest accuracy in recognizing moderate infection, as the complexity of the image background increased with healthy wheat and infected wheat intermingled, which interfered with the models' discriminative capabilities.

Additionally, we selected two examples from each of the three FHB severity levels to visualize the heatmaps for all models. As shown in Figure 7, ConvNext and ViT performed the worst, only focusing on areas with color traits similar to FHB. RepLKNet, MobileNet, MobileViT, and EfficientNet showed better results, but still had instances of missed and false detections, and, in some cases, failed to detect the wheat spikes at all. In contrast, FHBNet was able to precisely locate the majority of the wheat spikes and further pinpoint the lesions.

4 Discussion

In this study, we developed and validated the FHBNet model for automatically assessing the severity of wheat FHB with image-level

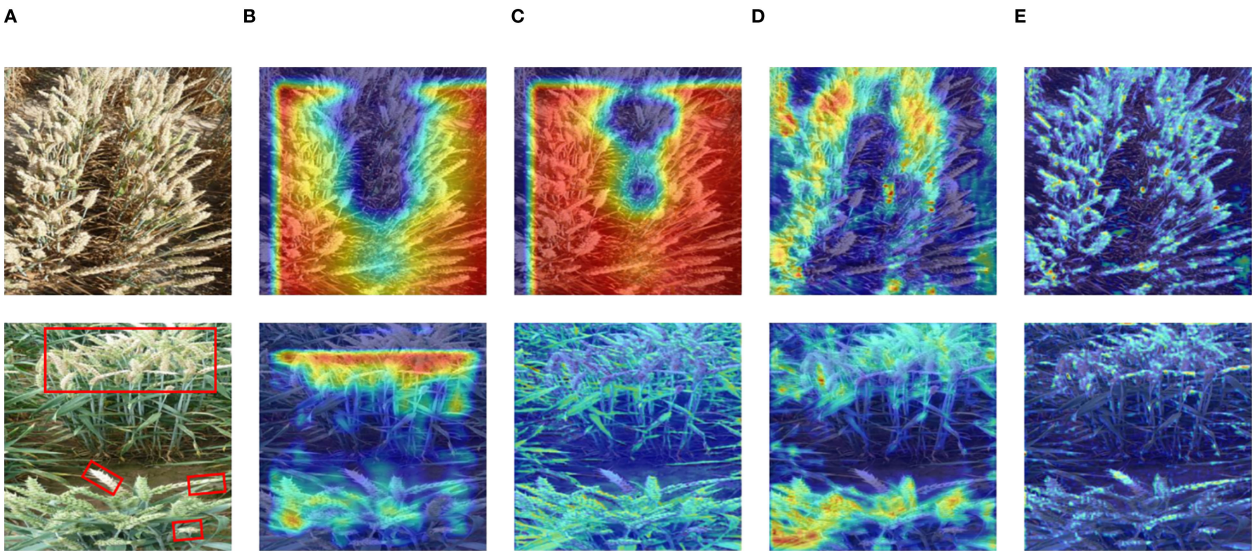


FIGURE 5
Heatmap visualization of FHBNet: **(A)** Original image. **(B)** Heatmap for Stage 1. **(C)** Heatmap for Stage 2. **(D)** Heatmap for Stage 3. **(E)** Heatmap for Stage 4. Note: the diseased wheat spikes are highlighted with red boxes in the second image.

annotated RGB images as an input. We introduced the MSCCA block, which integrated the FPSA module and the CCA module to enhance the capability of FHBNet for feature extraction and semantic analysis. The multi-scale convolution module adapted to variations in the orientation and scale distribution of field wheat spikes, thus improving the efficiency of spatial feature extraction. We enhanced the model's pixel-level representation by capturing the relationships between adjacent pixels. Subsequently, we introduced the BRA block to filter noise region from the feature map, achieving precise localization of small lesions on the spikes. In the experiments, FHBNet demonstrated superior performance, with accuracy and F1 score reaching 79.49% and 79.54%, respectively. FHBNet outperformed other advanced neural network models. Moreover, it maintained high recognition accuracy under various disturbances. These results indicated that FHBNet possessed strong generalization capability and robustness, enabling it to work stably under diverse field conditions.

We further analyzed models with the help of Table 3, Figure 7. The self-attention mechanism was a core advantage of ViT, enabling the model to capture relationships between different parts of the input image, regardless of their spatial positions in the image. This mechanism worked by computing the correlation between each image patch and all other patches, allowing ViT to understand the image globally rather than just locally. This was beneficial for the model to understand wheat images in complex backgrounds. However, by observing the result of ViT in Figure 7, we found that the heatmaps generated by ViT might have been too scattered, not clearly focusing on the details or pathological changes of wheat ears. The reason might

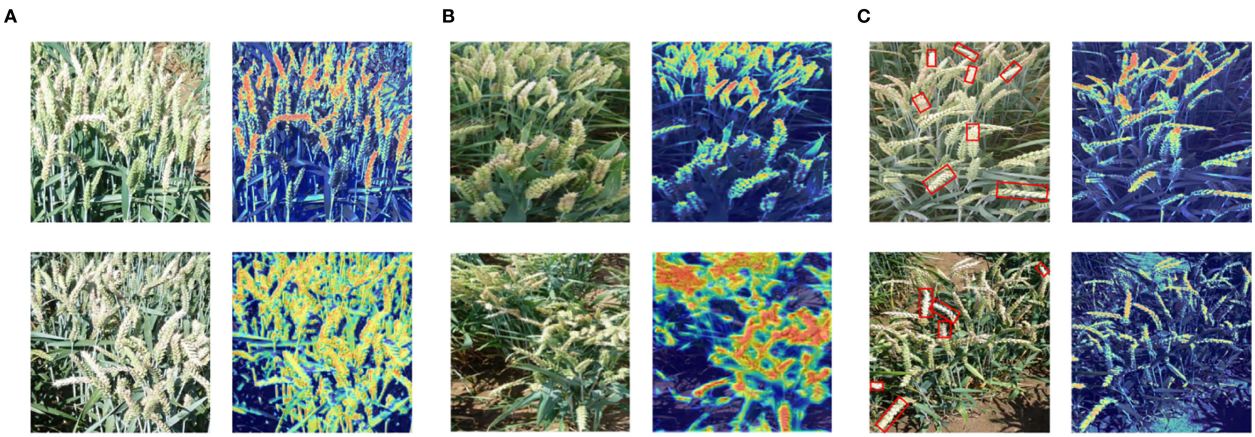


FIGURE 6
Heatmap visualization of FHBNet under different noise disturbances: **(A)** Strong natural light interference. **(B)** Different scales, with upper section indicated uneven distribution and the lower section denoted fallen wheat ears. **(C)** Intermingling of diseased wheat ears.

10.3389/fpls.2025.1549896

TABLE 3 Comparison results of SOTA models.

Model	Light		Moderate		Sever		F1/%	Acc/%
	P/%	R/%	P/%	R/%	P/%	R/%		
Ours	**86.75**	**74.86**	**71.16**	**76.12**	**82.02**	**85.78**	**79.54**	**79.49**
MobileViT	82.84	80.01	64.25	70.65	81.37	71.65	75.63	75.42
MobileNet	82.39	82.86	65.37	66.67	78.40	76.61	75.12	75.08
EfficientNet	85.99	77.14	61.29	71.54	79.81	77.98	75.84	75.59
RepLkNet	82.43	69.71	64.25	66.17	77.41	84.86	74.03	74.07
ViT	72.41	60.12	46.57	64.18	74.42	58.72	61.69	60.94
ConvNext	84.62	50.29	45.66	78.61	76.39	50.46	60.44	59.93

Bold fonts indicated the best performance.

have been that although the self-attention mechanism emphasized the global context, it might have overlooked the fine processing of extracting local features. When processing images of wheat ears, ViT might have overly focused on broad regions in the overall image rather than precisely focusing on subtle pathological changes on the ears. Additionally, compared to convolutional layers, self-attention lacked structured patterns of local receptive fields, which might have led to poor performance in extracting detailed features with local correlations.

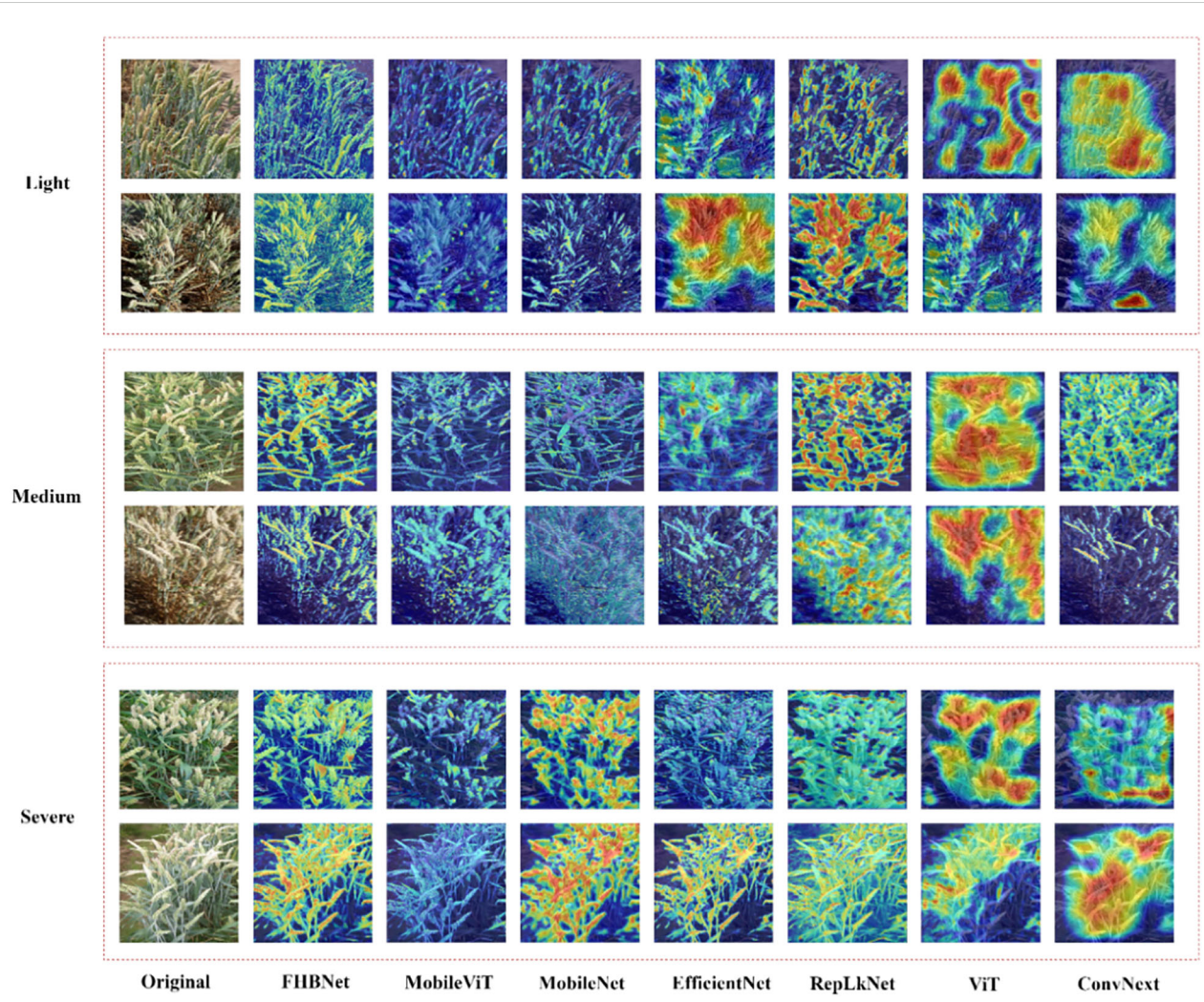


FIGURE 7
The heatmap visualization of all models.

The effect of MobileViT was slightly better than ViT. MobileViT combined the characteristics of ViT and convolutional networks, aiming to balance the global perceptual ability of Transformers with the efficient computation and local feature extraction ability of CNNs. This combination to some extent mitigated the drawback of ViT focusing only on global features. However, by observing the result of MobileViT in Figure 7, we found that some heatmaps displayed overly smooth activation regions, which might have reflected a 'blurring' phenomenon in the model's feature recognition process, meaning that the model failed to distinguish between different structures and boundaries in the image. Specifically, the global attention of the Transformer layers might have to some extent suppressed the local sensitivity of the convolutional layers, leading to visually smoother and blurrier feature representations. This was detrimental to the model's focus on wheat ears and their pathological features.

RepLKNet was a convolutional neural network that used large kernel convolutions to expand the model's receptive field and improve its ability to capture image features. This structure could better capture long-range dependencies in images, establish a good semantic space, and thus be able to locate wheat ears in complex backgrounds. Although the RepLKNet model successfully located the overall position of wheat to some extent, the range and contours of its activation areas were obviously larger than the actual wheat area in some samples. This might have been because RepLKNet used large kernel convolutions, which had a relatively wide receptive field. While this wide receptive field helped capture global semantic information in images, it might have also led to a decrease in accuracy in detail localization, especially in tasks that required precise delineation of small or fine features, which was detrimental in the severity discrimination task of FHB.

MobileNet achieved high recognition accuracy. However, as shown in the MobileNet column in Figure 7, it performed particularly poorly in the moderate class. The reason might have been that in the two examples of the moderate class, the spatial distribution of wheat was complex, and MobileNet was designed to be lightweight and efficient, it might not have been able to capture certain details and semantic information in complex images, especially for tasks that required understanding global context and were easily interfered by irrelevant background information. For example, MobileNet focused on the leaves, which was due to the model's failure to capture semantic information in the image, being interfered by irrelevant background information. Through the comparison in Figure 7, FHBNet could more completely and accurately extract wheat ears compared to other state-of-the-art models, which was important for further locating lesions on wheat ears and identifying the severity of FHB.

Additionally, our core advantage lied in the model's ability to directly learn features related to the severity of FHB from raw RGB images, without relying on annotated pixel-level annotations. This significantly reduced the cost of data labelling, while also enhancing the model's feasibility for practical applications (Kumar and Kukreja, 2022). Furthermore, the end-to-end nature of the model allowed for easy integration into existing agricultural monitoring systems, facilitating real-time disease prediction.

Despite the significant achievements of FHBNet in this study, there were still areas that warranted further exploration and improvement. First, under certain disturbances, there were still instances of missed detection of infected wheat spikes. As shown in Supplementary Figure S2, awns and weeds extensively cover the diseased spikes, and their color resembled that of the disease traits, making them difficult to distinguish. This complicated FHBNet to accurately assess the severity of FHB. In order to further enhance the prediction performance and robustness of FHBNet, we would further enhance the ability of extracting fine-grained features. For instance, to reduce the background interference on predictions, one could try the multi-granularity feature fusion approach (Xu et al., 2024; Su et al., 2024). Alternative approaches to enhance feature learning can also refer to knowledge distillation (Liu et al., 2024; Zhang et al., 2024). Second, FHBNet had a higher computational load compared to classic lightweight models and required more computing resources. Potential future directions for this research involve minimizing the model size through model compression and model pruning techniques to enable integration into edge devices (Yan et al., 2024; He et al., 2022). Such integration could aid in the onsite selection of FHB resistant varieties, thereby improving breeding efficiency. Last, optimizing the model for embedded devices could broaden its application in precision agriculture scenarios.

5 Conclusion

The model proposed in this study has achieved notable success in extracting FHB resistance phenotypes, providing a reliable basis for screening wheat resistant varieties. This research outcome is not only of theoretical importance but also provides practical decision support for wheat breeding, accelerating disease-resistant breeding efforts, and potentially providing new breakthrough directions for agricultural production.

Data availability statement

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

Author contributions

SXZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. HL: Conceptualization, Data curation, Formal analysis, Methodology,

Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. SZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – review & editing. HX: Funding acquisition, Investigation, Project administration, Supervision, Writing – review & editing. ZZ: Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This work was supported by the Natural Science Foundation of Jiangsu Province (Grant No. BK20231004), Jiangsu Province Key Research and Development Program (Grant No. BE2023369), and National Key Research and Development Program (2022YFD1900701).

Conflict of interest

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

Generative AI statement

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

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

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.2025.1549896/full#supplementary-material

References

Bhujel, A., Kim, N. E., Arulmozhi, E., Basak, J. K., and Kim, H. T. (2022). A lightweight attention-based convolutional neural networks for tomato leaf disease classification. *Agriculture-Basel* 12, 228. doi: 10.3390/agriculture12020228

Chen, P. (2020). Estimation of winter wheat grain protein content based on multisource data assimilation. *Remote Sens.* 12, 3201. doi: 10.3390/rs12193201

Ding, X., Chen, H., Zhang, X., Han, J., and Ding, G. (2022a). "Repmlpnet: Hierarchical vision mlp with re-parameterized locality," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* New Orleans, LA, United States.

Ding, X., Zhang, X., Han, J., and Ding, G. (2022b). "Scaling up your kernels to 31x31: Revisiting large kernel design in CNNs," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. New Orleans, LA, United States.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). "An image is worth 16x16 words: Transformers for image recognition at scale," in *International conference on learning representations*. Addis Ababa, Ethiopia.

Femenias, A., Gatius, F., Ramos, A. J., Sanchis, V., and Marín, S. (2020). Use of hyperspectral imaging as a tool for Fusarium and deoxynivalenol risk management in cereals: A review. *Food Control* 108, 106819. doi: 10.1016/j.foodcont.2019.106819

Gao, C., Gong, Z., Ji, X., Dang, M., He, Q., Sun, H., et al. (2022). Estimation of fusarium head blight severity based on transfer learning. *Agronomy* 12, 1876. doi: 10.3390/agronomy12081876

Gilbert, J., and Tekauz, A. (2000). Recent developments in research on Fusarium head blight of wheat in Canada. *Can. J. Plant Pathol.* 22, 1–8. doi: 10.1080/07060660009501155

Gruber, K. (2017). Agrobiodiversity: The living library. *Nature* 544, S8–S10. doi: 10.1038/544S8a

Gu, R., and Liu, L. (2025). An agricultural leaf disease segmentation model applying multi-scale coordinate attention mechanism. *Appl. Soft Computing* 172, 112904. doi: 10.1016/j.asoc.2025.112904

Hamila, O., Henry, C. J., Molina, O. I., Bidinosti, C. P., and Henriquez, M. A. (2024). Fusarium head blight detection, spikelet estimation, and severity assessment in wheat using 3d convolutional neural networks. *Can. J. Plant Sci* 4, 358–374. doi: 10.1139/cjps-2023-0127

He, J., Ding, Y., Zhang, M., and Li, D. (2022). Towards efficient network compression via Few-Shot Slimming. *Neural Networks* 147, 113–125. doi: 10.1016/j.neunet.2021.12.011

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. Las Vegas, NV, United States.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). "Mobilenets: Efficient convolutional neural networks for mobile vision applications," in *IEEE conference on computer vision and pattern recognition*. Honolulu, HI, United States.

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. Salt Lake City, UT, United States.

Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., and Liu, W. (2019). "Ccnet: Criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF international conference on computer vision*. Seoul, Korea.

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *Comput. Sci.* ArXiv. Available online at: https://arxiv.org/abs/1412.6980

Kong, Z., Ma, H., Yuan, G., Sun, M., Xie, Y., Dong, P., et al. (2022). Peeling the onion: hierarchical reduction of data redundancy for efficient vision transformer training. *Assoc. Advancement Artif. Intell* 939, 8360–8368. doi: 10.1609/aaai.v37i7.26008

Kumar, D., and Kukreja, V. (2022). Deep learning in wheat diseases classification: A systematic review. *Multimedia Tools Appl.* 81, 10143–10187. doi: 10.1007/s11042-022-12160-3

Li, J., Feng, Q., Zhang, J., and Yang, S. (2025). EMSAM: enhanced multi-scale segment anything model for leaf disease segmentation. *Front. Plant Sci.* 16, 1564079. doi: 10.3389/fpls.2025.1564079

Liu, Z., Lin, Y., Cao, Y., Hu, H., Zhang, Z., Lin, S., et al. (2021). "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*. Electric Network.

Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022). "A convnet for the 2020s," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. New Orleans, LA, United States.

Liu, B., Wei, S., Zhang, F., Guo, N., Fan, H., and Yao, W. (2024). Tomato leaf disease recognition based on multi-task distillation learning. *Front. Plant Sci.* 14, 1330527. doi: 10.3389/fpls.2023.1330527

Loshchilov, I., and Hutter, F. (2016). "Sgdr: Stochastic gradient descent with warm restarts," in *International Conference on Learning Representations*. Caribe Hilton, San Juan, Puerto Rico.

Mehta, S., and Rastegari, M. (2022). "Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer," in *International conference on learning representations*. Electric Network.

Mi, Z., Zhang, X., Su, J., Han, D., and Su, B. (2020). Wheat stripe rust grading by deep learning with attention mechanism and images from mobile devices. *Front. Plant Sci.* 11, 558126. doi: 10.3389/fpls.2020.558126

Mochida, K., Koda, S., Inoue, K., Hirayama, T., Tanaka, S., Nishii, R., et al. (2019). Computer vision-based phenotyping for improvement of plant productivity: a machine learning perspective. *GigaScience* 8, giy153. doi: 10.1093/gigascience/giy153

Rößle, D., Prey, L., Ramgraber, L., Hanemann, A., Cremers, D., Noack, P. O., et al. (2023). Efficient noninvasive FHB estimation using RGB images from a novel multiyear, multirater dataset. *Plant Phenomics* 5, 0068. doi: 10.34133/plantphenomics.0068

Sari, E., Knox, R. E., Ruan, Y., Henriquez, M. A., Kumar, S., Burt, A. J., et al. (2020). Historic recombination in a durum wheat breeding panel enables high-resolution mapping of Fusarium head blight resistance quantitative trait loci. *Sci. Rep.* 10, 7567. doi: 10.1038/s41598-020-64399-1

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*. Venice, Italy.

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., and Singh, A. (2018). Deep learning for plant stress phenotyping: trends and future perspectives. *Trends Plant Sci.* 23, 883–898. doi: 10.1016/j.tplants.2018.07.004

Srinivasan, S., Somasundharam, L., Rajendran, S., Singh, V. P., Mathivanan, S. K., and Moorthy, U. (2025). DBA-ViNet: an effective deep learning framework for fruit disease detection and classification using explainable AI. *BMC Plant Biol.* 25, 965. doi: 10.1186/s12870-025-07015-6

Su, Y., Zhao, H., Zheng, Y., and Wang, Y. (2024). Few-shot learning with multi-granularity knowledge fusion and decision-making. *IEEE Trans. Big Data* 10, 486–497. doi: 10.1109/TBDATA.2024.3350542

Tan, M., and Le, Q. (2019). "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. Long Beach, CA, United States.

Tian, L., Yuan, Y., En, Q., Ma, W., Zhang, G., and Liang, F. (2025). Attention-based unsupervised prompt learning for SAM in leaf disease segmentation. *Knowledge-Based Syst.* 320, 113652. doi: 10.1016/j.knosys.2025.113652

Toda, Y., and Okura, F. (2019). How convolutional neural networks diagnose plant disease. *Plant Phenomics* 2019, 9237136. doi: 10.34133/2019/9237136

Tolstikhin, I. O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., et al. (2021). Mlp-mixer: An all-mlp architecture for vision. *Adv. Neural Inf. Process. Syst.* 1857, 24261–24272.

Wang, Y. H., Li, J. J., and Su, W. H. (2023). An integrated multi-model fusion system for automatically diagnosing the severity of wheat fusarium head blight. *Agriculture* 13, 1381. doi: 10.3390/agriculture13071381

Wei, L., Tang, J., Chen, J., Mukamakuza, C. P., Zhang, D., and Zhang, T. (2025). A lightweight few-shot learning model for crop pest and disease identification. *Artif. Intell. Rev.* 58, 329. doi: 10.1007/s10462-025-11323-6

Weng, S., Han, K., Chu, Z., Zhu, Z., Zhang, Z., Zheng, L., et al. (2021). Reflectance images of effective wavelengths from hyperspectral imaging for identification of Fusarium head blight-infected wheat kernels combined with a residual attention convolution neural network. *Comput. Electron. Agric.* 190, 106483. doi: 10.1016/j.compag.2021.106483

Xu, Y., Wu, S., Yang, M., Wu, Z., Yao, Y., and Wei, Z. (2024). Two-stage fine-grained image classification model based on multi-granularity feature fusion. *Pattern Recognition* 146, 110042. doi: 10.1016/j.patcog.2023.110042

Yan, W., Feng, Q., Yang, S., Zhang, J., and Yang, W. (2024). Prune-FSL: Pruning-based lightweight few-shot learning for plant disease identification. *Agronomy* 14, 1878. doi: 10.3390/agronomy14091878

Yan, C., and Li, H. (2025). CAPNet: tomato leaf disease detection network based on adaptive feature fusion and convolutional enhancement. *Multimedia Syst.* 31, 178. doi: 10.1007/s00530-025-01756-y

Yang, Z., Sun, L., Liu, Z., Deng, J., Zhang, L., Huang, H. X., et al. (2025). FATDNet: A fusion adversarial network for tomato leaf disease segmentation under complex backgrounds. *Comput. Electron. Agric.* 234, 110270. doi: 10.1016/j.compag.2025.110270

Zang, H., Wang, C., Zhao, Q., Zhang, J., Wang, J., Zheng, G., et al. (2025). Constructing segmentation method for wheat powdery mildew using deep learning. *Front. Plant Sci.* 16, 1524283. doi: 10.3389/fpls.2025.1524283

Zhang, H., Huang, L., Huang, W., Dong, Y., Weng, S., Zhao, J., et al. (2022a). Detection of wheat Fusarium head blight using UAV-based spectral and image feature fusion. *Front. Plant Sci.* 13, 1004427. doi: 10.3389/fpls.2022.1004427

Zhang, X., Kun, L., and Zhang, Y. (2024). Plant pest and disease lightweight identification model by fusing tensor features and knowledge distillation. *Front. Plant Sci.* 15, 1443815. doi: 10.3389/fpls.2024.1443815

Zhang, D., Wang, D., Gu, C., Jin, N., Zhao, H., Chen, G., et al. (2019). Using neural network to identify the severity of wheat Fusarium head blight in the field environment. *Remote Sens.* 11, 2375. doi: 10.3390/rs11202375

Zhang, H., Zu, K., Lu, J., Zou, Y., and Meng, D. (2022b). "EPSANet: An efficient pyramid squeeze attention block on convolutional neural network," in *Proceedings of the asian conference on computer vision*. Macau SAR, China.

Zhu, L., Wang, X., Ke, Z., Zhang, W., and Lau, R. H. W. (2023). "Biformer: Vision transformer with bi-level routing attention," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Vancouver, Canada.

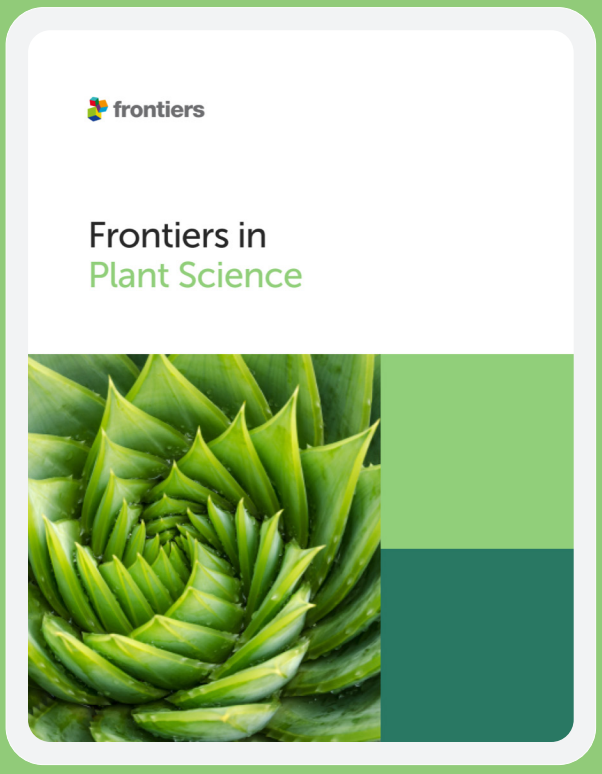
Frontiers in
Plant Science

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