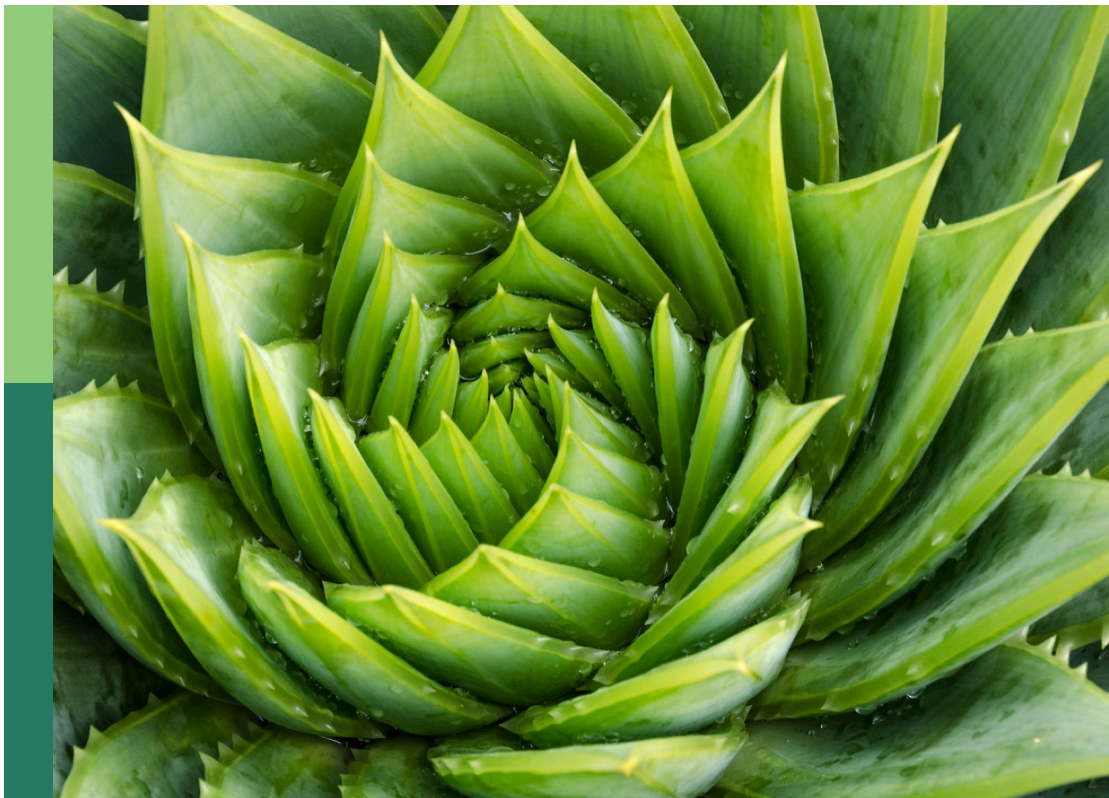
frontiers | Research Topics

Spectroscopy, imaging and machine learning for crop stress

Edited by
Shizhuang Weng, Baohua Zhang and Yijun Yan

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-3220-1
DOI 10.3389/978-2-8325-3220-1

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

Spectroscopy, imaging and machine learning for crop stress

Topic editors

Shizhuang Weng — Anhui University, China

Baohua Zhang — Nanjing Agricultural University, China

Yijun Yan — Robert Gordon University, United Kingdom

Citation

Weng, S., Zhang, B., Yan, Y., eds. (2023). *Spectroscopy, imaging and machine learning for crop stress*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-3220-1

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 28 July 2023
DOI 10.3389/fpls.2023.1240738

OPEN ACCESS

EDITED AND REVIEWED BY
Lisbeth Garbrecht Thygesen,
University of Copenhagen, Denmark

*CORRESPONDENCE
Shizhuang Weng
weng_1989@126.com

RECEIVED 15 June 2023
ACCEPTED 17 July 2023
PUBLISHED 28 July 2023

CITATION
Weng S (2023) Editorial: Spectroscopy, imaging and machine learning for crop stress.
Front. Plant Sci. 14:1240738.
doi: 10.3389/fpls.2023.1240738

Editorial: Spectroscopy, imaging and machine learning for crop stress

Shizhuang Weng*

National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, School of Electronic and Information Engnieering Anhui University, Hefei, China

KEYWORDS

spectroscopy, machine vision, hyperspectral imaging, machine learning, deep learning, crop stress

Editorial on the Research Topic
Spectroscopy, imaging and machine learning for crop stress

Crop stress poses a huge challenge to food security necessitating innovative approaches for early detection monitoring management of stress. In recent years the integration of spectroscopy imaging machine learning techniques has emerged as a promising avenue for detecting various types of crop stress. This editorial work introduces recent publications within the field included in the Research Topic "*Spectroscopy imaging machine learning for crop stress*." By exploring these cutting-edge research findings we can gain valuable insights into the application of these technologies to enhance agricultural resilience productivity.

Xiao et al. focused on the use of visible and near-infrared spectroscopy combined with deep learning to estimate leaf nitrogen concentration in cotton leaves. The study employed random frog, weighted partial least squares regression and saliency map to select characteristic wavelengths. The models based on convolutional neural network showed excellent performance for both qualitative and quantitative prediction of leaf nitrogen. These findings highlight the potential of deep learning and visible and near-infrared spectroscopy within accurate and real-time assessment of cotton leaf nitrogen content, enabling farmers to make applicable fertilization decisions

Wu et al. employed an unmanned aerial vehicle (UAV) to obtain multispectral images of a rice canopy and analyzed the response of multispectral reflectance features and physiological parameters including net photosynthetic rate (Pn), plant height (PH), and SPAD to different nitrogen treatments or leakage conditions at various growth stages of the crop. The study extensively analyzed the correlation between vegetation indices (VIs), texture indices (TIs), and Pn based on UAV multispectral images. The techniques and findings presented in this paper provide valuable insights within field-scale photosynthetic monitoring in rice cultivation and improve stress detection and yield prediction.

Choudhury et al. introduces two innovative algorithms, namely VisStressPredic and HyperStressPropagateNet, to predict the onset and propagation of drought stress in plants using camera-captured image sequences in visible light and hyperspectral modalities. The algorithms analyze visible light sequences at discrete intervals and utilize a deep neural network and hyperspectral images to propagate stress over time, which demonstrate a strong correlation between soil water content and the percentage of stressed plants. These

algorithms are evaluated on a dataset of cotton plant image sequences and offer valuable support for studying abiotic stresses in diverse plant species, promoting sustainable agricultural practices.

Hu X. et al. propose a novel approach called Class-Attention-based Lesion Proposal Convolutional Neural Network (CALP-CNN), utilizing a class response map to locate the main lesion object and identify discriminative lesion details in visual light images. Through a cascade architecture, CALP-CNN effectively handles complex background interference and misclassification of similar diseases. Experimental results on a self-built dataset demonstrate CALP-CNN achieves good classification performance and outperforms the existing fine-grained image recognition methods, highlighting its efficacy in field identification of strawberry diseases.

Hu Y. et al. presents an enhanced encoder-decoder framework based on DeepLab v3+ analysis of images to accurately identify crops with diverse planting patterns. The network utilizes ShuffleNet v2 as the backbone for feature extraction and incorporates a convolutional block attention mechanism to fuse attention features across channels and spatial dimensions. The enhanced network achieves significant improvements and requires fewer parameters and computational operations compared to other networks. This study demonstrates the effectiveness of Deep-agriNet in identifying crops with different planting scales, making it a valuable tool for crop identification in various regions and countries.

The combination of spectroscopy, imaging, and machine learning has a high potential for improving crop stress analysis and management. By utilizing these technologies, we can enhance our understanding of crop stress dynamics, develop precise and targeted stress detection methods, and improve decision-making processes for farmers. Ongoing research, technological advancements, and collaborative efforts are necessary to unlock the full potential of spectroscopy, imaging, and machine learning in mitigating crop stress and ensuring global food security.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

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.

Publisher's note

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 21 September 2022
DOI 10.3389/fpls.2022.1004427

OPEN ACCESS

EDITED BY
Daniel Cozzolino,
University of Queensland,
Australia

REVIEWED BY
Haikuan Feng,
Beijing Research Center for Information Technology in Agriculture, China
Zipeng Zhang,
Xinjiang University,
China

*CORRESPONDENCE
Yingying Dong
dongyy@aircas.ac.cn

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

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

RECEIVED 27 July 2022
ACCEPTED 29 August 2022
PUBLISHED 21 September 2022

CITATION
Zhang H, Huang L, Huang W, Dong Y, Weng S, Zhao J, Ma H and Liu L (2022) 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

Detection of wheat *Fusarium* head blight using UAV-based spectral and image feature fusion

Hansu Zhang[1†], Linsheng Huang[1†], Wenjiang Huang[2,3,4], Yingying Dong[2,3]*, Shizhuang Weng[1], Jinling Zhao[1], Huiqin Ma[2] and Linyi Liu[2]

[1]National Engineering Research Center for Agro-Ecological Big Data Analysis and Application, Anhui University, Hefei, China, [2]State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, [3]University of Chinese Academy of Sciences, Beijing, China, [4]Key Laboratory for Earth Observation of Hainan Province, Sanya, China

Infection caused by *Fusarium* head blight (FHB) has severely damaged the quality and yield of wheat in China and threatened the health of humans and livestock. Inaccurate disease detection increases the use cost of pesticide and pollutes farmland, highlighting the need for FHB detection in wheat fields. The combination of spectral and spatial information provided by image analysis facilitates the detection of infection-related damage in crops. In this study, an effective detection method for wheat FHB based on unmanned aerial vehicle (UAV) hyperspectral images was explored by fusing spectral features and image features. Spectral features mainly refer to band features, and image features mainly include texture and color features. Our aim was to explain all aspects of wheat infection through multi-class feature fusion and to find the best FHB detection method for field wheat combining current advanced algorithms. We first evaluated the quality of the two acquired UAV images and eliminated the excessively noisy bands in the images. Then, the spectral features, texture features, and color features in the images were extracted. The random forest (RF) algorithm was used to optimize features, and the importance value of the features determined whether the features were retained. Feature combinations included spectral features, spectral and texture features fusion, and the fusion of spectral, texture, and color features to combine support vector machine, RF, and back propagation neural network in constructing wheat FHB detection models. The results showed that the model based on the fusion of spectral, texture, and color features using the RF algorithm achieved the best performance, with a prediction accuracy of 85%. The method proposed in this study may provide an effective way of FHB detection in field wheat.

KEYWORDS

hyperspectral images, UAV, crop stress, feature fusion, classification models

Introduction

Wheat is the second largest grain crop in China. The stable and high yield of wheat has been the focus of agricultural production (Huang et al., 2020). *Fusarium* head blight (FHB), also known as scab, is a devastating wheat disease caused by the fungal plant pathogen *Fusarium graminearum* (*Gibberella*). *Fusarium*-infected wheat typically results in small, low mass, and shrunken grains, which can rapidly lead to very large crop losses and quality degradation (Bauriegel et al., 2011b). Furthermore, the fungus produces a large number of mycotoxins (deoxynivalenol, nivalenol and zearalenones etc.), among which the most toxic deoxynivalenol (DON) can disrupt normal cell function by inhibiting protein synthesis, posing a significant threat to human and animal health (Barbedo et al., 2015). In recent years, with global climate change, wheat FHB infection has become increasingly serious, resulting in severe damage to wheat quality and yields. Ineffective FHB management practices hinder the profitable and sustainable production of wheat, affecting its economic and social benefits in China. Therefore, the detection of disease development of wheat is important and essential for successful disease control.

Traditional FHB detection mainly relies on professionals to scout the development of wheat infection through visual interpretation, or scholars use chemical methods, such as gas chromatography (GC) (Simsek et al., 2012), high performance liquid chromatography (HPLC) (Simsek et al., 2012), enzyme-linked immunosorbent assay (ELISA) (Maragos et al., 2006), and polymerase chain reaction (PCR) (Amar et al., 2012; Atoui et al., 2012) to detect FHB and DON production. However, these methods are time-consuming, labor-intensive, unable to achieve large-scale monitoring, and are destructive to wheat. Remote sensing technology has been widely used in the monitoring and identification of wheat FHB with nondestructive inspections and rapid measurements. At present, monitoring of wheat FHB using remote sensing technology is mainly manifested in three aspects: (i) identify wheat kernels with varying degrees of damage under laboratory conditions to accurately judge the quality of wheat kernels (Delwiche et al., 2011; Barbedo et al., 2015; Jaillais et al., 2015; Alisaac et al., 2019; Femenias et al., 2020; Liang et al., 2020; Zhang D. Y. et al., 2020a; Zhang D. Y. et al., 2020b); (ii) use remote sensing technology to capture the information of individual or canopy wheat infected with FHB to accurately detect the disease (Dammer et al., 2011; Menesatti et al., 2013; Whetton et al., 2018a,b; Huang et al., 2019b; Zhang et al., 2019; Huang et al., 2020; Ma et al., 2020; Huang et al., 2021); and (iii) monitor wheat FHB on a regional scale with remote sensing (Liu et al., 2020b). However, there are many limitations in these studies. The inspection of wheat kernels has a time lag that only allows the use of kernels with different qualities and cannot fundamentally ameliorate wheat infection. Quantitative detection studies at the single plant scale or canopy scale only provides a theoretical reference without the spatial distribution of wheat infection to meet the needs of practical applications. Optical satellite images are at risk of being covered by clouds, and FHB may occur severely and frequently in cloudy and foggy areas, reducing the availability of remote sensing images (Liu et al., 2020a). Therefore, there is an urgent need for new technological means to solve the current problems.

Unmanned aerial vehicles (UAVs) are considered a practical detection method for crop pests and diseases. Unlike near-ground and satellite-based remote sensing platforms, applications of UAV have the advantages of large coverage, high efficiency, and flexibility (Fu et al., 2022; Zhu et al., 2022a). UAV can collect very high-resolution images and data in a cost-effective manner over a short period of time (Ye et al., 2020). As a new technological means, UAV technology has made significant progress in crop classification, growth monitoring, and identification of pests and diseases. UAV also allows for a proper balance between image quality, sensing efficiency, and operating cost (Li et al., 2019). At present, UAV images are mainly divided into multispectral images and hyperspectral images. Hyperspectral images have dozens to hundreds of continuous and subdivided spectral bands in the ultraviolet, visible, near-infrared, and mid-infrared regions, making them more sensitive to the reflected energy of light and increasingly available (Liu et al., 2020a). Hyperspectral images can provide image and spectral data of each pixel, thus detecting the internal chemical compositions and external phenotypic traits of objects (Zhang D. Y. et al., 2020b). Currently, there are few reports on the detection of FHB infection in wheat using UAV hyperspectral technology (Liu et al., 2020a; Ma et al., 2021; Xiao et al., 2021). We attempted to use UAV hyperspectral technology to explore wheat FHB detection methods in our study. What's more, scholars have primarily mined spectral features that could characterize physiological and biochemical changes (such as moisture, pigment, etc.), as well as considered texture features that can represent spatial changes of wheat to detect FHB. In fact, the infected wheat tissue usually transitions from green (healthy tissue) to yellow–white (diseased tissue) as the disease progresses. Color has been proven to be the most effective means to distinguish different image objects and realize object recognition among phenotypic traits (e.g., color, texture, and size) extracted from images (Zhang et al., 2018). However, the application of UAV color features in FHB detection has not been explored. Therefore, this study combines color features to further explore the wheat FHB detection methods based on UAV hyperspectral images.

Our study investigated the potential of fusing spectral and image features of UAV hyperspectral images to improve the ability of detecting wheat FHB in the field. The overall technical flow chart is shown in Figure 1. First, we determined the most suitable sensitive spectral features to identify FHB; these features reflect the disease stress of the host. Second, we extracted texture features that could represent the disease distribution based on band images containing the most disease information. Finally, we calculated the color features that characterize disease incidence. We combined multiple algorithms to construct classification models and examine the effect of multi-features on the detection accuracy of FHB. Our goals were to (1) evaluate the performance

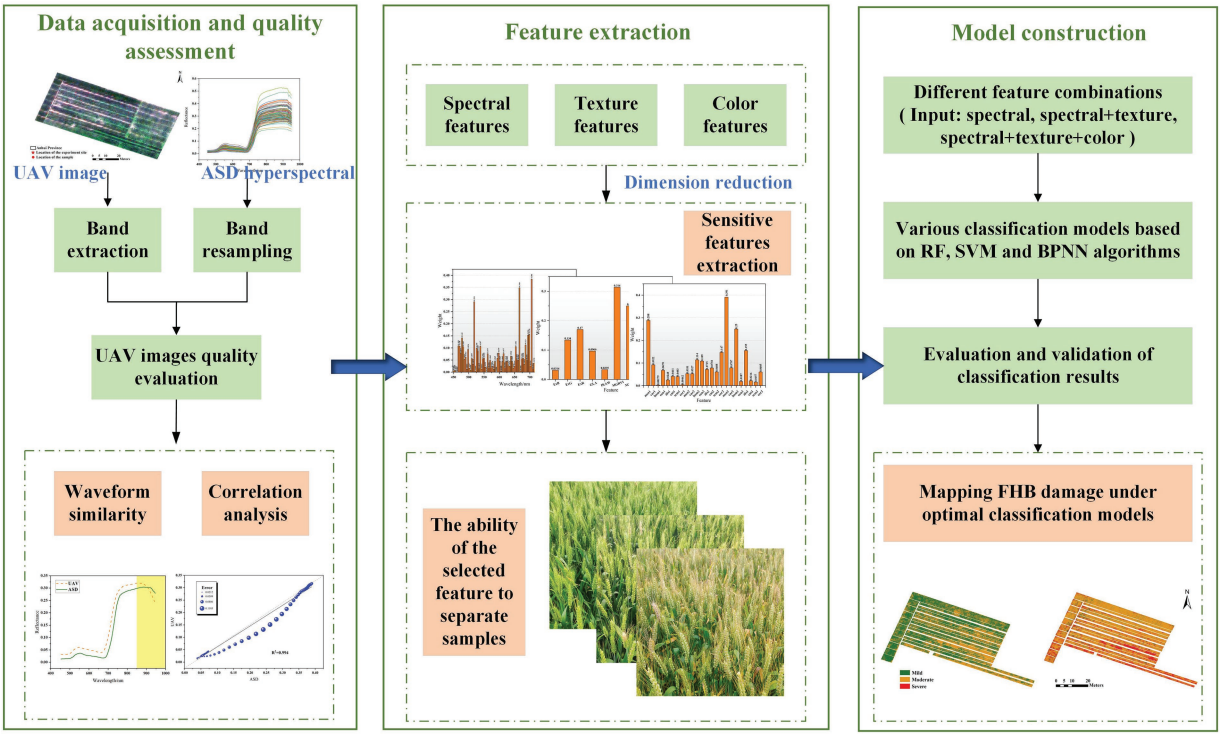


FIGURE 1
Methodological framework.

of UAV hyperspectral images in identifying wheat FHB occurrence; (2) evaluate the potential of multi-features in FHB detection; (3) explore the best classification method for UAV images; and (4) map the occurrence of FHB in a wheat field using the optimal model. In general, we developed a novel method for FHB detection based on UAV images, which forms a basis for the precise prevention and control of FHB.

Materials and methods

Experiment site and data acquisition

Our experiment site was situated in the Anhui Agricultural University Production Base (31°290 N, 117°130E) in Lujiang County, Anhui Province, China (Figure 2). The main wheat variety in this area is Yangmai 25, which is susceptible to FHB. Zero tillage and a typical subtropical humid monsoon climate provide favorable conditions for the occurrence of wheat FHB in this region. According to the Anhui Meteorological Service, the average temperature from April to early May 2019 in Lujiang County was about 20°C, accompanied by several days of rainfall. The wheat was in the flowering period in April. Sufficient fungus sources and climatic conditions caused natural wheat FHB in the experiment site.

Data were sourced from UAV image acquisition and field investigation. The UAV images were obtained using an M600 Pro aircraft of Dajang Innovations (DJI) during the wheat filling stage on May 3 and 8, 2019. This system was equipped with a Cubert S185 FireflEYE SE hyperspectral imaging camera (Cubert GmbH, Ulm, Baden-Württemberg, Germany), which can collect the reflected radiation in the 450–950 nm range. The spectral sampling interval was 4 nm, and there were 125 bands in total. The UAV flew at a speed of 3 m/s at an altitude of 60 m. The camera triggers at a frequency of 0.8 s, with a forward overlap of 80% and a side overlap of 65%. All UAV images were collected under clear weather and cloudless skies between 11 a.m. and 1 p.m. (local time). Before capturing hyperspectral images, radiometric correction of the camera was required. A panchromatic image with high spatial resolution and hyperspectral cube image with a low spatial resolution were fused and spliced for subsequent analysis. The final hyperspectral images had a spatial resolution of 4 cm. Field investigation experiments were carried out while capturing the UAV images. Fifty plots (each with an area of 1 m^2) were evenly selected across the experiment field. These plots were used as ground sample points to verify the quality of the UAV images. To accurately locate the sampling points, we fixed a flagpole next to each point. The canopy spectral reflectance of the sample points was collected using an ASD FieldSpec Pro spectrometer (Analytical Spectral Devices, Inc., Boulder, CO, USA), which has a spectral resolution of 3 nm in the range of 350–1,000 nm and 10 nm in the range of 1,000–2,500 nm. All canopy spectral measurements were carried out at a height of about 1.3 m above the ground, and 10 measurements were taken at each sample point. A BaSO4 calibration panel was used before

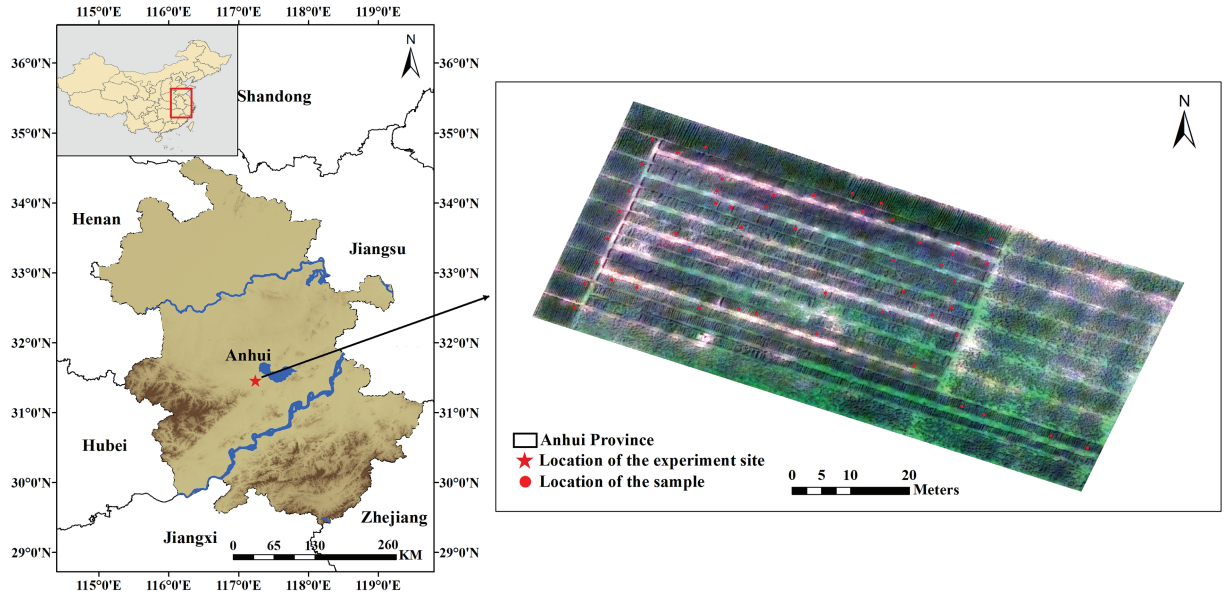


FIGURE 2
Location of the experiment site and field investigation samples. The star in the left map represents the location of the experiment site, and the right map is the experiment field photographed by the UAV, where red marks the location of the field investigation point.

each measurement to correct for changes in illumination conditions, and the average was used as the final canopy spectrum. According to the rules for monitoring and forecasting wheat head blight suggested by the National Plant Protection Department of China (Chinese Standard: GB/T 15796–2011), the diseased ear ratio (DER) in each plot can be expressed by the ratio of diseased ears to the total investigated ears. The wheat planting density in the study area was relatively uniform. Then, we randomly selected 50 wheat plants at every sample point and recorded the number of diseased wheat plants by visual interpretation. DER was divided into five classes: $0.1\% < DER \leq 10\%$ (Class 1), $10\% < DER \leq 20\%$ (Class 2), $20\% < DER \leq 30\%$ (Class 3), $30\% < DER \leq 40\%$ (Class 4), and $DER > 40\%$ (Class 5). Actually, wheat fields with more than 30% infected wheat are severely damaged, and those with less than 10% are mildly damaged. Therefore, we reclassified DER into three grades: mild infection ($0.1\% < DER \leq 10\%$), moderate infection ($10\% < DER \leq 30\%$), and severe infection ($DER > 30\%$) for subsequent analysis.

Data processing and analysis

Data quality assessment of UAV

UAV hyperspectral images are obtained by fusing and splicing a panchromatic image and hyperspectral cube image. UAV is prone to the impact of objective factors such as shaking in flight. Therefore, it is necessary to evaluate the image quality before identifying wheat FHB in the field. ASD spectrometers are widely used in agricultural remote sensing monitoring, and their spectral information is often used as an important basis for monitoring crop pests and diseases (Cao et al., 2013; Ashourloo et al., 2014; Zheng et al., 2018; Huang et al., 2019a; Ma et al., 2020). In this study, we used ASD spectral data as a criterion to evaluate the quality of UAV images (Bareth et al., 2015; Gao et al., 2016; Chen et al., 2018). First, we extracted and averaged the spectral reflectance of all pixels in the sample points to obtain the UAV spectral information in the same region as the ASD measurement. Second, we analyzed the spectral variations between the two data sets by resampling the ASD canopy spectrum and determining the differences in the waveforms. Finally, we calculated the correlation between the resampled ASD spectrum and the UAV spectrum in the 450–950 nm range. If there is a strong correlation between the data measured by the two sensors and the same spectral curve, then the UAV data are considered reliable.

Optimal feature selection for wheat FHB detection

The UAV hyperspectral images captured in this study contain 125 spectral bands, from visible to near infrared, which reflect the internal physiological and biochemical changes of wheat after pathogen infection (Li et al., 2014). In addition, wheat presents different spatial distributions as FHB severity increases, as indicated by the texture and color features of an image. Here, we detected FHB of the wheat field by extracting key features from images captured on May 3 and 8. The wheat was at the same growth stage on both dates; therefore, any feature changes between the two dates were mainly due to disease development rather than the wheat growth. It should be noted that the extracted features are not only spectral features but also include image features (texture and color features). Moreover, the extracted features may

contain invalid information and thus be insensitive to wheat FHB. A random forest (RF) algorithm was adopted to further reduce data redundancy and develop efficient models.

The spectral features for each plot were extracted from hyperspectral images using the region of interest (ROI) tool in ENVI 5.3 software. The feature extraction method is the same as that used in data quality assessment of UAV. We extracted and averaged the spectral reflectance of all pixels contained in the sample point as the final spectral value of each sample point. The texture features were extracted by the gray level co-occurrence matrix (GLCM) method (Zhang et al., 2017). The GLCM method is a classical statistical analysis technique that describes texture by studying the spatial correlation characteristics of the gray level (Guo et al., 2020). The mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation were extracted for FHB detection analysis. Table 1 describes the texture features. Before texture feature extraction, the principal component analysis (PCA) method was used to reduce the dimensionality of the hyperspectral images and generate principal component images containing only three bands. The first three bands contain most of the information (the cumulative variance exceeds 97%); thus, the texture features were extracted from the gray images corresponding to the three bands. The extraction of texture features was completed with ENVI 5.3 software, and the specific process occurred in four steps (Fu et al., 2022): (1) select the gray images in "Texture Input File" dialog, (2) select the necessary texture features in the check box, (3) set the processing window size to 3×3 (the smallest window size guarantees the highest resolution), and (4) set the output path and calculate the texture values. A total of 24 texture features were calculated.

TABLE 1 The texture feature used in the study and descriptions.

Texture feature	Abbreviation	Content
Mean	mea	Average of grey levels
Variance	var	Change in greyscale
Homogeneity	hom	Local homogeneity, as opposed to contrast
Contrast	con	Clarity of texture
Dissimilarity	dis	Similarity of the pixels
Entropy	ent	Diversity of the pixels
Second Moment	sem	Uniformity in greyscale
Correlation	cor	Ductility of grey value

For the color features selection, we calculated color indices through band combinations to indicate different aspects of wheat infection (Li et al., 2019; Huang et al., 2020; Ge et al., 2021). Color feature is the most widely used visual feature in image retrieval; it is usually related to the object or scene contained in the image; at the same time, color feature is less dependent on the size, orientation, and perspective of the image itself, making it highly robust (Huang et al., 2020). During the mild infection stage, several wheat plants were withered and yellowed in the field. As the infection worsened, the damaged area gradually increased (Figure 3). Chromatic aberration can be used to distinguish the severity of FHB. In this study, three wavelengths (694, 542, and 482 nm) of the hyperspectral images were used to synthesize RGB images and extract color features. The extracted color features mainly included Excess Blue Vegetation Index (ExB), Excess Green Vegetation Index (ExG), Excess Red Vegetation Index (ExR), Green Leaf Algorithm (GLA), Kawashima Index (IKAW), Modified Green Red Vegetation Index (MGRVI), Normalized Green-Red Difference Index (NGRDI), Red Green Blue Vegetation Index (RGBVI), Visible Atmospherically Resistant Index (VARI), and Woebbecke Index (WI). Details of the 10 color features mentioned in this paper are shown in Table 2.

Rational selection of the important features in wheat FHB detection is the most critical step in image analysis. RF consists of multiple decision trees, which can calculate the importance of individual feature variables. The feature evaluation method is called "embedding," which integrates the features of the filter and wrapper methods (Pal and Foody, 2010). We evaluated the importance of features by calculating the contribution rate of each feature in the random forest, as measured by the Gini index (Deng

FIGURE 3
Different incidences of wheat in the field: mild infection (left), moderate infection (center), and severe infection (right).

TABLE 2 The color feature used in the study and descriptions.

Color feature (abbreviation)	Full name	Formula	Reference
ExB	Excess Blue Vegetation Index	1.4B-G	Li et al. (2019)
ExG	Excess Green Vegetation Index	2G-R-B	Woebbecke et al. (1995)
ExR	Excess Red Vegetation Index	1.4R-G	Meyer and Neto (2008)
GLA	Green Leaf Algorithm	(2G-R-B)/(2G + R + B)	Louhaichi et al. (2001)
IKAW	Kawashima Index	(R-B)/(R + B)	Kawashima and Nakatani (1998)
MGRVI	Modified Green Red Vegetation Index	$(G^2\text{-}R^2)/(G^2 + R^2)$	Tucker (1979)
NGRDI	Normalized Green-Red Difference Index	(G-R)/(G + R)	Tucker (1979)
RGBVI	Red Green Blue Vegetation Index	$(G^2\text{-}B \times R)/(G^2 + B \times R)$	Bendig et al. (2015)
VARI	Visible Atmospherically Resistant Index	(G-R)/(G + R-B)	Gitelson et al. (2002)
WI	Woebbecke Index	(G-B)/(R-G)	Woebbecke et al. (1995)

and Runger, 2013). To reduce the random error generated during the operation of the random forest algorithm, an average of 20 algorithms was set as the final importance score of each feature. Analysis of variance (ANOVA) was used to further test the ability of selected features to separate mild, moderate, and severe disease samples.

Classification model construction and evaluation

Using MATLAB R2016b (MathWorks, Natick, MA, USA), three algorithms, support vector machine (SVM), RF, and a back propagation neural network (BPNN) were the basis for the detection models of wheat FHB.

SVM is a supervised learning algorithm that realizes the best generalization ability and prevents overfitting by trying to find a compromise between the minimum calibration set error and the maximum edge error; it is one of the most powerful classifiers (Faris et al., 2017). SVM is expected to find an optimal hyperplane to divide the samples and ultimately create a convex quadratic programming problem that only provides global minima (avoiding local minima). When the variables cannot be separated linearly, SVM can use the kernel function to project variables into higher-dimensional feature space, which makes linear division easier (Xia et al., 2016). Compared with other classifiers that require a large number of samples, SVM can find the optimal solution on the basis of existing samples, so it has better applicability to limited samples, lower computational complexity, and less training time. The kernel function, kernel parameter size, and penalty parameter are important factors affecting the performance of the SVM model. We chose the radial basis function as the kernel function and used the grid optimization method to search for the best parameters to obtain better model accuracy.

The RF algorithm, proposed by Breiman (2001), is a popular ensemble learning algorithm in classification, prediction, and feature selection (Breiman, 2001). When using the RF algorithm for classification, the final label of the input sample is determined by voting for each decision tree in the random forest (Guo et al., 2011; Zhu et al., 2022b). Random resampling and node random splitting techniques are used to train the RF model (Gislason et al., 2006). RF is advantageous in remote sensing image processing (Rodriguez-Galiano et al., 2012): (1) RF is less computationally intensive than other tree ensemble methods (such as Boosting) and less prone to overfitting; (2) RF has a strong ability to resist noise and outliers, can tolerate a certain amount of data loss, and has good robustness to noise and outliers; (3) RF can analyze complex classification features and measure the importance of variables; (4) RF supports high dimensional data and generates an internal unbiased estimate of generalization error ("out of bag" error). In this study, the number of model decision trees was set to 200, and other parameters were kept as the default.

BPNN is one of the most widely used network models in remote sensing (Yang et al., 2011). It is a multi-layer feedforward neural network based on error backpropagation algorithm training, usually including an input layer, hidden layer, and output layer. When a set of information is inputted, the network can achieve the target accuracy through continuous repeated training and adjustment so as to produce satisfactory results. The algorithm continuously collects the errors generated by the model during the training period, returns these errors as output values through back propagation, and then continuously adjusts the weight of each neuron according to the error value. Finally, the best classification by the model is achieved.

A total of 100 samples with mild, moderate, and severe disease progression were randomly divided into the calibration set and prediction set (4:1 ratio). The calibration set was used for model construction, and the prediction set was preliminarily used to evaluate the capabilities of the model classification. To further evaluate the accuracy and prevent the model from overfitting, the validation set was used to verify the generalization ability of the model. We employed a five-fold cross-validation method to equally divide the dataset into five parts, each of which was an independent validation set. The accuracy of each validation set was evaluated, and the average was used as the final model validation accuracy. The calibration accuracy, prediction accuracy, and validation accuracy demonstrated the model's ability to detect wheat FHB in our study. Using ArcGIS 10.6 software to map the damage of wheat FHB, and calculate the ratio of the number of

infected pixels to the number of healthy pixels to statistics the proportion of wheat areas with different infection grades.

Results

UAV data quality verification based on canopy data

Figure 4A shows the original mean spectrum of canopy wheat measured by the ASD spectrometer and extracted from the UAV images over 450–950 nm. From the perspective of waveform similarity, the variations of the two spectra in the visible to near-infrared region (450–850 nm) are consistent, with significant peaks (near 550 nm) and troughs (near 680 nm). However, the spectrum measured by the ASD spectrometer is lower than the spectrum extracted by the UAV images overall. Above 850 nm, the spectral reflectance of the UAV images gradually decreases, and the spectral curve shows a significant downward trend compared with that of ASD, while the ASD spectral curve has little fluctuation. Figure 4B shows the correlation between UAV spectrum and ASD spectrum in the range of 450–950 nm. The two spectra are highly correlated, with R^2 above 0.97, which indicates that the image quality of UAV is trustworthy. The correlation between the UAV spectrum and ASD spectrum within 450–850 nm was further analyzed: R^2 reached 0.99 (Figure 4C). Thus, the band greater than 850 nm greatly influences the UAV images. Therefore, the last 100 bands of the UAV images were excluded from post-processing.

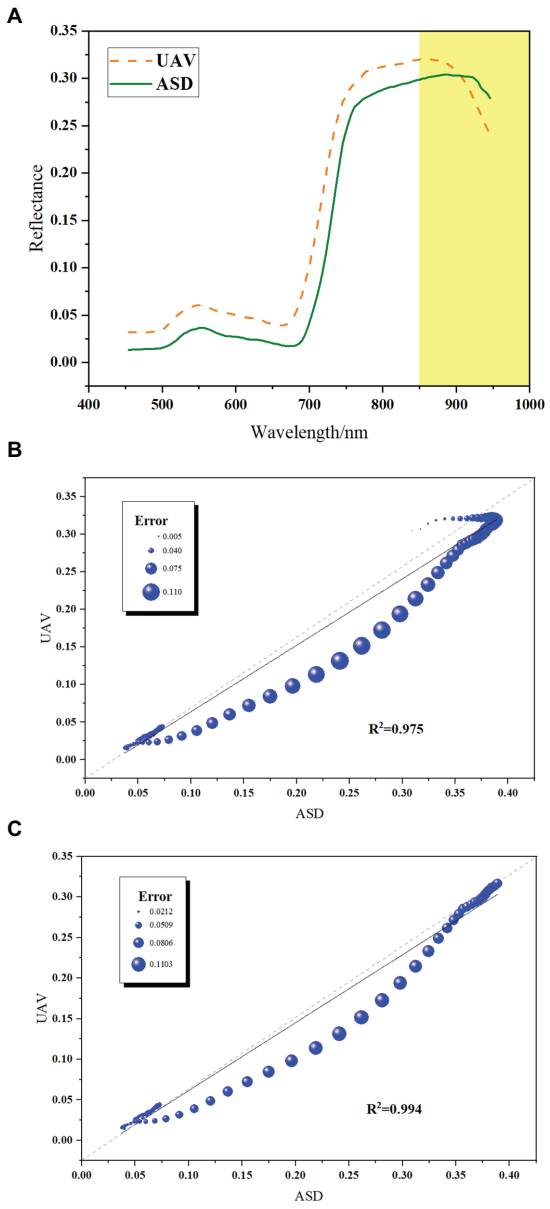


FIGURE 4
Curve comparison and correlation of UAV and ASD spectra. **(A)** Curves of ASD and UAV spectra. **(B)** Correlation between the two types of curves at 450–950nm. **(C)** Correlation between the two types of curves at 450–850nm.

Optimal spectral and image features

The RF algorithm was used to evaluate the importance of each feature in the FHB detection models to filter out redundant features. Figure 5 depicts the importance distribution of spectral and image features. The greater the weight, the more important the corresponding features. According to sequential backward elimination, all features with weights greater than 0.2 were selected to detect wheat samples with mild, moderate, and severe infection; the result was five spectral features, three texture features, and two color features (Table 3). The weights of the selected spectral features were much higher than remaining spectral features (Figure 5A). One selected spectral feature was located in the visible region, three were located in the red edge region, and one was located in the near-infrared region. For the image features, three texture features and two color features were selected to illustrate the distributions of disease and the degree of infection in wheat; the maximum weight of selected image features reached 0.39. Table 4 demonstrates the separation ability of these selected features to detect mild, moderate, and severe samples by ANOVA. In general, the selected features show different mean and standard deviation values among multi-class samples. There were significant differences among the mild, moderate, and severe samples of all features, and the significance level reached 0.95. Therefore, the selected features have strong separation ability to detect infected samples in this study.

Model construction

The purpose of our study is to effectively identify field FHB based on the fusion of spectral and image features of UAV images to be able to control the development of field diseases in a timely

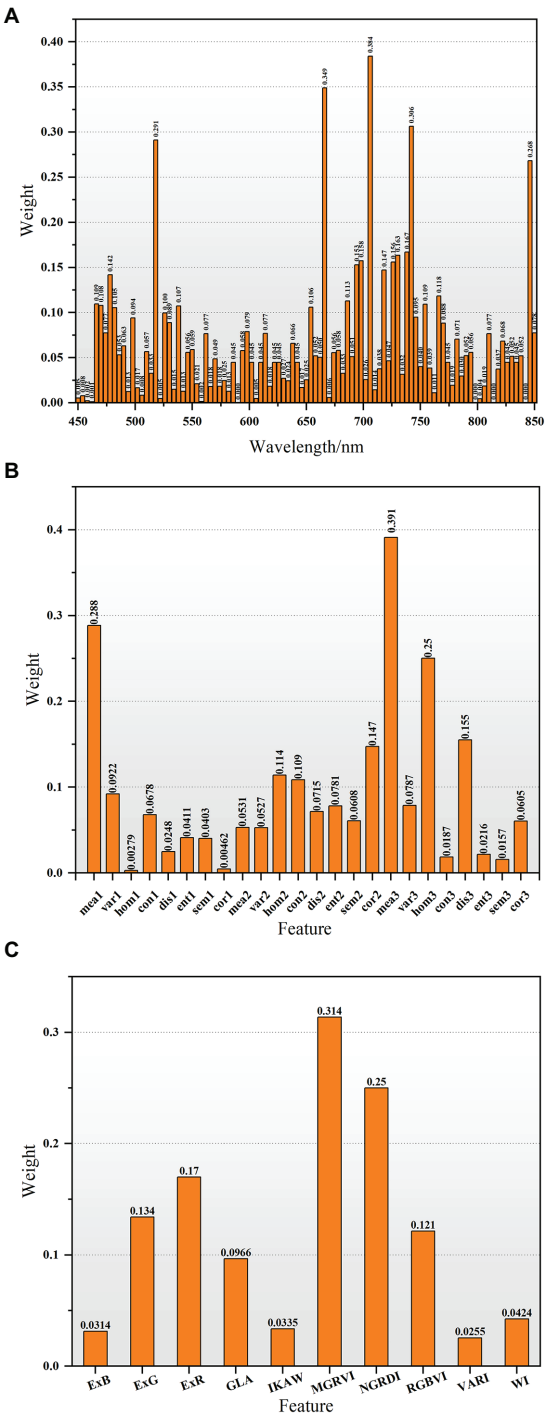


FIGURE 5
The importance distributions of various features based on the RF algorithm. **(A–C)** represent the weights of spectral features, texture features, and color features, respectively.

TABLE 3 The features selected by importance ranking.

Type	Variable number	Selected Features
Spectral features	5	band1(518 nm), band2(666 nm), band3(706 nm), band4(742 nm) and band5(846 nm)
Texture features	3	mean1, mean3 and hom3
Color features	2	MGRVI and NGRDI

TABLE 4 Statistical characteristics of feature values of the mild, moderate, and severe disease samples.

Feature	Sample category	Mean of feature	Std. deviation	*P*-Value (ANOVA)
band1	Mild	0.059	0.013	0.002
	Moderate	0.063	0.014	
	Severe	0.076	0.011	
band2	Mild	0.054	0.017	0.035
	Moderate	0.057	0.019	
	Severe	0.071	0.016	
band3	Mild	0.155	0.041	0.001
	Moderate	0.168	0.039	
	Severe	0.209	0.030	
band4	Mild	0.312	0.070	0.000
	Moderate	0.345	0.065	
	Severe	0.418	0.054	
band5	Mild	0.374	0.084	0.000
	Moderate	0.411	0.079	
	Severe	0.496	0.062	
mea1	Mild	21.88	3.383	0.038
	Moderate	21.32	4.067	
	Severe	18.56	2.238	
mea3	Mild	37.03	11.207	0.003
	Moderate	32.66	10.748	
	Severe	24.12	6.161	
hom3	Mild	0.78	0.112	0.019
	Moderate	0.80	0.083	
	Severe	0.76	0.100	
MGRVI	Mild	−0.32	0.037	0.031
	Moderate	−0.35	0.059	
	Severe	−0.40	0.046	
NGRDI	Mild	−0.16	0.024	0.030
	Moderate	−0.18	0.032	
	Severe	−0.21	0.027	

manner. Therefore, the classification models were developed by combining different feature fusion with SVM, RF, and BPNN for the analysis of wheat FHB detection. The calibration accuracy and prediction accuracy of the models are shown in Table 5. The precisions of models constructed based on different feature variables are significantly different. The integration of spectral, texture, and color features seems to achieve the best accuracy. When spectral features were used as model inputs, the RF model performed best with a prediction accuracy of 70%, followed by BPNN and SVM with prediction accuracies of 65 and 60%, respectively. When considering the integration of spectral and texture features, the accuracy of the three classification models

TABLE 5 Model classification accuracy based on different features and algorithms.

Feature	Classification algorithm	Calibration accuracy (%)	Prediction accuracy (%)	Validation accuracy (%)
Spectral	RF	100	70	70
	SVM	63	60	59
	BPNN	78	65	72
Spectral + texture	RF	100	80	79
	SVM	70	70	60
	BPNN	76	75	76
Spectral + texture + color	**RF**	**100**	**85**	**83**
	SVM	74	70	63
	BPNN	84	80	83

Bold values indicate the optimal algorithm and highest accuracy.

was improved by 10%, and the RF model achieved the highest accuracy at 80%. When spectral, texture, and color features were integrated as input variables, the prediction accuracy of the RF model was further improved to 85%. The prediction accuracy of the SVM model remained unchanged, but the prediction accuracy of the BPNN model was also improved by 5%. The calibration accuracy of the model also shows the same trend as the prediction accuracy. Among all models, the calibration accuracy of the RF model reached 100%. With the addition of image features, the calibration accuracy of the model continued to improve. The above results indicate that the fusion of spectral and image features can improve the performance of the model through texture and color features in terms of identifying wheat FHB. The five-fold cross-validation method was used to further verify the model to prove its universality. The validation results are shown in Table 5. The results show that the highest validation accuracy was 83%, which is reflected in the integration of spectral feature, texture feature, color feature, and RF algorithm. The above results show that the spectral and image feature fusion combined with the RF algorithm can benefit the rapid detection and accurate analysis of a wheat field with mild, moderate, and severe infection.

To understand the spatial distribution of FHB-infected wheat in the study area, models based on different feature integrations and the optimal RF algorithm were adapted to map the damage of wheat FHB on May 3 and 8, 2019. The results are shown in Figure 6. From the mapping results, the wheat infection degree increased over time. Although the infection had spread over the entire farmland on May 3, the wheat in the field showed mild and moderate infection, and severe infection was almost zero (sporadic distribution). However, on May 8, almost all wheat in the study area showed moderate or severe infection, indicating that a large outbreak rapidly occurred. Table 6 summarizes the proportions of wheat area with mild infection, moderate infection, and severe infection corresponding to each figure in Figure 6. Moderate infection impacted more than 75% of the wheat on May 8, and the severe infection impacted more than 10%. The addition of image features improved the model in terms of detecting severely infected wheat. The proportion of severely infected wheat on May 8 in Figure 6B (18.12%) and Figure 6C (18.85%) is higher than that in Figure 6A (11.73%); these results are mainly reflected in the presence of some severe infection along the edge of the plot. This severe infection phenomenon is consistent with our field survey results.

Discussion

In the present work, the detailed information contained in UAV hyperspectral images were fully exploited to help identify wheat FHB in the field. FHB can change the pigment, water content, and cell structure of wheat, as well as the structure, shape, and color of the wheat canopy. Therefore, we fused the spectral features that represent internal physiological changes with the image features that represent the spatial information of wheat to effectively detect wheat FHB.

Before analyzing the UAV images, we first evaluated the quality of the UAV hyperspectral images, which is a critical step to ensure that the UAV images accurate identify FHB. We evaluated the quality of the UAV images by comparing and analyzing the data obtained from an ASD spectrometer. The spectral curves of the wheat sample points extracted from the UAV images share a common trend with those of the ASD spectral: a peak and a trough in the VIS–NIR region. However, the values of the ASD spectra were lower than the spectral values obtained from the UAV images, which is likely due to the influence of the bidirectional reflectance distribution function (BRDF) caused by the difference in the geometrical positions of the sun-target-sensors of the two data sets. Some studies have proven that BRDF has a significant impact on UAV hyperspectral data (Burkart et al., 2015). Above 850 nm, the UAV spectral curve shows a downward trend compared with the ASD spectral curve, while the ASD spectral curve has little variation, which is consistent with other scholars' observations (Gao et al., 2016; Chen et al., 2018). The sensor may have too much noise at the detection boundary. Furthermore, it need cloud-free conditions on the measurement day, so there is a long time interval between the UAV flight and ASD information collection, as well as changes in light conditions. According to Figure 4, the spectral reflectance of sample points obtained by different sensors is significantly correlated within 450–850 nm (R^2 of 0.99). ASD hyperspectral data are an extensive

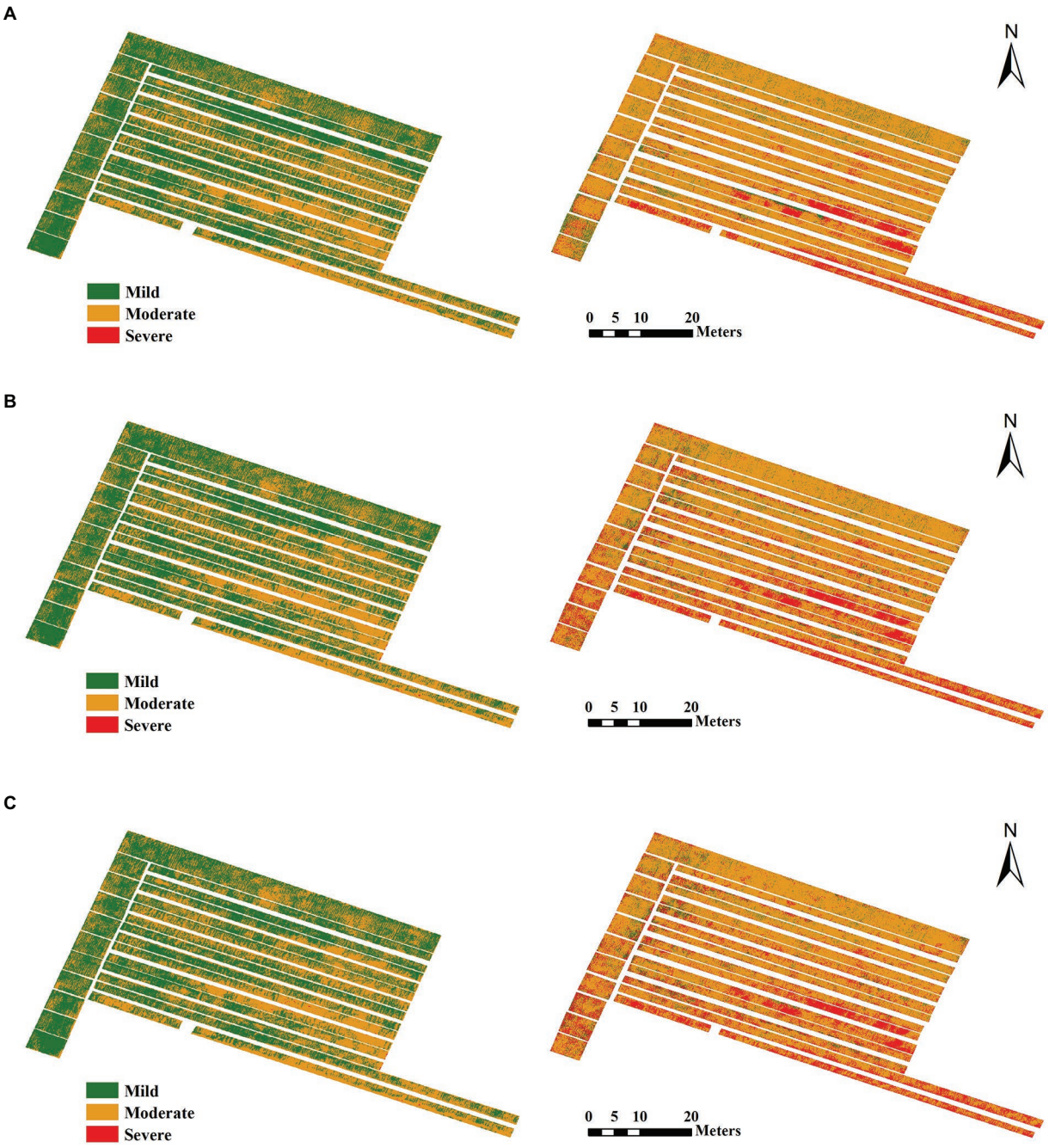


FIGURE 6
Damage maps for May 3 (left) and May 8 (right) based on different feature combinations and the RF algorithm. **(A)** Spectral features. **(B)** Spectral and texture features. **(C)** Spectral, texture, and color features.

remote sensing identification method for crop pests and diseases (Cao et al., 2013; Shi et al., 2018; Guo et al., 2020). This spectrum has been used to accurately identify field wheat FHB (Huang et al., 2019a,b; Ma et al., 2020). The high correlation between UAV and ASD spectra further proves the reliability of UAV images.

Next, we extracted the band features, texture features, and color features contained in the hyperspectral images. Table 3 shows the details of the extracted features. The band features we extracted are mainly located in the green edge, red edge, and near-infrared region. The green edge is mainly related to the content of wheat pigments (including carotenoids and chlorophyll) (Al Masri et al., 2017; Zhang Z. P. et al., 2020b), and the position of the red edge is sensitive to the movement of the red edge caused by the change of chlorophyll concentrations (Zhang Z. P. et al., 2020a). Near-infrared wavelengths are primarily related to wheat moisture content,

TABLE 6 The percentages of mildly, moderately, and severely infected wheat corresponding to the damage maps.

Feature	Data	Mild (%)	Moderate (%)	Severe (%)	Sum (%)
Spectral	May 3	57.16	42.76	0.08	100
	May 8	5.72	82.55	11.73	100
Spectral + texture	May 3	55.45	44.45	0.10	100
	May 8	5.67	76.21	18.12	100
Spectral + texture + color	May 3	53.82	46.11	0.08	100
	May 8	5.26	75.88	18.85	100

as FHB-infected wheat is accompanied by a temporary increase in transpiration and tissue desiccation (Bauriegel et al., 2011a). The VIS–NIR bands in hyperspectral images are proposed to overcome visual symptom disassociations with DON contamination. Because the DON concentration of wheat is at a low level in the early stages and the typical symptoms of *Fusarium* damage cannot be detected visually, the spectral features are more conducive to observing the early symptoms of wheat infection (Femenias et al., 2020; Zhang et al., 2021). The GLCM-based texture feature extraction method was based on Fu et al. (2022). The GLCM method describes the texture by studying the spatial correlation characteristics of the gray levels (Haralick and Shanmugam, 1973). In fact, texture information can help distinguish the spatial information independent of tone to identify objects or regions of interest in an image, but it is not recommended to use it by itself due to the poor performance of texture parameters (Sarker and Nichol, 2011). Previously, auxiliary texture information was effectively combined with spectral information to significantly improve the accuracy of wheat GPC estimation (Fu et al., 2022). Therefore, in this study, we attempted to fuse texture and spectral features to improve the detection accuracy of field FHB. The results demonstrate that texture features can serve as complementary information to increase the dimensionality of UAV hyperspectral image data (Table 5). In addition, we calculated some color features by band combinations to indicate different aspects of wheat infection. While texture features may add additional information to FHB estimation, crop infection is more directly related to color information rather than the spatial arrangement of colors (Li et al., 2019). What's more, since color images highlight specific vegetation greenness and are considered to be less sensitive to changes in light conditions, color features extracted from RGB images have the potential to provide crop growth and nutritional status, immediately providing researchers and farmers with a realistic and intuitive visualization of crop growth status (Du and Noguchi, 2017; Ge et al., 2021). At present, some scholars use color features to estimate the nitrogen density of winter wheat leaves (Rorie et al., 2011), estimate the leaf area index of rice (Li et al., 2019), monitor the growth status of wheat (Du and Noguchi, 2017), and accurately detect wheat FHB at the spikes scale (Huang et al., 2020). However, the effect of FHB detection of field wheat based on color features has not been explored yet. Therefore, in this study, we further supplemented color features in the input models based on spectral and texture features to identify wheat FHB. Actually, color features in UAV digital images are usually based on RGB cameras because UAV systems with RGB cameras are inexpensive, compact, and convenient. In the future, a UAV system suitable for FHB monitoring in the field should be considered. The RGB bands in this study are a basis for future RGB cameras, avoiding the complexity of hyperspectral data processing.

Table 5 shows the model classification results of field wheat with different degrees of infection according to different input variables. The addition of texture and color features can further improve the accuracy of the model compared to methods that use spectral features to detect wheat FHB. As seen in Table 6, the improvement of accuracy is mainly manifested in the difference in the model's detection of mild, moderate, and severe disease samples. In the early stage of wheat FHB infection (May 3), 57.16% and 42.76% of the field wheat with mild and moderate infection, respectively, could be identified by the model using spectral features. With the addition of texture features and color features, the proportion of mildly infected wheat in the field identified by the model gradually decreased, and the proportion of moderately infected wheat increased. In the late stage of wheat FHB infection (May 8), the model indicated that the proportions of mildly and moderately infected wheat gradually decreased with the addition of image features, and the severely infected area gradually increased. That is to say, before the image features are added, the model always misses the wheat with more severe disease. In fact, as the wheat infection spread, the dry and white areas of the wheat ears became larger until the wheat died (Huang et al., 2020). The addition of image features can enable the model to capture this process. When the information contained in the model increases, the detection of samples with severe disease improves, which is consistent with the results in our study.

The research shows that the fusion of spectral and image features can distinguish the disease incidence of wheat in the field; this method can help future precision agriculture and large area wheat FHB monitoring. However, current research still exists limitations. In addition to considering spectral features, texture features, and color features, some vegetation features, such as Structure Insensitive Pigment Index (SIPI), Anthocyanin Reflectance Index (ARI), Normalized Difference Vegetation Index (NDVI), and Plant Senescence Reflection Index (PSRI), are often used to reflect plant disease stress status (Xiao et al., 2021). Hence, the effectiveness of the vegetation features in wheat FHB detection based on UAV hyperspectral images is worth considering. Additionally, only three machine learning algorithms (RF, BPNN, and SVM) were used in this study. The generalization ability of the models in the temporal and spatial dimensions must be verified. Further consideration can be given to combining data augmentation and deep learning methods to develop more stable and independent models, as well as reduce

the uncertainty of model applicability in other regions. Scale has become a popular topic in remote sensing research, and the information contained in a single pixel under different resolutions will change significantly. Appropriate spatial resolution images for agricultural monitoring are needed (Na et al., 2016). Our study only used two images with a spatial resolution of 4 cm to detect FHB, which is relatively simple. Various spatial resolution images are worth considering in the future. Finally, the occurrence of wheat FHB is related to the time of infection (Alisaac et al., 2020) and meteorological factors, such as temperature and humidity. In the future, we will aim to consider wheat infection time and meteorological factors to explore early FHB detection methods and effectively prevent and control FHB occurrence and outbreak. The influence of wheat varieties and the development of various pests and diseases on the spread of wheat FHB cannot be ignored. More researches are needed to investigate the influence of varieties and multiple infections on the model performance in the future.

Conclusion

In this study, the quantitative detection of wheat with mild, moderate, and severe FHB infection in the field was achieved by fusing spectral and image features extracted from the UAV hyperspectral images. After obtaining the hyperspectral images, we first evaluated the quality of the images and identified the data in the 450–850 nm band for subsequent analysis by comparing waveform similarity and correlation with ASD hyperspectral data. Then, we extracted the spectral features that reflect the physiological and biochemical changes within the host, as well as the texture and color features that characterize the spatial changes of wheat. The RF algorithm was used to further eliminate redundant features and improve the operating efficiency of the model. Finally, FHB quantitative detection models, based on different combinations of spectral features, texture features, and color features were formulated by combining BPNN, SVM, and RF algorithms. We evaluated the classification results of the different models, and the FHB-related wheat damage was mapped using the best algorithm. The results show that the spectral features can potentially determine the damage level of FHB, but the performance of the models is not satisfactory. The fusion of spectral features and texture features can improve the model detection level, but the maximum prediction accuracy of the models was only 80%. The model based on the fusion of spectral, texture, and color features was best, and the prediction accuracy of the RF algorithm reached 85%. The damage map illustrates that wheat FHB developed very rapidly over a short time, causing destruction of the crop. This study builds upon previous models in terms of feature types, monitoring methods, and monitoring areas and provides a new methodology for FHB detection in the field by deeply mining features in UAV images and combining multiple spectral advantages.

Data availability statement

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

Author contributions

HZ: data curation, methodology, and writing-original draft. HM and LL: investigation and data acquisition. JZ and SW: formal analysis, methodology, and supervision. LH: conceptualization and funding acquisition. YD and WH: writing-review and editing. All authors contributed to the article and approved the submitted version.

Funding

The work presented here was supported by National Key R&D Program of China (2021YFB3901303), National Natural Science Foundation of China (42071423), Program of Bureau of International Cooperation, Chinese Academy of Sciences (183611KYSB20200080), Alliance of International Science Organizations (Grant No. ANSO-CR-KP-2021-06), Natural Science Research Project of Anhui Provincial Education Department (KJ2019A0030).

Acknowledgments

The authors thank Kehui Ren from the University of Chinese Academy of Sciences for his hard work in the experiment and Hong Chang from the National Engineering Research Center for Information Technology in Agricultural for her guidance.

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

References

Al Masri, A., Hau, B., Dehne, H. W., Mahlein, A. K., and Oerke, E. C. (2017). Impact of primary infection site of fusarium species on head blight development in wheat ears evaluated by IR-thermography. *Eur. J. Plant Pathol.* 147, 855–868. doi: 10.1007/s10658-016-1051-2

Alisaac, E., Rathgeb, A., Karlovsky, P., and Mahlein, A. K. (2020). Fusarium head blight: effect of infection timing on spread of fusarium graminearum and spatial distribution of deoxynivalenol within wheat spikes. *Microorganisms* 9:1 doi: 10.3390/microorganisms9010079

Alisaac, E., Behmann, J., Rathgeb, A., Karlovsky, P., Dehne, H.-W., and Mahlein, A.-K. (2019). Assessment of fusarium infection and mycotoxin contamination of wheat kernels and flour using hyperspectral imaging. *Toxins* 11:556. doi: 10.3390/toxins11100556

Amar, A. B., Oueslati, S., Ghorbel, A., and Mliki, A. (2012). Prediction and early detection of mycotoxigenic fusarium culmorum in wheat by direct PCR-based procedure. *Food Control* 23, 506–510. doi: 10.1016/j.foodcont.2011.08.021

Ashourloo, D., Mobasheri, M., and Huete, A. (2014). Evaluating the effect of different wheat rust disease symptoms on vegetation indices using hyperspectral measurements. *Remote Sens.* 6, 5107–5123. doi: 10.3390/rs6065107

Atoui, A., El Khoury, A., Kallassy, M., and Lebrihi, A. (2012). Quantification of fusarium graminearum and fusarium culmorum by real-time PCR system and zearalenone assessment in maize. *Int. J. Food Microbiol.* 154, 59–65. doi: 10.1016/j.ijfoodmicro.2011.12.022

Barbedo, J. G., Tibola, C. S., and Fernandes, J. M. (2015). Detecting fusarium head blight in wheat kernels using hyperspectral imaging. *Biosyst. Eng.* 131, 65–76. doi: 10.1016/j.biosystemseng.2015.01.003

Bareth, G., Aasen, H., Bendig, J., Gnyp, M. L., Bolten, A., Jung, A., et al. (2015). Low-weight and UAV-based hyperspectral full-frame cameras for monitoring crops: spectral comparison with portable spectroradiometer measurements. *Photogramm. Fernerkund. Geoinf.* 103:7. doi: 10.1127/pfg/2015/0256

Bauriegel, E., Giebel, A., and Herppich, W. B. (2011a). Hyperspectral and chlorophyll fluorescence imaging to analyse the impact of fusarium culmorum on the photosynthetic integrity of infected wheat ears. *Sensors* 11, 3765–3779. doi: 10.3390/s110403765

Bauriegel, E., Giebel, A., Geyer, M., Schmidt, U., and Herppich, W. (2011b). Early detection of fusarium infection in wheat using hyper-spectral imaging. *Comput. Electron. Agric.* 75, 304–312. doi: 10.1016/j.compag.2010.12.006

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., et al. (2015). Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Obs. Geoinf.* 39, 79–87. doi: 10.1016/j.jag.2015.02.012

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

Burkart, A., Aasen, H., Alonso, L., Menz, G., Bareth, G., and Rascher, U. (2015). Angular dependency of hyperspectral measurements over wheat characterized by a novel UAV based goniometer. *Remote Sens.* 7, 725–746. doi: 10.3390/rs70100725

Cao, X. R., Luo, Y., Zhou, Y. L., Duan, X. Y., and Cheng, D. F. (2013). Detection of powdery mildew in two winter wheat cultivars using canopy hyperspectral reflectance. *Crop Prot.* 45, 124–131. doi: 10.1016/j.cropro.2012.12.002

Chen, P. F., Li, G., Shi, Y. J., Xu, Z. T., Yang, F. T., and Cao, Q. J. (2018). Validation of an unmanned aerial vehicle hyperspectral sensor and its application in maize leaf area index estimation. *Sci. Agric. Sin.* 51, 1464–1474. doi: 10.3864/j.issn.0578-1752.2018.08.004

Dammer, K.-H., Möller, B., Rodemann, B., and Heppner, D. (2011). Detection of head blight (fusarium ssp.) in winter wheat by color and multispectral image analyses. *Crop Prot.* 30, 420–428. doi: 10.1016/j.cropro.2010.12.015

Delwiche, S. R., Kim, M. S., and Dong, Y. (2011). Fusarium damage assessment in wheat kernels by Vis/NIR hyperspectral imaging. *Sens. Instrumen. Food Qual.* 5, 63–71. doi: 10.1007/s11694-011-9112-x

Deng, H., and Runger, G. (2013). Gene selection with guided regularized random forest. *Pattern Recogn.* 46, 3483–3489. doi: 10.1016/j.patcog.2013.05.018

Du, M., and Noguchi, N. (2017). Monitoring of wheat growth status and mapping of wheat yield's within-field spatial variations using color images acquired from UAV-camera system. *Remote Sens.* 9:3. doi: 10.3390/rs9030289

Faris, H., Hassonah, M. A., Al-Zoubi, A. M., Mirjalili, S., and Aljarah, I. (2017). A multi-verse optimizer approach for feature selection and optimizing SVM parameters based on a robust system architecture. *Neural Comput. Appl.* 30, 2355–2369. doi: 10.1007/s00521-016-2818-2

Femenias, A., Gatius, F., Ramos, A. J., Sanchis, V., and Marín, S. (2020). Standardisation of near infrared hyperspectral imaging for quantification and classification of don contaminated wheat samples. *Food Control* 111:107074. doi: 10.1016/j.foodcont.2019.107074

Fu, Z. P., Yu, S. S., Zhang, J. Y., Xi, H., Gao, Y., Lu, R. H., et al. (2022). Combining UAV multispectral imagery and ecological factors to estimate leaf nitrogen and grain protein content of wheat. *Eur. J. Agron.* 132. doi: 10.1016/j.eja.2021.126405

Gao, L., Yang, G. J., Yu, H. Y., Xu, B., Zhao, X. Q., Dong, J. H., et al. (2016). Retrieving winter wheat leaf area index based on unmanned aerial vehicle hyperspectral remote sensing. *Trans. Chin. Soc. Agric. Eng.* 32, 113–120. doi: 10.11975/j.issn.1002-6819.2016.22.016

Ge, H. X., Xiang, H. T., Ma, F., Li, Z. W., Qiu, Z. C., Tan, Z. Z., et al. (2021). Estimating plant nitrogen concentration of rice through fusing vegetation indices and color moments derived from UAV-RGB images. *Remote Sens.* 13:9. doi: 10.3390/rs13091620

Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recogn. Lett.* 27, 294–300. doi: 10.1016/j.patrec.2005.08.011

Gitelson, A. A., Kaufman, Y. J., Stark, R., and Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* 80, 76–87. doi: 10.1016/S0034-4257(01)00289-9

Guo, A. T., Huang, W. J., Ye, H. C., Dong, Y. Y., Ma, H. Q., Ren, Y., et al. (2020). Identification of wheat yellow rust using spectral and texture features of hyperspectral images. *Remote Sens.* 12:9. doi: 10.3390/rs12091419

Guo, L., Chehata, N., Mallet, C., and Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using random forests. *ISPRS J. Photogramm. Remote Sens.* 66, 56–66. doi: 10.1016/j.isprsjprs.2010.08.007

Haralick, R. M., and Shanmugam, K. (1973). Textural features for image classification. *IEEE Trans. Syst. Man Cybern.-Syst.* SMC-3, 610–621. doi: 10.1109/TSMC.1973.4309314

Huang, L. S., Wu, Z. C., Huang, W. J., Ma, H. Q., and Zhao, J. L. (2019a). Identification of fusarium head blight in winter wheat ears based on fisher's linear discriminant analysis and a support vector machine. *Appl. Sci.* 9:3894. doi: 10.3390/app9183894

Huang, L. S., Zhang, H. S., Ding, W. J., Huang, W. J., Hu, T. G., and Zhao, J. L. (2019b). Monitoring of wheat scab using the specific spectral index from ASD hyperspectral dataset. *J. Spectrosc.* doi: 10.1155/2019/9153195

Huang, L. S., Li, T. K., Ding, C. L., Zhao, J. L., Zhang, D. Y., and Yang, G. J. (2020). Diagnosis of the severity of fusarium head blight of wheat ears on the basis of image and spectral feature fusion. *Sensors* 20:2887. doi: 10.3390/s20102887

Huang, L. S., Wu, K., Huang, W. J., Dong, Y. Y., Ma, H. Q., Liu, Y., et al. (2021). Detection of fusarium head blight in wheat ears using continuous wavelet analysis and PSO-SVM. *Agriculture* 11:998. doi: 10.3390/agriculture11100998

Jaillais, B., Roumet, P., Pinson-Gadais, L., and Bertrand, D. (2015). Detection of fusarium head blight contamination in wheat kernels by multivariate imaging. *Food Control* 54, 250–258. doi: 10.1016/j.foodcont.2015.01.048

Kawashima, S., and Nakatani, M. (1998). An algorithm for estimating chlorophyll content in leaves using a video camera. *Ann. Bot.* 81, 49–54. doi: 10.1006/anbo.1997.0544

Li, S. Y., Yuan, F., Ata-Ui-Karim, S. T., Zheng, H. B., Cheng, T., Liu, X. J., et al. (2019). Combining color indices and textures of UAV-based digital imagery for rice LAI estimation. *Remote Sens.* 11:1763. doi: 10.3390/rs11151763

Li, X. C., Zhang, Y. J., Bao, Y. S., Luo, J. H., Jin, X. L., Xu, X. G., et al. (2014). Exploring the best hyperspectral features for LAI estimation using partial least squares regression. *Remote Sens.* 6, 6221–6241. doi: 10.3390/rs6076221

Liang, K., Huang, J. N., He, R. Y., Wang, Q. J., Chai, Y. Y., and Shen, M. X. (2020). Comparison of Vis-NIR and SWIR hyperspectral imaging for the non-destructive detection of don levels in fusarium head blight wheat kernels and wheat flour. *Infrared Phys. Techn.* 106:103281. doi: 10.1016/j.infrared.2020.103281

Liu, L. Y., Dong, Y. Y., Huang, W. J., Du, X. P., and Ma, H. Q. (2020a). Monitoring wheat fusarium head blight using unmanned aerial vehicle hyperspectral imagery. *Remote Sens.* 12:22. doi: 10.3390/rs12223811

Liu, L. Y., Dong, Y. Y., Huang, W. J., Du, X. P., Ren, B. Y., Huang, L. S., et al. (2020b). A disease index for efficiently detecting wheat fusarium head blight using sentinel-2 multispectral imagery. *IEEE Access* 8, 52181–52191. doi: 10.1109/ACCESS.2020.2980310

Louhaichi, M., Borman, M. M., and Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto Int.* 16, 65–70. doi: 10.1080/10106040108542184

Ma, H., Huang, W., Jing, Y., Pignatti, S., Laneve, G., Dong, Y., et al. (2020). Identification of fusarium head blight in winter wheat ears using continuous wavelet analysis. *Sensors* 20:20. doi: 10.3390/s20010020

Ma, H. Q., Huang, W. J., Dong, Y. Y., Liu, L. Y., and Guo, A. T. (2021). Using UAV-based hyperspectral imagery to detect winter wheat fusarium head blight. *Remote Sens.* 13:15. doi: 10.3390/rs13153024

Maragos, C., Busman, M., and Sugita-Konishi, Y. (2006). Production and characterization of a monoclonal antibody that cross-reacts with the mycotoxins nivalenol and 4-deoxynivalenol. *Food Addit. Contam.* 23, 816–825. doi: 10.1080/02652030600699072

Menesatti, P., Antonucci, F., Pallottino, F., Giorgi, S., Matere, A., Nocente, F., et al. (2013). Laboratory vs. in-field spectral proximal sensing for early detection of fusarium head blight infection in durum wheat. *Biosyst. Eng.* 114, 289–293. doi: 10.1016/j.biosystemseng.2013.01.004

Meyer, G. E., and Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Comput. Electron. Agric.* 63, 282–293. doi: 10.1016/j.compag.2008.03.009

Na, L., Gaodi, X., Demin, Z., Changshun, Z., and Cuicui, J. (2016). Remote sensing classification of marsh wetland with different resolution images. *J. Resour. Ecol.* 7, 107–114. doi: 10.5814/j.issn.1674-764x.2016.02.005

Pal, M., and Foody, G. M. (2010). Feature selection for classification of hyperspectral data by SVM. *IEEE T. Geosci. Remote Sens.* 48, 2297–2307. doi: 10.1109/TGRS.2009.2039484

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* 67, 93–104. doi: 10.1016/j.isprsjprs.2011.11.002

Rorie, R. L., Purcell, L. C., Mozaffari, M., Karcher, D. E., King, C. A., Marsh, M. C., et al. (2011). Association of "greenness" in corn with yield and leaf nitrogen concentration. *Agron. J.* 103, 529–535. doi: 10.2134/agronj2010.0296

Sarker, L. R., and Nichol, J. E. (2011). Improved forest biomass estimates using ALOS AVNIR-2 texture indices. *Remote Sens. Environ.* 115, 968–977. doi: 10.1016/j.rse.2010.11.010

Shi, Y., Huang, W., González-Moreno, P., Luke, B., Dong, Y., Zheng, Q., et al. (2018). Wavelet-based rust spectral feature set (WRSFS): a novel spectral feature set based on continuous wavelet transformation for tracking progressive host–pathogen interaction of yellow rust on wheat. *Remote Sens.* 10:4. doi: 10.3390/rs10040525

Simsek, S., Burgess, K., Whitney, K. L., Gu, Y., and Qian, S. Y. (2012). Analysis of deoxynivalenol and deoxynivalenol-3-glucoside in wheat. *Food Control* 26, 287–292. doi: 10.1016/j.foodcont.2012.01.056

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8, 127–150. doi: 10.1016/0034-4257(79)90013-0

Whetton, R. L., Waine, T. W., and Mouazen, A. M. (2018a). Hyperspectral measurements of yellow rust and fusarium head blight in cereal crops: part 2: on-line field measurement. *Biosyst. Eng.* 167, 144–158. doi: 10.1016/j.biosystemseng.2018.01.004

Whetton, R. L., Hassall, K. L., Waine, T. W., and Mouazen, A. M. (2018b). Hyperspectral measurements of yellow rust and fusarium head blight in cereal crops: part 1: laboratory study. *Biosyst. Eng.* 166, 101–115. doi: 10.1016/j.biosystemseng.2017.11.008

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

Xia, J. S., Chanussot, J., Du, P. J., and He, X. Y. (2016). Rotation-based support vector machine ensemble in classification of hyperspectral data with limited training samples. *IEEE T. Geosci. Remote Sens.* 54, 1519–1531. doi: 10.1109/TGRS.2015.2481938

Xiao, Y., Dong, Y., Huang, W., Liu, L., and Ma, H. (2021). Wheat fusarium head blight detection using UAV-based spectral and texture features in optimal window size. *Remote Sens.* 13:13. doi: 10.3390/rs13132437

Yang, Y., Zhu, J., Zhao, C., Liu, S., and Tong, X. (2011). The spatial continuity study of NDVI based on kriging and BPNN algorithm. *Math. Comput. Model.* 54, 1138–1144. doi: 10.1016/j.mcm.2010.11.046

Ye, H. C., Huang, W. J., Huang, S. Y., Cui, B., Dong, Y. Y., Guo, A. T., et al. (2020). Recognition of banana fusarium wilt based on UAV remote sensing. *Remote Sens.* 12:6. doi: 10.3390/rs12060938

Zhang, D. Y., Zhou, X. G., Zhang, J., Lan, Y. B., Xu, C., and Liang, D. (2018). Detection of rice sheath blight using an unmanned aerial system with high-resolution color and multispectral imaging. *PLoS One* 13:e0187470. doi: 10.1371/journal.pone.0187470

Zhang, D. Y., Wang, Q., Lin, F. F., Yin, X., Gu, C. Y., and Qiao, H. B. (2020a). Development and evaluation of a new spectral disease index to detect wheat fusarium head blight using hyperspectral imaging. *Sensors* 20:2260. doi: 10.3390/s20082260

Zhang, D. Y., Wang, D. Y., Gu, C. Y., Jin, N., Zhao, H. T., 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, D. Y., Chen, G., Zhang, H. H., Jin, N., Gu, C. Y., Weng, S. Z., et al. (2020b). Integration of spectroscopy and image for identifying fusarium damage in wheat kernels. *Spectrochim. Acta Part A* 236:118344. doi: 10.1016/j.saa.2020.118344

Zhang, X., Cui, J. T., Wang, W. S., and Lin, C. (2017). A study for texture feature extraction of high-resolution satellite images based on a direction measure and gray level co-occurrence matrix fusion algorithm. *Sensors* 17:7. doi: 10.3390/s17071474

Zhang, Z. P., Ding, J. L., Wang, J. Z., and Ge, X. Y. (2020a). Prediction of soil organic matter in northwestern China using fractional-order derivative spectroscopy and modified normalized difference indices. *Catena* 185:104257. doi: 10.1016/j.catena.2019.104257

Zhang, Z. P., Ding, J. L., Zhu, C. M., and Wang, J. Z. (2020b). Combination of efficient signal pre-processing and optimal band combination algorithm to predict soil organic matter through visible and near-infrared spectra. *Spectrochim. Acta Part A* 240:118553. doi: 10.1016/j.saa.2020.118553

Zhang, Z. P., Ding, J. L., Zhu, C. M., Wang, J. Z., Ma, G. L., Ge, X. Y., et al. (2021). Strategies for the efficient estimation of soil organic matter in salt-affected soils through Vis-NIR spectroscopy: optimal band combination algorithm and spectral degradation. *Geoderma* 382:114729. doi: 10.1016/j.geoderma.2020.114729

Zheng, Q., Huang, W. J., Cui, X. M., Dong, Y. Y., Shi, Y., Ma, H. Q., et al. (2018). Identification of wheat yellow rust using optimal three-band spectral indices in different growth stages. *Sensors* 19:1. doi: 10.3390/s19010035

Zhu, C. M., Ding, J. L., Zhang, Z. P., and Wang, Z. (2022a). Exploring the potential of UAV hyperspectral image for estimating soil salinity: effects of optimal band combination algorithm and random forest. *Spectrochim. Acta Part A* 279:121416. doi: 10.1016/j.saa.2022.121416

Zhu, C. M., Ding, J. L., Zhang, Z. P., Wang, J. J., Wang, Z., Chen, X. Y., et al. (2022b). SPAD monitoring of saline vegetation based on Gaussian mixture model and UAV hyperspectral image feature classification. *Comput. Electron. Agric.* 200:107236. doi: 10.1016/j.compag.2022.107236

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 December 2022
DOI 10.3389/fpls.2022.1080745

OPEN ACCESS

EDITED BY
Baohua Zhang,
Nanjing Agricultural University, China

REVIEWED BY
Zichen Huang,
Kyoto University, Japan
Jiangbo Li,
Beijing Academy of Agriculture and Forestry Sciences, China
Cedric Okinda,
Masinde Muliro University of Science and Technology, Kenya

*CORRESPONDENCE
Yong He
yhe@zju.edu.cn

†These two authors contribute equally to this manuscript

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

RECEIVED 26 October 2022
ACCEPTED 24 November 2022
PUBLISHED 20 December 2022

CITATION
Xiao Q, Wu N, Tang W, Zhang C, Feng L, Zhou L, Shen J, Zhang Z, Gao P and He Y (2022) Visible and near-infrared spectroscopy and deep learning application for the qualitative and quantitative investigation of nitrogen status in cotton leaves. *Front. Plant Sci.* 13:1080745. doi: 10.3389/fpls.2022.1080745

Visible and near-infrared spectroscopy and deep learning application for the qualitative and quantitative investigation of nitrogen status in cotton leaves

Qinlin Xiao[1,2†], Na Wu[3†], Wentan Tang[1,2], Chu Zhang[4], Lei Feng[1,2], Lei Zhou[5], Jianxun Shen[6], Ze Zhang[7], Pan Gao[8] and Yong He[1,2*]

[1]College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China, [2]Key Laboratory of Spectroscopy Sensing, Ministry of Agriculture and Rural Affairs, Hangzhou, China, [3]School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Huzhou, China, [4]School of Information Engineering, Huzhou University, Huzhou, China, [5]College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing, China, [6]Hangzhou Raw Seed Growing Farm, Hangzhou, China, [7]Key Laboratory of Oasis Eco-Agriculture, College of Agriculture, Shihezi University, Shihezi, China, [8]College of Information Science and Technology, Shihezi University, Shihezi, China

Leaf nitrogen concentration (LNC) is a critical indicator of crop nutrient status. In this study, the feasibility of using visible and near-infrared spectroscopy combined with deep learning to estimate LNC in cotton leaves was explored. The samples were collected from cotton's whole growth cycle, and the spectra were from different measurement environments. The random frog (RF), weighted partial least squares regression (WPLS), and saliency map were used for characteristic wavelength selection. Qualitative models (partial least squares discriminant analysis (PLS-DA), support vector machine for classification (SVC), convolutional neural network classification (CNNC) and quantitative models (partial least squares regression (PLSR), support vector machine for regression (SVR), convolutional neural network regression (CNNR)) were established based on the full spectra and characteristic wavelengths. Satisfactory results were obtained by models based on CNN. The classification accuracy of leaves in three different LNC ranges was up to 83.34%, and the root mean square error of prediction (RMSEP) of quantitative prediction models of cotton leaves was as low as 3.36. In addition, the identification of cotton leaves based on the predicted LNC also achieved good results. These results indicated that the nitrogen content of cotton leaves could be effectively detected by deep learning and visible and near-infrared spectroscopy, which has great potential for real-world application.

KEYWORDS

cotton, leaf nitrogen content, spectra, deep learning, visible and near-infrared spectroscopy (Vis-NIR)

1 Introduction

Cotton (*Gossypium* spp), as one of the important economic crops in the world, is widely used in the textile industry because of its excellent natural properties. Nitrogen is an essential plant macronutrient, taking an important part in crop photosynthesis, which provides necessary nutritional support for the growth and development of crops (Ma et al., 2022). Observations have shown that nitrogen fertilization has an important effect on cotton yield. Rational nitrogen fertilization is beneficial to increase cotton yield, while both deficit and excessive nitrogen fertilization have a negative impact on cotton growth and development (Liu et al., 2010; Gospodinova and Panayotova, 2019). Gospodinova and Panayotova (2019) summarized the research on the effects of mineral fertilization on cotton yield and concluded that nitrogen should be applied at different development stages as needed. Optimizing the nitrogen fertilizer application scheme is conducive to improving nitrogen utilization efficiency and cotton yield. Knowing the nutritional status of cotton is the prerequisite to realizing on-demand nitrogen application. Therefore, rapid and accurate evaluation and detection of cotton nitrogen is of great significance for monitoring plant nutrition status, as well as making fertilization decisions.

Leaf nitrogen concentration (LNC), a critical indicator of nitrogen nutrient status, is widely used in crop nutrient status evaluation (Wan et al., 2022). A study conducted by Kergoat et al. (2008) has shown that LNC is an essential factor affecting canopy light utilization efficiency and photosynthetic rate. Generally, LNC is determined by destructive analysis methods, such as the Kjeldahl-digestion method. Although the destructive approaches are objective, they have disadvantages such as being time-consuming, labor-intensive, high cost and strong destructiveness. It is also difficult to meet the actual needs of rapid and real-time detection and diagnosis of LNC in a wide range. In recent years, non-destructive techniques, such as visible and near-infrared (VNIR) spectroscopy (Mishra et al., 2021) and multi-spectral and hyperspectral imaging (Tahmasbian et al., 2021; Guo et al., 2022), have been developed to detect crop nutrition status. Multi-spectral and hyperspectral images usually carry more information than spectra data. However, the acquisition of spectral images generally requires expensive and bulky sensors, the amount of data is enormous, and there is more information redundancy, which requires more storage space and tedious data processing. Multi-spectral and hyperspectral imaging are not economically feasible when many samples need to be examined and evaluated. VNIR allows rapid acquisition of spectral information related to samples' physiological state and internal components at a relatively low cost. In the past few years, VNIR has attracted extensive attention and has been used in qualitative and quantitative research in plants (Zhang et al., 2020a; Xia et al., 2021; Luo et al., 2022).

For the studies aiming at employing VNIR for nitrogen detection, Mishra (Mishra et al., 2021) et al. demonstrated the feasibility of using VNIR to quantitatively predict the nitrogen and potassium concentration in bell pepper leaves. The results showed that VNIR allowed accurate prediction of nitrogen with an RMSEP of 0.44%. Sun et al. (2013) used VNIR to identify the fertilized nitrogen level of lettuce leaves and achieved a high classification accuracy of 100%. Zhang et al. (2022a) explored the performance of using spectra of different ranges to estimate nitrogen content in cotton leaves and obtained a R^2_c = 0.794~0.909 and R^2_P = 0.774 ~ 0.899. Relationships between cotton leaf spectra curves (380-700 nm, 700-1300 nm, and 1300-2500 nm) and nitrogen content contributed to satisfactory predictions for nitrogen content detection. There are indeed many researches on the detection of LNC (Sun et al., 2013; Wang et al., 2018; Gao et al., 2022; Pourdarbani et al., 2022; Tang et al., 2022; Zhang et al., 2022b). Although the studies focusing on the LNC classification achieved good results (Sun et al., 2013; Wang et al., 2018; Pourdarbani et al., 2022), the samples in these studies were classified according to different nitrogen fertilization levels or different nitrogen fertilization days. It should be noted that there is a large difference between the fertilization of nitrogen and its actual uptake for the plant. Therefore, the adaptability of the classification models according to the nitrogen fertilization division is greatly limited by the uncertainty of the actual LNC. What's more, in practice, it is always hard to get accurate fertilization data and estimate the fertilization condition. In addition, the studies focusing on LNC prediction are mainly for a specific cultivar or a specific spectral data collection environment (Rotbart et al., 2013), which may limit the scope of the applicability of the established models.

Deep learning is a method that simulates the human brain for analysis and learning. It forms abstract features to represent the data distribution. Deep learning has the advantages of strong self-learning and feature-extraction ability and great capability of processing spectra data (Xiao et al., 2020). In recent years, deep learning has been applied to conduct various tasks in spectral and image data processing (Steinbrener et al., 2019; Zhou et al., 2019). Convolutional neural network (CNN) is one of the typical deep learning models. CNN has been proven effective in processing spectra data and establishing classification and regression models for various agricultural tasks (Zhang et al., 2020b; Zhang et al., 2020c; Gai et al., 2022).

The objective of the present study was to explore the feasibility of qualitative diagnosis and quantitative detection of LNC based on VNIR combined with deep learning. The goals include (1) exploring the laws of the spectra of leaves with different LNC; (2) classifying nitrogen levels according to the measured LNC; (3) detecting LNC for two cotton cultivars under the condition that the spectra were collected in different measurement environments. The specific content includes (1)

extracting characteristic wavelengths by random frog (RF), weighted partial least squares regression (WPLS), and saliency map for qualitative discrimination and quantitative detection tasks, respectively; (2) building partial least squares discriminant analysis (PLS-DA), support vector machine for classification (SVC), and convolutional neural network classification (CNNC) models based on full spectra and characteristic wavelengths to identify cotton leaves with different LNC qualitatively; (3) developing partial least squares regression (PLSR), support vector machine for regression (SVR), convolutional neural network regression (CNNR) models to quantitatively detect LNC in cotton leaves.

2 Materials and methods

2.1 Sample preparation

Cotton was planted in an experimental field at the Hangzhou Raw Seed Growing Farm (30°22'58.85" N, 119°56'7.80" E), Hangzhou, Zhejiang province, China. Cotton cultivars Lumianyan 24 (LMY24) and Xinluzao 53 (XLZ53) were sampled in this experiment. Thirty-six experimental plots of 4×2 m were used with six nitrogen rates (0, 120, 240, 360, 480, 278 kg/hm^2). Each nitrogen level was set with three replicates. Leaf sampling was conducted during the whole growth stage. Leaves at different leaf positions were selected from the experimental plots. Finally, a total of 1400 leaves were acquired. It is worth mentioning that the spectra of 648 leaves were collected in the laboratory, and the spectra of the remaining leaves were collected in the field. For the samples measured in the laboratory, the leaves were cut, placed in the black bags and stored in a cooler with a temperature of about 4°C. These samples were transported to the laboratory immediately. The time of transit was within one hour.

2.2 Spectra acquisition

Leaf spectra acquisition was conducted by a spectroradiometer (Fieldspec4, Analytical Spectral Devices - ASD, Boulder, CO, USA) system. This spectroradiometer consists of a leaf clip, which provides a light source. During the measurement, the leaves were clamped up for spectra acquisition. Three different positions of each leaf were measured, and five scans were conducted for each measurement. The spectra of five scans were averaged as the spectra of the leaf region, and the average spectra of three leaf regions were taken as the spectra of each leaf. The regions of spectra acquisition for each leaf is shown in Figure 1. The collected spectra cover the visible and near-infrared region (400 ~ 1000 nm) and the short-wave near-infrared region (1000 ~ 2500 nm), and the spectral resolution are 3nm and 8nm, respectively. Considering the noise in the beginning, spectra between 430-2500 nm were used in this study.

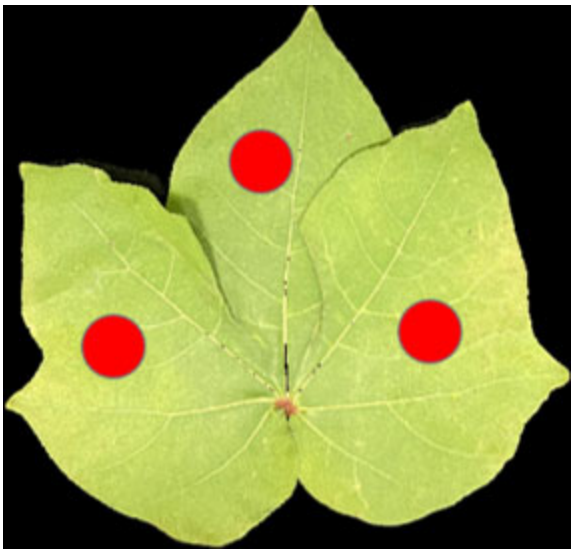


FIGURE 1
The regions of spectra acquisition for cotton leaves.

2.3 Measurements of leave nitrogen concentration

After finishing the spectra collection, the leaves were placed in an oven, dried at 105°C for half an hour, and then cooled to 80°C until the sample weight recorded a constant weight. Then, the dried leaves were ground into a fine powder and sieved through a 40-mesh. A uniform dry leaf sample of fixed mass was taken, and the nitrogen concentration was determined by the Kjeldahl method after acid digestion (Kjeldahl, 1883). According to the measured LNC (mg/g), cotton leaf samples were divided into three categories: low-level LNC, medium-level LNC, and high-level LNC. The detailed statistical information on sample composition is presented in Table 1. Cotton leaves with different LNC levels are shown in Figure 2. It can be seen that the leave with high-level LNC has a deeper green color.

2.4 Data analysis methods

2.4.1 Convolutional neural network

In this study, two self-developed CNN architectures were applied for building classification and regression models, and their structures are shown in Figures 3A, B, respectively. For the classification task, two convolution layers were set, both followed by a max pooling layer and a batch normalization layer. The

TABLE 1 Statistical information of composition of the cotton leaves.

LNC level	total number of samples	Range of LNC (g/kg)	Mean (g/kg)	Standard Deviation	number of samples incal/val/pre set
Low	230	14.99-25.00	21.73	2.37	138/46/46
Medium	601	25.02-34.96	30.09	2.84	360/120/121
High	569	35.02-52.46	40.19	3.60	341/114/114

cal/val/pre set means the calibration set, the validation set and the prediction set, respectively.

number of filters, kernel size, and strides of the two convolution layers were both set as 16, 3, and 1, respectively. The rectified linear unit (ReLU) was used as the activation for computing the outputs of the convolutional layers. The max pooling layer served as down-sampling and dimensionality reduction to form the features of the next layer. Then, a fully connected network with 64 neurons was added, followed by a batch normalization layer. The dropout layer was used to avoid overfitting. The fully connected layer at the end was used for output. For the regression task, two batch normalization layers followed by convolution layers were employed. The number of filters, kernel size, and strides of the two convolution layers were both set as 32, 3, and 1, respectively. Same as the proposed CNN for classification, the rectified linear unit (ReLU) was used as the activation. A batch normalization layer was added before the features were outputting to the fully connected layer. In the end, two fully connected layers with 64 and 16 neurons were used for building non-linear regression models to predict the LNC of different leaves. The fully connected layer at the end was used for output.

For the training phase, the Softmax cross-entropy loss function combined with stochastic gradient descent (SGD) optimizer was applied to train the CNN developed for the classification task. The L1 loss function and the adaptive moment estimation (Adam) optimizer were used for the regression task. The detailed information about SGD optimizer and Adam optimizer could be found on the website https://pytorch.org/docs/stable/optim.html . For both training tasks, the batch size was set as 64, and a scheduled learning rate was used. In the beginning, the learning rate was set to 0.05. The learning rate was reduced ten times after every 200 epochs. According to this rule, the training phase was terminated when the loss was stable.

2.4.2 Conventional models

PLS-DA and SVC models were established to classify cotton leaves with different LNC. PLS-DA is a linear discriminant algorithm developed from PLSR (Yuan et al., 2021). PLS-DA algorithm can effectively extract the variables helpful for classification and realize data recognition. PLS-DA can deal with irreversible matrices and select the number of latent variables so that the model achieves the best balance between underfitting and overfitting (Li et al., 2021). SVC is a pattern recognition algorithm based on a support vector machine (SVM) for classification. It achieves the classification goal by exploring the hyperplane that maximizes the distance between different classes (Xiao et al., 2020). In this study, the radial basis function (RBF) was used as the kernel function. The regularization parameter c and kernel function parameter g were determined through a grid-search procedure. The search range of c and g were both assigned as 2^{-8} to 2^{8}. PLSR and SVR models were used to establish the quantitative analysis model of LNC. Detailed information on PLS and SVR details can be found in our previous article (Xiao et al., 2022). For both qualitative and quantitative analysis models, five-fold cross-validation was adopted.

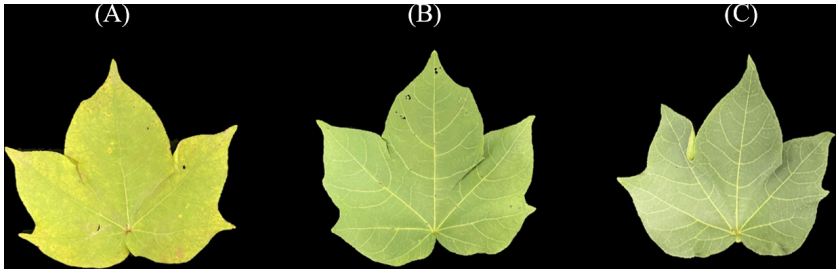


FIGURE 2
Cotton leaves with different LNC levels: **(A)** low-level LNC, **(B)** medium-level LNC, **(C)** high-level LNC.

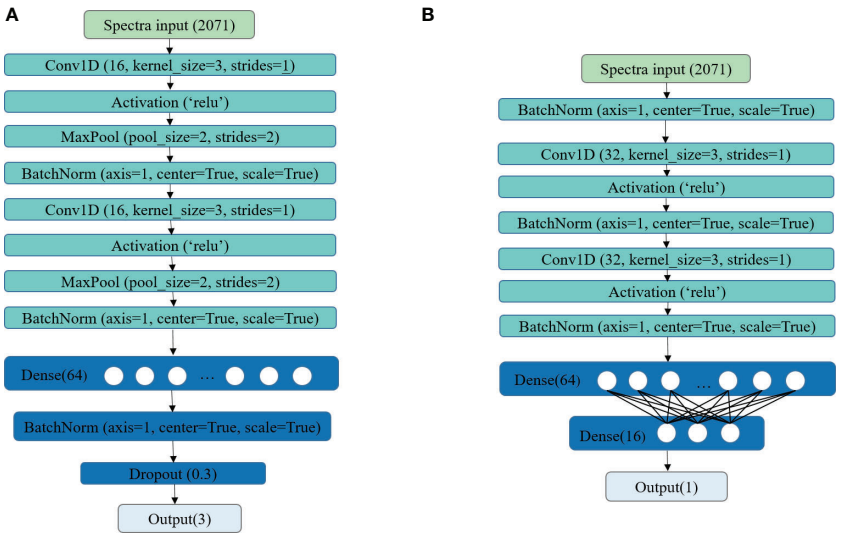


FIGURE 3
The CNN structure for classification model **(A)** and regression model **(B)**.

2.4.3 Wavelengths selection

Hyperspectral data contains massive amounts of information, which also exists information redundancy, collinearity, and noise that are not conducive to data processing. To make effective use of data, characteristic wavelengths extraction is a common strategy. In this study, RF, WPLS and saliency map were used to extract characteristic wavelengths. For the RF algorithm, based on the idea of inversible-jump Markov Monte Carlo, PLS-DA and PLSR are selected as modeling methods for classification and regression, respectively. Models are established by constantly updating the subset of variables according to the defined criteria. The frequency of each variable selected in the modeling subset is calculated after reaching the number of iterations (Yun et al., 2013; Sun et al., 2021). The top 40 wavelengths with the highest frequency were selected as the characteristic wavelengths. When using WPLS for wavelength selection, a PLS regression model is first established, and each variable's regression coefficient was calculated. The wavelengths with the larger absolute value of the regression coefficient at the crest and trough were selected (Mehmood et al., 2012). Saliency map is a popular method for computing the contribution of each variable to the model performance. In this study, for classification tasks, CNNC model was first established and calculated the saliency based on the method proposed in Feng's study (Feng et al., 2021). Similarly, as for regression tasks, CNNR model was first established and saliency map was applied following the way in our previous study (Xiao et al., 2022). The first 40 critical wavelengths with the highest frequency for both tasks were selected as the characteristic wavelengths.

2.4.4 Software and model evaluation

For model establishment, PLS-DA and PLSR were performed in R2019b (The MathWorks, Natick, MA, USA). SVC, and SVR were conducted in the scikit-learn 0.23.1 (Anaconda, Austin, TX, USA) using python 3.1. The CNN models were conducted in MXNet 1.4.0 (MXNetAmazon, Seattle, WA, USA). For feature selection, RF was performed in R2019b (The MathWorks, Natick, MA, USA). WPLS was carried out in the Unscrambler X 10.1 (Camo AS, Oslo, Norway). Saliency map was conducted in MXNet1.4.0 (MXNetAmazon, Seattle, WA, USA).

It is critical to evaluate the model performance with appropriate indicators. Classification accuracy is used for assessing the qualitative analysis models. Classification accuracy is calculated as the ratio of correctly classified samples to the total number of samples. The closer it is to 100%, the better the model's performance. The coefficients of determination (R^2) and root mean square error (RMSE) of calibration, validation, and prediction set were applied to assess the performance of quantitative analysis models. The closer R^2 of the model is to 1, the closer RMSE is to 0, indicating that the model performance is more satisfactory.

3 Results

3.1 Spectra features

The spectra of all the cotton leaves and leaves with different LNC are shown in Figures 4A, B. As shown in Figure 4A, the spectra of all the leaves present a consistent change tendency.

Four peaks (550, 1650, 1820, and 2225 nm) and three valleys (670, 1432, and 1950 nm) were observed. The analysis of the chemical bonds which may be assigned to the peaks and valleys can be found in our previous study (Xiao et al., 2022). Figure 4B presents the spectral curves of cotton leaves with different LNC. It can be seen that the reflectance of leaves with high LNC in the range of 430 ~ 520 nm is slightly higher than that of leaves with low LNC. There is a slight increase in the reflectance between 520 ~ 610 nm, and the reflectance of leaves with low LNC becomes higher. After 700 nm, the reflectance curves increase sharply and form a high reflectivity plateau between 775 and 1300 nm, between which the reflectance of leaves with high LNC is larger. The variation trend of the reflectance with LNC between the range of 1400 ~ 1900 nm and 2000 ~ 2500 nm is the opposite from that in the range of 775-1300 nm. The variation of reflectance in different spectral intervals makes it possible to identify leaves with different LNC content. As the results demonstrated in our previous study (Xiao et al., 2022), the model based on the spectra processed by first derivative (FD) and standard normal variate transformation (SNV) demonstrated great generalization ability. Therefore, the method of FD+SNV was used to preprocess the spectra and the processed spectra were used for subsequent modelling. The transformed curves are shown in Figure 4C.

3.2 Wavelengths selection

Hyperspectral data contains amounts of information such as redundancy, collinearity, background, and other information unrelated to LNC detection. The irrelevant information will significantly increase the burden of data processing, affect the analysis and extraction of effective data, and directly affect the model's performance. Therefore, in this study, RF, WPLS, and saliency map were used to select the characteristic wavelengths.

The optimal wavelengths selected by different methods for regression models and classification models are shown in Tables 2 and 3, respectively. It can be seen that the number and the location of the selected wavelengths on spectral curves varied from different methods. For the wavelengths selected for the classification model, compared with full spectra, the number of variables chosen by RF, WPLS, and saliency map was reduced by 98.07%, 98.02%, and 98.07%, respectively. For the wavelengths selected for the regression model, the number of variables selected by RF, WPLS, and saliency map was reduced by 98.07%, 97.73%, and 98.07%, respectively. Obviously, wavelengths selection significantly reduces data computation and alleviates the model dependence on high-performance computing instruments, which will contribute to the popularization and application of the model.

The position of the selected wavelengths in the spectral curve is displayed in Figure 5. For specific wavelength selection methods, the position of characteristic wavelength selected for classification and regression is largely coincidental. It indicated that the characteristic wavelengths related to nitrogen detection selected by wavelength selection method were consistent even in the tasks with different purposes. The specific number of selected variables might be related to the calculation protocol of wavelength selection method. Although there are differences in

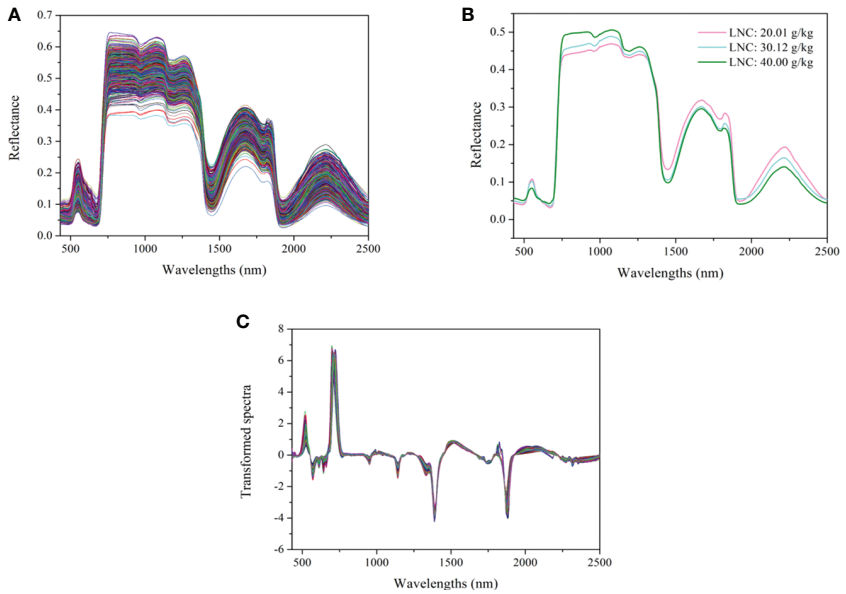


FIGURE 4
Spectra of cotton leaves: **(A)** the spectra of all the cotton leaves, **(B)** the spectra of leaves with different LNC, **(C)** the spectra transformed by FD +SNV.

the number and location of wavelengths selected by RF, WPLS, and saliency map, some bands were chosen as the optimal wavelengths for LNC detection by more than one method, such as the bands around 554 nm, 595 nm, 1179nm, 1490 nm, 1671nm, 1673 nm, 1746 nm, 2046 nm, 2154 nm, 2230 nm, and 2459 nm. These bands are likely to have a strong correlation with nitrogen detection. Among the wavelengths selected by more than one algorithm, the bands in visible range was related with the color of the leaves (Malacara, 2011). The spectral response near 1490 nm was associated with N-H amide with N-R group, which can be connected with protein content (Salzer, 2008). The reflectance around 1673 nm and 1746 nm were associated with C-H methyl (Salzer, 2008). The bands around 2046 nm was due to symmetrical NH stretching and amide II (Salzer, 2008).

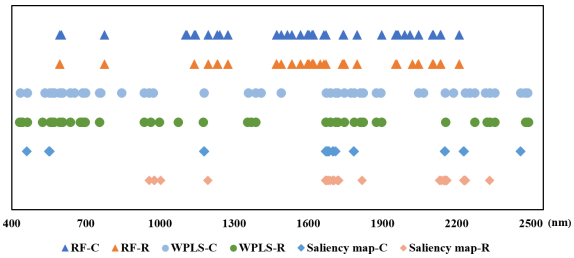


FIGURE 5
The position of the optimal wavelengths selected for classification and regression models (-C means for classification tasks, -R means for regression tasks).

3.3 Classification models

3.3.1 Classification models using full spectra

In this study, PLS-DA, SVC, and CNNC models were built using full spectra. The results are shown in Table 4. All the models obtained decent results, with the accuracy of the prediction set exceeding 82%. Compared with PLS-DA and SVC models, the CNNC model achieved a more satisfactory result. The accuracy of the prediction set was 84.70%, which illustrates the good performance of CNNC model. The confusion matrix for all datasets of the CNNC model is displayed in Figure 6. In three data sets, about 19-28% of leaf samples with low LNC were easily confused with leaves with medium LNC. Leaves with high LNC were easily confused with those with medium LNC, and the proportion of misclassified leaves was 8%-11%. Overall, leaves with high LNC and leaves with low LNC can be accurately separated, as only one or two leaves with low LNC were misclassified as high LNC, and none samples with high LNC were distinguished as low LNC.

3.3.2 Classification models using optimal wavelengths

Table 5 demonstrates the classification results of the PLS-DA, SVC, and CNNC models based on the optimal wavelengths selected by RF, WPLS, and saliency map. The overall performance of the models based on wavelengths selected by RF was superior to the performance of the models constructed on the wavelengths selected by WPLS and saliency map, as the accuracy for the prediction set was slightly higher. For the models based on the optimal wavelengths selected by RF and WPLS, the performance of the CNNC model was superior to that of the PLS-DA and SVC model, accomplishing an accuracy of 84.34% and 83.27% for the prediction set. Regarding the

TABLE 2 The optimal wavelengths selected by RF, WPLS, and saliency map for regression models.

Method	Number	Optimal wavelength (nm)
RF	40	594, 602, 776, 1104, 1109, 1139, 1144, 1196, 1232, 1242, 1274, 1472, 1473, 1490, 1516, 1534, 1568, 1597, 1600, 1604, 1615, 1619, 1620, 1663, 1671, 1741, 1742, 1797, 1897, 1953, 1958, 1963, 1989, 2012, 2047, 2103, 2106, 2134, 2135, 2211
WPLS	47	436, 464, 535, 555, 565, 573, 594, 607, 638, 655, 688, 699, 755, 760, 845, 936, 957, 973, 1179, 1357, 1387, 1409, 1490, 1673, 1690, 1711, 1720, 1746, 1773, 1785, 1810, 1822, 1876, 1894, 2046, 2066, 2154, 2188, 2236, 2254, 2273, 2317, 2335, 2355, 2459, 2478, 2488
Saliency map	40	461, 551, 552, 553, 554, 1177, 1178, 1179, 1670, 1671, 1672, 1673, 1674, 1675, 1676, 1677, 1678, 1679, 1680, 1681, 1682, 1683, 1684, 1699, 1700, 1701, 1702, 1710, 1783, 1784, 1785, 2151, 2152, 2153, 2227, 2228, 2229, 2230, 2458, 2459

TABLE 3 The optimal wavelengths selected by RF, WPLS, and saliency map for classification models.

Method	Number	Optimal wavelength (nm)
RF	40	594, 595, 776, 1139, 1195, 1232, 1233, 1275, 1471, 1472, 1490, 1533, 1534, 1568, 1597, 1600, 1604, 1615, 1616, 1619, 1620, 1621, 1648, 1663, 1671, 1738, 1741, 1745, 1746, 1798, 1953, 1958, 2021, 2044, 2047, 2103, 2106, 2134, 2135, 2211
WPLS	41	432, 437, 444, 464, 525, 556, 564, 571, 595, 607, 638, 678, 688, 698, 755, 936, 963, 997, 1074, 1176, 1355, 1370, 1387, 1672, 1690, 1711, 1720, 1746, 1785, 1810, 1822, 1876, 1897, 2155, 2273, 2324, 2335, 2355, 2479, 2487, 2490
Saliency map	40	957, 976, 977, 978, 1002, 1193, 1670, 1671, 1672, 1674, 1680, 1682, 1690, 1700, 1702, 1703, 1704, 1718, 1720, 1722, 1723, 1818, 2131, 2133, 2137, 2147, 2149, 2152, 2155, 2156, 2157, 2160, 2228, 2230, 2231, 2232, 2234, 2235, 2236, 2333

10.3389/fpls.2022.1080745

TABLE 4 The results of the classification models based on full spectra.

Model	Accuracy		
	Calibration set	Validation set	Prediction set
PLS-DA	88.56%	77.14%	82.56%
SVC	85.10%	79.29%	83.99%
CNN	86.17%	82.86%	84.70%

models based on the wavelengths selected by saliency map, SVC performed better than the PLS-DA and CNNC model, with the accuracy of the prediction set reaching 79.36%. Although the CNNC model based on full spectra achieved the best classification accuracy of 84.70% for the prediction set, the CNNC model based on the optimal wavelengths chosen by RF obtained quite similar results. Considering the number of variables used in modeling, the results of CNNC models constructed on the optimal wavelengths selected by RF were reasonably acceptable, which realized comparable performance with the model based on full spectra with less computation.

3.4 Regression models

3.4.1 Regression models using full spectra

Figure 7 shows the results of different regression models using full spectra for the nitrogen detection of cotton leaves. All the models obtained satisfactory performance, with R^2_c (coefficients of determination of calibration set), R^2_v (coefficients of determination of validation set), and R^2_p (coefficients of determination of prediction set) all exceeding 0.75. Compared with PLSR and SVR models, the CNNR model performed slightly better, achieving the smallest RMSE for the prediction set. These results indicated that VNIR combined with the CNNR model was conducive to effectively characterizing the LNC of cotton leaves.

3.4.2 Regression models using optimal wavelengths

Different regression models were constructed based on the optimal wavelengths for LNC estimation. The results are shown in Table 6. It can be found that the performance of the model established using the optimal wavelengths selected by different methods showed a difference. The models built on the optimal wavelengths selected by WPLS performed slightly better than the models based on the optimal wavelengths selected by RF and saliency map, with higher R^2_p and lower RMSEP. In addition, it can be found that the overall performance of CNNR models was better than that of PLS and SVR models. All the CNNR model achieved good results, with R^2_c, R^2_v, and R^2_p were over 0.779, 0.724, and 0.711, indicating the robustness of the CNNR model based on optimal wavelengths. Specifically, the CNNR model based on the wavelengths chosen by WPLS obtained the best result. The R^2_p and RMSEP were 0.766 and 3.389, respectively. Besides, a comparison was made between the models based on full spectra and those using the selected optimal wavelengths. Overall, the models established on the chosen variables performed less well than those based on full spectra. The performance of the CNNR model based on optimal wavelengths selected by WPLS was quite close to that based on full spectra. To some extent, the reduced computation compensates for the slight performance deficit, indicating that the CNNR model equipped with optimal wavelength selection methods is effective for cotton LNC estimation.

3.4.3 Identification of leaf nitrogen status based on the predicted LNC

The predicted LNC values of all samples were calculated by regression model and then classified according to the rules mentioned in section 2.3. Then, the identification accuracy was calculated by comparing the categories corresponding to the predicted and actual values to evaluate the model's effectiveness. The results are shown in Table 7. It can be seen that the identification of cotton samples based on the predicted

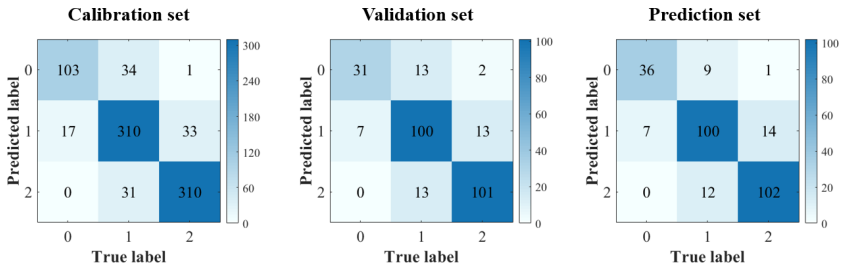


FIGURE 6
Confusion matrix of CNN model using full spectra. (Notes: Number 0, 1, 2 means leaf samples with low-level LNC, medium-level LNC and high-level LNC, respectively.).

10.3389/fpls.2022.1080745

TABLE 5 The results of the classification models based on optimal wavelengths.

Data type	Model	Accuracy		
		Calibration set	Validation set	Prediction set
RF	PLS-DA	80.93%	75.71%	79.00%
	SVC	88.08%	80.71%	83.99%
	CNN	86.53%	79.64%	84.34%
WPLS	PLS-DA	78.90%	77.86%	77.22%
	SVC	84.03%	80.36%	82.92%
	CNN	82.84%	79.29%	83.27%
Saliency map	PLS-DA	74.37%	71.89%	74.29%
	SVC	84.51%	78.21%	79.36%
	CNN	83.08%	77.50%	77.58%

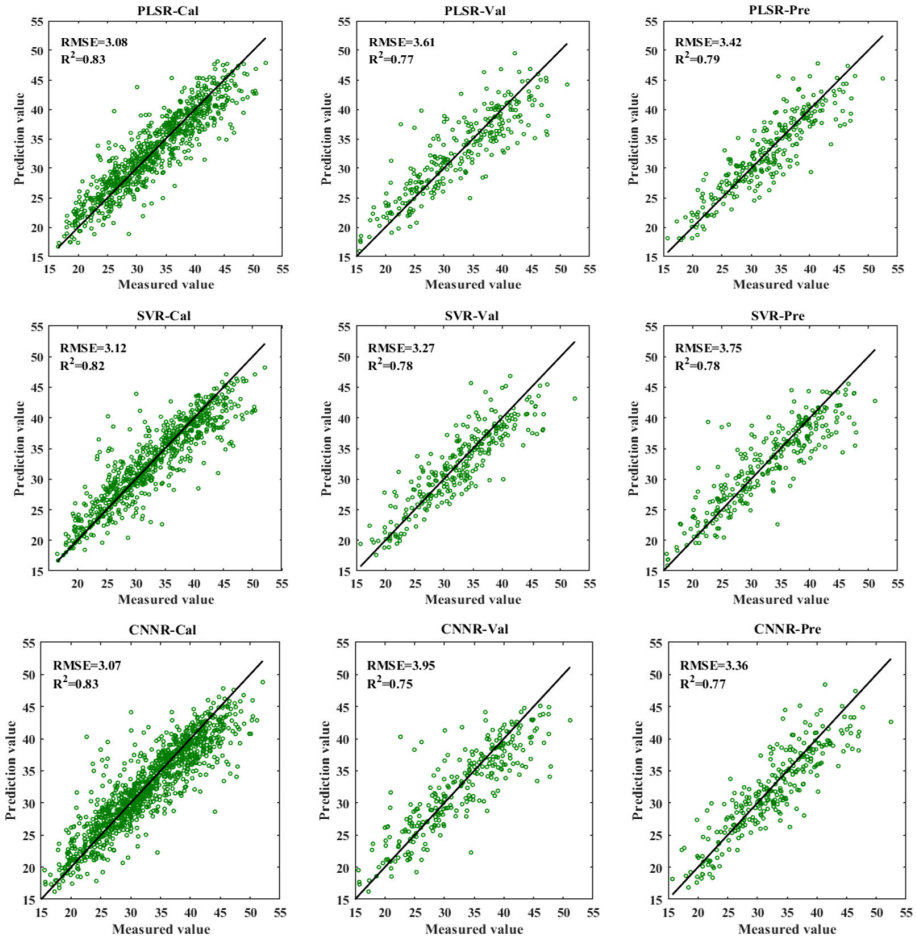


FIGURE 7
The results of PLSR, SVR, CNNR models based on full spectra for LNC detection.

TABLE 6 The results of the regression models based on optimal wavelengths.

Data Type	Model	Calibration set		Validation set		Prediction set	
		R^2_C	RMSEC	R^2_V	RMSEV	R^2_P	RMSEP
RF	PLSR	0.76	3.67	0.75	3.95	0.74	3.59
	SVR	0.80	3.30	0.73	3.62	0.73	4.12
	CNNR	0.82	3.19	0.74	4.02	0.75	3.53
WPLS	PLSR	0.78	3.52	0.76	3.93	0.76	3.42
	SVR	0.80	3.37	0.76	3.42	0.77	3.83
	CNNR	0.79	3.38	0.76	3.91	0.77	3.39
Saliency map	PLSR	0.66	4.36	0.60	2.02	0.67	4.03
	SVR	0.81	3.27	0.71	3.77	0.72	4.17
	CNNR	0.78	3.50	0.72	4.17	0.71	3.76

value could achieve good results, similar to that of the results based on classification models, which in turn reflects the effectiveness of the regression model.

4 Discussion

Visible and near-infrared spectral techniques combined with deep learning can be used for nitrogen-level estimation. Some previous studies addressing the nitrogen-level classification of plant leaves were discussed and compared with our results with other spectral imaging works. It must be noted a rough comparison is not rigorous as the papers relate to different plants, techniques, and datasets. Pourdarbani et al. (2022) investigated the feasibility of using hyperspectral imaging to detect excess nitrogen content in tomato plants. Artificial neural networks and the particle swarm optimization algorithm were proposed and achieved a satisfactory classification accuracy of 92.6% for leaves at different nitrogen levels. The leaves in this work (Pourdarbani et al., 2022) were classified according to different days of nitrogen application. Sun et al. (2013) used VNIR to identify the nitrogen level of lettuce leaves. Adaptive boosting was applied with K nearest neighbor and SVM, which could achieve a high classification accuracy of 100%. The samples in this study were divided according to the fertilized nitrogen level. Wang et al. (2018) employed hyperspectral imaging to discriminate nitrogen fertilizer levels of the tea plant. The leaves from three nitrogen fertilizer levels were sampled, and up to 100% accuracy was achieved by the SVM model based on spectral data and textural data. The excellent performance might benefit from the texture information provided by the image. Although the methods mentioned above achieved good results, the samples in these studies were classified according to different nitrogen fertilization levels or different nitrogen fertilization days (Sun et al., 2013; Wang et al., 2018; Pourdarbani et al., 2022). There is a large difference between the fertilization of nitrogen and its actual uptake for the plant. Therefore, the adaptability of the classification models according to the nitrogen fertilization

TABLE 7 The identification results of cotton leaves based on the predicted LNC.

Data type	Model	Accuracy		
		Calibration set	Validation set	Prediction set
Full spectra	PLSR	83.81%	80.71%	85.00%
	SVR	85.71%	83.21%	79.29%
	CNNR	84.17%	78.21%	83.21%
RF	PLSR	81.43%	76.79%	83.21%
	SVR	85.00%	81.79%	80.00%
	CNNR	83.57%	77.50%	81.43%
WPLS	PLSR	81.07%	78.57%	81.79%
	SVR	82.50%	81.79%	78.21%
	CNNR	85.12%	78.93%	81.07%
Saliency map	PLSR	75.60%	64.64%	73.93%
	SVR	84.40%	78.21%	78.21%
	CNNR	81.90%	76.07%	77.14%

division is greatly limited by the uncertainty of the actual LNC. Thus, in this study, the chemical analysis for nitrogen measurement was conducted, and the leaves were divided into three categories according to the true LNC value of three ranges. Among similar studies which also measured actual LNC, Nativ et al. (Rotbart et al., 2013) used VNIR to estimate the nitrogen concentration in olive leaves. The leaves were divided into three groups according to the measured nitrogen content, and an overall accuracy of 83% was obtained. Du et al. (2016) explored the feasibility of using hyperspectral LiDAR to detect nitrogen content in rice leaves. The accuracy of 83% was obtained when 32 wavelengths were considered. The results in above studies (Rotbart et al., 2013) (Du et al., 2016) are slightly lower than the accuracy of 84.342% achieved by RF-CNN model in this study. It can be observed that although a perfect classification is not achieved, the method used in this study has a relatively higher accuracy of 84.342% in the best case. The performance is quite close to and even higher than the result obtained by other existing methods, which demonstrates that it is feasible to classify cotton leaves with different LNC by VNIR and deep learning algorithm.

Regarding the regression task, in a similar study on the LNC prediction of cotton leaves, Zhang et al. (2022a) explored the potential of using spectra of different ranges to estimate nitrogen content in cotton leaves, and obtained a R^2_c = 0.794~ 0.909 and R^2_P = 0.774 ~ 0.899. The prediction results of the best model are better than those in this research. The possible reason was that the samples used in this study (Zhang et al., 2022) were acquired at the flower and boll stage of cotton, which only covered two growing stages. The leaves used in our study cover the whole growing stage. Different thicknesses and textures of leaf samples would also cause spectral differences, which may affect the accuracy of nitrogen detection. Besides, the spectra in this paper were collected under two environmental conditions, covering the laboratory environment and the field environment. The difference in the measurement environment would also lead to the difference in the spectra. When the measurement was conducted in the field, there were more interference factors, which was also the reason for the relatively less satisfactory results. However, in practice, due to the diversity of application scenarios and the need for nutrient monitoring over the whole growth cycle of plants, it is critical to develop the models presented in this paper to enhance their applicability.

Besides, deep learning with VNIR performed well in estimating LNC in plant leaves. Table 6 and Figure 7 show that CNNR outperformed PLSR and SVR models, achieving a relatively lower RMSE for the prediction set. The study demonstrated that the CNN model established for regression tasks could achieve good results, which previous studies have confirmed. Weng et al. (2022) combined CNNR and visible and near-infrared reflectance spectroscopy to determine the behenic acid in edible vegetable oils, with R^2_P = 0.8843 and RMSEP = 0.1182, outperforming PLSR and SVR model. Wu et al. (2022) applied CNNR and Raman spectroscopy to identify the amount of olive oil in a corn-olive oil blend, with R^2_P = 0.9908 and RMSEP = 0.7183. In addition, one-dimension deep learning regression models based on spectral data are performed well in soluble solid content estimation in cherry tomato (Xiang et al., 2022) and oil content prediction of single maize kernel (Zhang et al., 2022c). Hence, the spectral analysis model developed by CNN can be expected to provide a simple, rapid, and accurate analysis of LNC in cotton leaves.

5 Conclusion

In this study, visible and near-infrared spectroscopy combined with deep learning was used to detect LNC in cotton leaves qualitatively and quantitatively. RF, WPLS, and saliency map were used to extract characteristic wavelengths, classification models (PLS-DA, SVC, CNNC) and regression models (PLSR, SVR, CNNR) were established based on full spectra and characteristic wavelengths, respectively. Overall, the models based on CNN architecture performed better than other models for both classification and regression tasks. For the classification task, CNNC model based on full spectra performed best, with the classification accuracy reaching 84.70%. For the regression task, the performance of CNNR model developed on full spectra was superior, achieving an R^2_P of 0.77 and an RMSEP of 3.36. The good performance of visible and near-infrared spectroscopy assisted by deep learning demonstrated its effectiveness for nitrogen content prediction of cotton leaves. This approach is helpful for farmers to accurately identify the nutritional status of cotton plants in the field and make reasonable fertilization decisions in time.

Data availability statement

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

Author contributions

YH, LF, and NW, funding acquisition, conceptualization, and supervision. QX, WT, and NW, data curation and validation. CZ, LF, LZ, ZZ, and PG, formal analysis, and writing, project administration review and editing. CZ and QX, investigation, methodology, and software. JS, resources. QX, NW, and CZ, visualization and writing, original draft. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by XPCC Science and Technology Projects of Key Areas (2020AB005) and National Natural Science

Foundation of China (61965014), and the Shenzhen Science and Technology Projects (CJGJZD20210408092401004).

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

Du, L., Gong, W., Shi, S., Yang, P., Sun, J., Zhu, B., et al. (2016). Estimation of rice leaf nitrogen contents based on hyperspectral lidar. *Int. J. Appl. Earth Observation Geoinformation* 44, 136–143. doi: 10.1016/j.jag.2015.08.008

Feng, L., Wu, B., He, Y., and Zhang, C. (2021). Hyperspectral imaging combined with deep transfer learning for rice disease detection. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.693521

Gai, Z., Sun, L., Bai, H., Li, X., Wang, J., and Bai, S. (2022). Convolutional neural network for apple bruise detection based on hyperspectral. *Spectrochimica Acta Part a-Molecular Biomolecular Spectrosc.* 279, 121432. doi: 10.1016/j.saa.2022.121432

Gao, Z., Luo, N., Yang, B., and Zhu, Y. (2022). Estimating leaf nitrogen content in wheat using multimodal features extracted from canopy spectra. *Agronomy-Basel* 12 (8), 1915. doi: 10.3390/agronomy12081915

Gospodinova, G., and Panayotova, G. (2019). Strategies for nitrogen fertilization of cotton (gossypium hirsutum l.). a review. *Bulgarian J. Agric. Sci.* 25, 59–67.

Guo, Y., Chen, S., Li, X., Cunha, M., Jayavelu, S., Cammarano, D., et al. (2022). Machine learning-based approaches for predicting spad values of maize using multi-spectral images. *Remote Sens.* 14 (6), 1337. doi: 10.3390/rs14061337

Kjeldahl, J. (1883). A new method for the estimation of nitrogen in organic compounds. *Analytical Chem.* 22, 366–382. doi: 10.1007/BF01338151

Kergoat, L., Lafont, S., Arneth, A., Le Dantec, V., and Saugier, B. (2008). Nitrogen controls plant canopy light-use efficiency in temperate and boreal ecosystems. *J. Geophys. Res. Biogeosci.* 113 (G4), G04017. doi: 10.1029/2007JG000676

Li, X., Jiang, H., Jiang, X., and Shi, M. (2021). Identification of geographical origin of chinese chestnuts using hyperspectral imaging with 1d-cnn algorithm. *Agriculture-Basel* 11 (12), 1274. doi: 10.3390/agriculture11121274

Liu, S., Zhang, S., and Zhang, L. (2010). Above-ground dry matter accumulation of cotton genetics at different nitrogen applications. *Cotton Sci.* 22 (1), 77–82.

Luo, W., Tian, P., Fan, G., Dong, W., Zhang, H., Liu, X. J. I. P., et al. (2022). Non-destructive determination of four tea polyphenols in fresh tea using visible and near-infrared spectroscopy. *Infrared Phys. Technol.* 123, 104037. doi: 10.1016/j.infrared.2022.104037

Ma, L., Chen, X., Zhang, Q., Lin, J., Yin, C., Ma, Y., et al. (2022). Estimation of nitrogen content based on the hyperspectral vegetation indexes of interannual and multi-temporal in cotton. *Agronomy-Basel* 12 (6), 1319. doi: 10.3390/agronomy12061319

Malacara, D. (2011). *Color vision and colorimetry: Theory and applications. 2nd ed* (Bellingham, WA, USA: SPIE Press).

Mehmood, T., Liland, K. H., Snipen, L., and Saebo, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometrics Intelligent Lab. Syst.* 118, 62–69. doi: 10.1016/j.chemolab.2012.07.010

Mishra, P., Herrmann, I., and Angileri, M. J. T. (2021). Improved prediction of potassium and nitrogen in dried bell pepper leaves with visible and near-infrared spectroscopy utilising wavelength selection techniques. *Talanta* 225, 121971. doi: 10.1016/j.talanta.2020.121971

Pourdarbani, R., Sabzi, S., Rohban, M. H., Garcia-Mateos, G., Molina-Martinez, J. M., Paliwal, J., et al. (2022). Metaheuristic algorithms in visible and near infrared spectra to detect excess nitrogen content in tomato plants. *J. near Infrared Spectrosc.* 30, 197–207. doi: 10.1177/09670335221098527

Rotbart, N., Schmilovitch, Z., Cohen, Y., Alchanatis, V., Erel, R., Ignat, T., et al. (2013). Estimating olive leaf nitrogen concentration using visible and near-infrared spectral reflectance. *Biosyst. Eng.* 114, 426–434. doi: 10.1016/j.biosystemseng.2012.09.005

Salzer, R. (2008). *Practical guide to interpretive near-infrared spectroscopy* Vol. 47 (Boca Raton, FL, USA: CRC Press).

Steinbrener, J., Posch, K., and Leitner, R. (2019). Hyperspectral fruit and vegetable classification using convolutional neural networks. *Comput. Electron. Agric.* 162, 364–372. doi: 10.1016/j.compag.2019.04.019

Sun, H., Feng, M., Xiao, L., Yang, W., Ding, G., Wang, C., et al. (2021). Potential of multivariate statistical technique based on the effective spectra bands to estimate the plant water content of wheat under different irrigation regimes. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.631573

Sun, J., Jin, X., Mao, H., Wu, X., Tang, K., and Zhang, X. (2013). Identification of lettuce leaf nitrogen level based on adaboost and hyperspectrum. *Spectrosc. Spectral Anal.* 33, 3372–3376. doi: 10.3964/j.issn.1000-0593(2013)12-3372-05

Tahmasbian, I., Morgan, N. K., Bai, S. H., Dunlop, M. W., and Moss, A. F. (2021). Comparison of hyperspectral imaging and near-infrared spectroscopy to determine nitrogen and carbon concentrations in wheat. *Remote Sens.* 13 (6), 1128. doi: 10.3390/rs13061128

Tang, R., Luo, X., Li, C., and Zhong, S. (2022). A study on nitrogen concentration detection model of rubber leaf based on spatial-spectral information with nir hyperspectral data. *Infrared Phys. Technol.* 122, 104094. doi: 10.1016/j.infrared.2022.104094

Wang, Y., Hu, X., Hou, Z., Ning, J., and Zhang, Z. (2018). Discrimination of nitrogen fertilizer levels of tea plant (camellia sinensis) based on hyperspectral imaging. *J. Sci. Food Agric.* 98, 4659–4664. doi: 10.1002/jsfa.8996

Wan, L., Zhou, W., He, Y., Wanger, T. C., and Cen, H. (2022). Combining transfer learning and hyperspectral reflectance analysis to assess leaf nitrogen concentration across different plant species datasets. *Remote Sens. Environ.* 269, 112826. doi: 10.1016/j.rse.2021.112826

Weng, S., Chu, Z., Wang, M., Han, K., Zhu, G., Liu, C., et al. (2022). Reflectance spectroscopy with operator difference for determination of behenic acid in edible vegetable oils by using convolutional neural network and polynomial correction. *Food Chem.* 367, 130668. doi: 10.1016/j.foodchem.2021.130668

Wu, X., Gao, S., Niu, Y., Zhao, Z., Ma, R., Xu, B., et al. (2022). Quantitative analysis of blended corn-olive oil based on raman spectroscopy and one-dimensional convolutional neural network. *Food Chem.* 385, 132655. doi: 10.1016/j.foodchem.2022.132655

Xiang, Y., Chen, Q., Su, Z., Zhang, L., Chen, Z., Zhou, G., et al. (2022). Deep learning and hyperspectral images based tomato soluble solids content and firmness estimation. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.860656

Xiao, Q., Bai, X., Gao, P., and He, Y. (2020). Application of convolutional neural network-based feature extraction and data fusion for geographical origin identification of radix astragali by visible/short-wave near-infrared and near infrared hyperspectral imaging. *Sensors* 20, 4940. doi: 10.3390/s20174940

Xiao, Q., Tang, W., Zhang, C., Zhou, L., Feng, L., Shen, J., et al. (2022). Spectral preprocessing combined with deep transfer learning to evaluate chlorophyll content in cotton leaves. *Plant Phenomics* 2022, 9813841. doi: 10.34133/2022/9813841

Xia, J., Zhang, W., Zhang, W., Yang, Y., Hu, G., Ge, D., et al. (2021). A cloud computing-based approach using the visible near-infrared spectrum to classify greenhouse tomato plants under water stress. *Comput. Electron. Agric.* 181, 105966. doi: 10.1016/j.compag.2020.105966

Yuan, R., Liu, G., He, J., Wan, G., Fan, N., Li, Y., et al. (2021). Classification of lingwu long jujube internal bruise over time based on visible near-infrared hyperspectral imaging combined with partial least squares-discriminant analysis. *Comput. Electron. Agric.* 182, 106043. doi: 10.1016/j.compag.2021.106043

Yun, Y., Li, H., Wood, L. R. E., Fan, W., Wang, J., Cao, D., et al. (2013). An efficient method of wavelength interval selection based on random frog for multivariate spectral calibration. *Spectrochimica Acta Part a-Molecular Biomolecular Spectrosc.* 111, 31–36. doi: 10.1016/j.saa.2013.03.083

Zhang, D., Chen, G., Yin, X., Hu, R., Gu, C., Pan, Z., et al. (2020c). Integrating spectral and image data to detect fusarium head blight of wheat. *Comput. Electron. Agric.* 175, 105588. doi: 10.1016/j.compag.2020.105588

Zhang, J., Cheng, T., Shi, L., Wang, W., Niu, Z., Guo, W., et al. (2022b). Combining spectral and texture features of uav hyperspectral images for leaf nitrogen content monitoring in winter wheat. *Int. J. Remote Sens.* 43, 2335–2356. doi: 10.1080/01431161.2021.2019847

Zhang, J., Liu, Y., He, Y., Hu, G., and Bai, N. (2020a). Characterization of deep green infection in tobacco leaves using a hand-held digital light projection based near-infrared spectrometer and an extreme learning machine algorithm. *Analytical Lett.* 53 (14), 2266–2277. doi: 10.1080/00032719.2020.1738452

Zhang, J., Yang, Y., Feng, X., Xu, H., Chen, J., and He, Y. (2020b). Identification of bacterial blight resistant rice seeds using terahertz imaging and hyperspectral imaging combined with convolutional neural network. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00821

Zhang, L., Wang, Y., Wei, Y., and An, D. (2022c). Near-infrared hyperspectral imaging technology combined with deep convolutional generative adversarial network to predict oil content of single maize kernel. *Food Chem.* 370, 131047. doi: 10.1016/j.foodchem.2021.131047

Zhang, Q., Ma, L., Chen, X., Lin, J., Yin, C., Yao, Q., et al. (2022). Estimation of nitrogen in cotton leaves using different hyperspectral region data. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 50 (1), 12595. doi: 10.15835/nbha50112595

Zhou, L., Zhang, C., Liu, F., Qiu, Z., and He, Y. (2019). Application of deep learning in food: A review. *Compr. Rev. Food Sci. Food Saf.* 18, 1793–1811. doi: 10.1111/1541-4337.12492

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 13 January 2023
DOI 10.3389/fpls.2022.1096619

OPEN ACCESS

EDITED BY
Shizhuang Weng,
Anhui University, China

REVIEWED BY
Alireza Sanaeifar,
Zhejiang University, China
Yongqian Ding,
Nanjing Agricultural University, China

*CORRESPONDENCE
Xiaodan Ma
mxd@cau.edu.cn

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

RECEIVED 12 November 2022
ACCEPTED 28 December 2022
PUBLISHED 13 January 2023

CITATION
He H, Ma X, Guan H, Wang F and Shen P (2023) Recognition of soybean pods and yield prediction based on improved deep learning model.
Front. Plant Sci. 13:1096619.
doi: 10.3389/fpls.2022.1096619

Recognition of soybean pods and yield prediction based on improved deep learning model

Haotian He, Xiaodan Ma*, Haiou Guan, Feiyi Wang and Panpan Shen

College of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, Da Qing, China

As a leaf homologous organ, soybean pods are an essential factor in determining yield and quality of the grain. In this study, a recognition method of soybean pods and estimation of pods weight per plant were proposed based on improved YOLOv5 model. First, the YOLOv5 model was improved by using the coordinate attention (CA) module and the regression loss function of boundary box to detect and accurately count the pod targets on the living plants. Then, the prediction model was established to reliably estimate the yield of the whole soybean plant based on back propagation (BP) neural network with the topological structure of 5-120-1. Finally, compared with the traditional YOLOv5 model, the calculation and parameters of the proposed model were reduced by 17% and 7.6%, respectively. The results showed that the average precision (AP) value of the improved YOLOv5 model reached 91.7% with detection rate of 24.39 frames per millisecond. The mean square error (MSE) of the estimation for single pod weight was 0.00865, and the average coefficients of determination R^2 between predicted and actual weight of a single pod was 0.945. The mean relative error (MRE) of the total weight estimation for all potted soybean plant was 0.122. The proposed method can provide technical support for not only the research and development of the pod's real-time detection system, but also the intelligent breeding and yield estimation.

KEYWORDS

soybean pods, deep learning model, pod recognition, phenotypic calculation, yield prediction model

1 Introduction

Soybean is not only one of the five major crops in the world, but also an essential high protein grain and oil crop (Yu et al., 2022). As a leaf homologous organ (Lu et al., 2022), soybean pods are an important factor in determining grain yield and quality. Therefore, it is necessary to detect the pod's quality of soybean plants in different growth stages and analyze the phenotypic characters of different varieties of pods. At the same time, it is also one of the important methods for identification and screening of soybean varieties (Zhao C. et al., 2021).

In recent years, the target detection technology based on deep learning had been applied to detect the traits of crop ecology and morphology, and had achieved good results in

measuring and analyzing crop fruit, disease, stem, growth and yield estimation (Jia et al., 2021; Zhu et al., 2021; Lee and Shin, 2021; Sharma et al., 2022; Fu et al., 2022a). Kong et al. (2021) proposed a fruit target detection and positioning method based on Darknet depth framework YOLOv4, which realized the accurate positioning and recognition of the different kinds of fruits in the picture. Gao et al. (2022) used the Yolov4-micro network integrating with the channel spatial reliability discriminant correlation filter (CSR-DCF) to detect and count the apple fruit. Mirhaji et al. (2021) selected the yolov4 model to recognize the oranges on the fruit trees in the image, realizing the estimation of the orange yield in the orchard. Chen et al. (2022) used a new revolution bottleneck module and added an SE module on the basis of the original YOLOv5 network model to identify plant diseases under the natural conditions, accurately. Anam et al. (2021) trained the convolutional neural network (CNN) to segment the affected area of rice leaf disease based on the local threshold segmentation. Mathew and Mahesh (2022) used YOLOv5 network model to classify the acquired pictures of pepper leaf diseases, and realized the effective detection of bacterial spot disease on pepper leaves in the farm. Liu H. et al. (2020) built a CNN model with multi-scale hierarchical features based on the deep learning framework Tensor Flow, and realized the accurate identification of corn seedling stems. Ma et al. (2021) introduced the image augmentation technology in the original Mask R-CNN network layer to expand the image samples, and proposed an effective segmentation method of rice stem impurities based on improved Mask R-CNN. Fu et al. (2022b) used the YOLOv4 network model to realize the rapid real-time detection of banana bunches and stems in the natural environment. Zhou et al. (2021) used UAV images and convolution neural network to estimate the yield of soybean breeding varieties under drought stress. Yang et al. (2022) proposed a pod length and width calculation method based on Mask R-CNN network structure, which realized rapid segmentation of pods from pictures and effective calculation of the pod's shape traits. Uzal et al. (2018) estimated the number of seeds per pod in plant breeding based on customized feature extraction (FE), support vector machine (SVM) and convolutional neural network (CNN). Yan et al. (2020) used five kinds of deep learning network models to identify one pod, two pods, three pods and four pods of mature soybean in the picture, accurately. Guo et al. (2021) improved the yolov4 target detection algorithm by integrating the K-means clustering algorithm and the attention mechanism module, and realized the detection of the number of pods per plant and the number of seeds in pods. Li et al. (2021) proposed a set of SPM-IS soybean phenotype measurement framework composed of the characteristic pyramid network, principal component analysis algorithm and instance segmentation network, which realized the effective measurement of the pod length, pod width, node length, main stem length, grain length, grain width, number of pods, number of nodes and number of nodes. Ning et al. (2021) proposed a phenotypic information extraction method for the soybean plant based on IM-SSD+ACO algorithm, which realized the extraction of soybean phenotypic traits including the number of pods, plant height, number of branches, main stem and plant type, effectively. Lu et al. (2022) proposed a method based on the YOLOv3 algorithm to predict soybean yield according to the number of pods and leaves. Zhang et al. (2021) constructed a soybean yield prediction model based on skew parameters using the color of soybean canopy leaves at different growth stages as input values.

At present, the recognition and counting of soybean pods as well as the detection of morphological and physiological phenotypes are not systematic, and there is a lack of methods for the recognition of soybean pods under natural growth and the estimation of pods weight per plant. In order to overcome the shortcomings of traditional artificial pods counting and yield estimation, such as time-consuming, error-prone and subjective factors, in this paper, a pod counting method based on improved YOLOv5 model was proposed. And then the pod phenotypic traits of pods were calculated including length, width, area, chord length and convex arc length. On this basis, the prediction model was established to reliably estimate the yield of the whole soybean plant based on Back Propagation (BP) neural network with the topological structure of 5-120-1.

2 Data acquisition and image preprocessing

2.1 Experimental materials

The cultivation of soybean plants and the acquisition of pods information were carried out in Heilongjiang Bayi Agricultural University of China. Soybean cultivation experiment was based on the agronomic background of exploring the changes in the ecological and morphological growth process of various organs of soybean under drought stress. The objective of this study was to establish a yield estimation method based on the number of soybean pods. The experiment was conducive to accurate control of soil, fertilizer, water and other environmental factors, reflecting the differences of plant characteristics under different growth stages. The cultivation experiment was carried out under the outdoor condition of 20~34° C, and the soybean varieties Suinong 26 and Heihe 49 were selected. During soybean planting, the medium sized soil block was paved on the bottom of the basin made of Polyvinyl chloride (PVC) with a diameter of 0.3m and a height of 0.18m, and then the screened non saline alkali fine soil was loaded until the basin weight was 5kg. After the compound fertilizer was paved evenly, the fine soil shall be put into the basin to 8.32kg. A total of 60 pots of two varieties were planted, with 6 holes in each pot and 2 soybeans in each hole. At the first trifoliolate stage (V1), 24 pots of single soybean and 36 pots of multi soybean were reserved.

2.2 Construction of image acquisition system

A soybean plants image acquisition system based on digital camera was constructed to dynamically acquire the digital image data of soybean plants in different growth stages. The growth stages of soybean studied included the beginning seed stage (R5), the full seed stage (R6), the beginning maturity stage (R7) and the full maturity stage (R8). The soybean plant image acquisition system device was shown in Figure 1. The acquisition system was composed of a digital camera, a lifting tripod support and a lifting platform. The acquisition platform used Canon

10.3389/fpls.2022.1096619

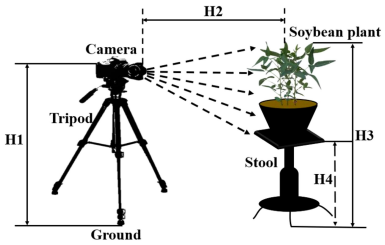


FIGURE 1
Schematic diagram of acquisition device for soybean plants images.

700D camera and Canon IS USM zoom lens, and the resolution of the captured image was 5184 pixels × 3456 pixels. The vertical distance H1 between the camera and the ground was 80 cm, and the horizontal distance H2 between the camera and the soybean plant was always kept at 100 cm. The vertical distance H4 was adjusted between the lifting platform and the ground according to the height of the soybean plant, so that the vertical distance H3 between the soybean plant and the ground was kept at 95 cm. In order to make the acquired the images of soybean plants in the natural growth state, the windless and sunny weather conditions were selected to complete the data acquisition.

2.3 Data acquisition method

For the integrity of experimental data, soybean plants of two varieties in different growth stages were acquired as experimental samples from 2020 to 2022. In order to avoid the overfitting phenomenon caused by the insufficient diversity of sample data, all soybean potted plants were rotated 360 ° for three times with the marked point as the starting position, and each rotation angle was 120 °. In this experiment, soybeans were collected from R5. According to the growth status, the data were collected every 4-10 days. A total of 863 digital images of soybean plants in different growth stages were acquired, including 216 at R5, 216 at R6, 216 at R7 and 215 at R8. Figure 2 showed the

FIGURE 2
Soybean plants' images of two varieties in different growth stages **(A)** Single soybean plant of Suinong 26 ; **(B)** Multi soybean plants of Suinong 26; **(C)** Single soybean plant of Heihe 49; **(D)** Multi soybean plants of Heihe 49.

10.3389/fpls.2022.1096619

images of soybean plants of two varieties in different growth stages.

2.4 Data annotations

In order to ensure the accuracy of the data and effectively train the detection model, it was necessary to manually label the data before data training. In this study, the LabelImg, an image annotation tool, was used to locate and mark the pods on 863 soybean plants images acquired in different growth stages. When labeling, the smallest circumscribed rectangle of the pods was taken as the real box, so as to reduce the useless pixels on the background in the box. The marked results were stored in .xml file, which contained information such as the position of the pods, the size of the anchor boxes and the label of the pods in the image. A total of 21650 pods were marked on 863 pictures in different growth stages. Figure 3 showed the labeling of single and multiple soybean pods of two varieties in different growth stages.

2.5 Data augmentation and dataset partitioning

In this paper, according to the distribution characteristics of pods in soybean plants, the raw images were rotated, and then Gaussian noise was added and brightness was changed without changing the original data characteristics, so as to improve the detection accuracy and realize the robustness of the model.

After data augmentation of the original data, due to the impact of the quality of the raw image, some picture labels might exceed the pod's calibration range or suffer serious quality loss. Therefore, it was necessary to manually select pictures. Finally, 322 poor quality pictures were removed, and 3130 pictures were obtained, including 863 raw images and 2267 enhanced images. In this paper, the raw images and the enhanced images were combined into a data set, including images and labels, a total of 6260 files. The images in all data sets were divided into training set, testing set and validation set according to ratio of 7:2:1. The specific distribution was shown in Table 1.

FIGURE 3
Soybean pod's marking images of two varieties in different growth stages **(A)** Single soybean plant of Suinong 26; **(B)** Multi soybean plants of Suinong 26; **(C)** Single soybean plant of Heihe 49; **(D)** Multi soybean plants of Heihe 49.

TABLE 1 Distribution of soybean plants in each data set in different growth stages.

Data set	Total number of images	The number of pictures in different growth stages			
		R5	R6	R7	R8
Training set	2191	550	549	549	543
Testing set	626	157	159	156	154
Validation set	313	83	78	77	75

3 Principle and method

3.1 Overall framework of pods' recognition and yield prediction methods

First, the soybean plants data in different growth stages (R5~R8) were dynamically obtained by using the acquisition system of soybean plants and data augmentation and dataset expansion were carried out by using image processing algorithms, including random rotation, Gaussian filtering and adjusting image primary color. On this basis, the YOLOv5 model was improved by embedding CA module and using EIOU Loss instead of GIOU Loss as the regression loss function based on boundary box, so as to realize the recognition and counting of soybean pods in different growth stages. Then, the pod's length, width, area, chord length and convex arc length were calculated by using the minimum circumscribed rectangle method, the maximum inscribed circle method, the regional pixel counting and template calibration method, and the combination of contour convex hull and endpoint detection method. The validity of the pod's phenotype calculation method was verified by establishing the linear correlation between measured and calculated values. Finally, the weight estimation model of single pod was constructed based on BP neural network with topological structure of 5-120-1. The average weight of a single pod of two varieties was estimated. Combined with the number of pods per plant identified by the improved YOLOv5 algorithm, the weight of the whole plants pods was estimated. Figure 4 showed the overall framework of the soybean pod's recognition and estimation of all pods weight per plant.

3.2 Recognition method of the pod based on improved YOLOv5 algorithm

3.2.1 YOLOv5 network model

YOLOv5 network model was mainly divided into four parts: input end, backbone network, characteristic network architecture and output end (Gu et al., 2022). The input end was mainly used to preprocess the image. Mosaic data augmentation operation (Zhao B. et al., 2021), adaptive anchor box calculation (Gao et al., 2019) and adaptive image scaling were used to scale the input image to the input size of the network, and normalization was performed at the same time. The backbone network structure mainly included the Focus module that sliced the images, the Bottleneck Cross Stage Partial (CSP) module and the Spatial Pyramid Pooling (SPP) module were used to fix the image size. Feature network architecture mainly solved the problem of multi-scale detection in target detection. SPP module and FPN+PAN module located in the middle of backbone network and input terminal were used to further improve the diversity and robustness of target features. The output included a classification branch and a regression branch, which were mainly used to output the target recognition results.

3.2.2 Embedding CA attention mechanism

Attention mechanism mimics biological vision and mainly scans the whole image quickly to screen out the regions of interest, and invests more attention resources to suppress other useless information and improves the efficiency and accuracy of visual information processing (Qi et al., 2022). Because the soybean pods target was small and easy to be disturbed by background factors, YOLOv5 model was easy to lose the characteristic information of small targets during convolution sampling, resulting in missed detection and false detection. Therefore, the CA module (Xu X. et al., 2022) was embedded in YOLOv5 model. Through the channel attention module and spatial attention module, the weight of small targets in the whole feature map could be increased effectively. The effective extraction of the target pod's feature information was realized, and the accuracy of soybean pods' recognition was improved in different growth stages. Figure 5 showed the structure of CA module.

First, the global pooling operation was performed in Channel Attention Module on the spatial dimension of the input characteristic graph with the size of C×H×W. The size of feature map after operation was C×H×1. The eigenvalues obtained by the average pooling operation mainly described the background information of the image, and the eigenvalues obtained by the maximum pooling operation mainly described the texture information of the image. Then, the pooled results were sent to the two shared parameters of neural networks respectively. After spliced in the channel dimensions, the two groups of pooled output results were multiplied and added separately, and the weight range was constrained to the (0,1) interval with the help of the activation function. Finally, the input feature map was weighted to obtain the channel attention feature map, so as to enhance the expression of pod information, suppress the expression of useless background information, and improve the recognition effect. The output channel representation of the height and width of the target box was as follows:

$$Z_c^h(h) = \frac{1}{w} \sum_{0<i\le w} X_c(h, i) \tag{1}$$

$$Z_c^w(w) = \frac{1}{H} \sum_{0<i\le H} X_c(j, w) \tag{2}$$

Where, Z_c^h represented the output of the c channel with a height of h, and Z_c^w represented the output of the c channel with a width of w.

10.3389/fpls.2022.1096619

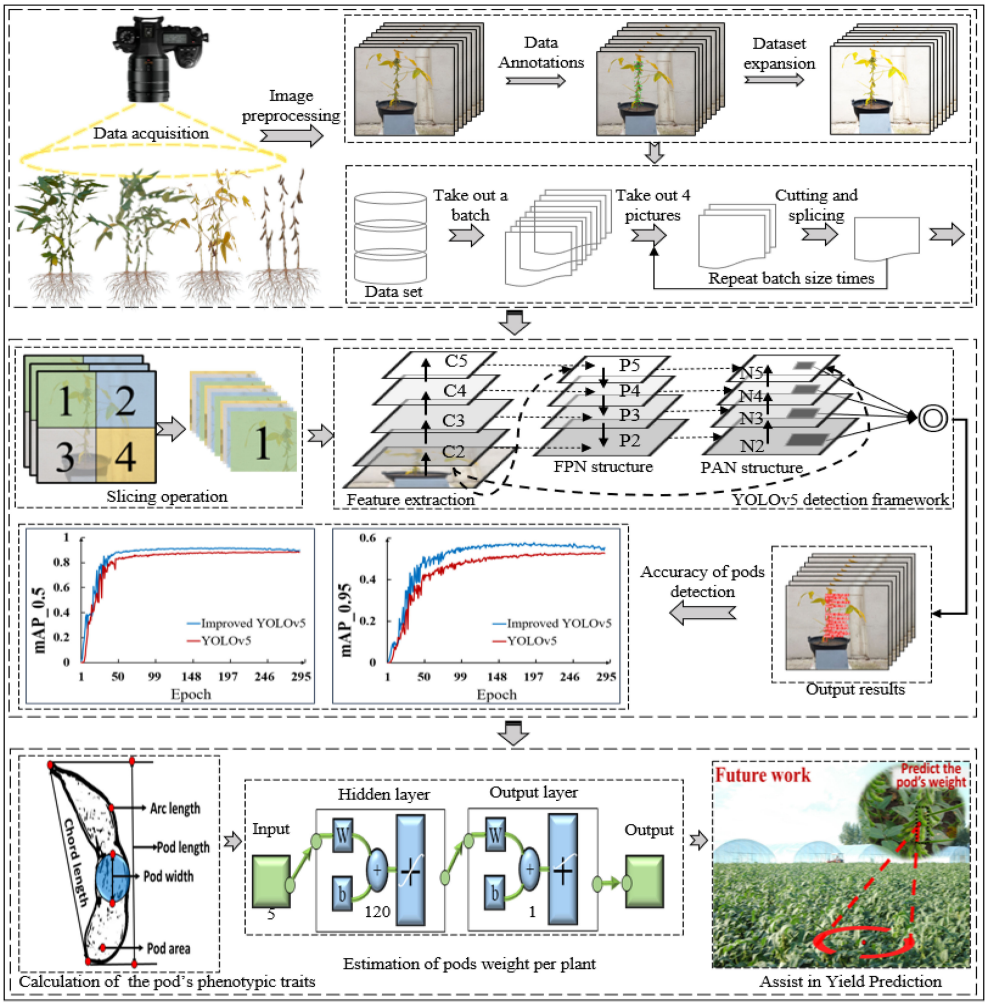


FIGURE 4
Overall framework of the soybean pod's recognition and yield estimation.

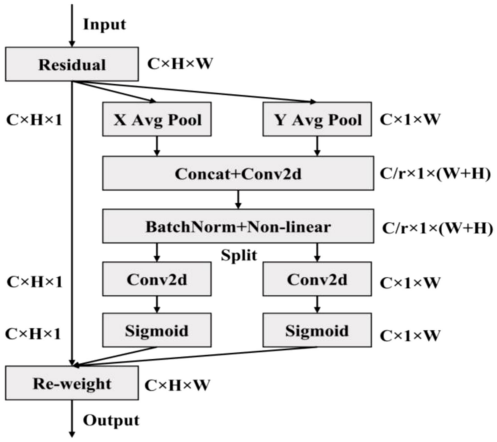


FIGURE 5
Structure of the CA module.

Equation (1) and (2) aggregated features along two spatial directions, returning a pair of directional perceptual attention features Z_c^h and Z_c^w, and generated a pair of feature maps at the same time, making the target location to be detected more accurate.

Different from the channel attention module, the spatial attention module extracted features through the spatial dimension information of the feature map. First, the average pooling and maximum pooling operations were performed in the spatial attention module on the channel dimension for the input characteristic graph of size C×H×W. The size of the two feature maps after operation was 1, which reduced the increase of parameters. The two one-dimensional feature maps were spliced into a two-dimensional feature map based on channel dimensions. Then, 7×7 convolution layers were used to extract the mask map that described the location information of the feature space in the feature map. After constrained by the activation function, the mask map was applied to the input feature map to obtain the spatial attention feature map enhanced according to the spatial location information, so as to improve the expression of pod shape, size, color, texture and other features. The specific Equation was as follows:

$$f = \delta(F_1(\left[z^h, z^w\right])) \tag{3}$$

$$g^h = \sigma(F_h(f^h)) \tag{4}$$

$$g^w = \sigma(F_w(f^w)) \tag{5}$$

$$y_c(i,j) = x_c(i,j) \times g_c^w(j) \tag{6}$$

Where, F_1 represented 1*1 convolution, f represented the intermediate feature image obtained through the down sampling operation δ, and two separate tensors f^h and f^w could be obtained after segmentation along the spatial dimension. Then, g^h and g^w with the same channel number as the input image X could be obtained by 1*1 convolution F_h, F_w and σ transformation. After expansion, they were added to the input as attention weights, and y represented the final output image.

The CSP module before and after embedded the attention mechanism was shown in Figure 6. The CA module decomposed the channel attention into two one-dimensional features along different spatial directions for coding, which could not only capture the long-range dependence along one spatial direction, but also saved the accurate location information along the other direction, while expanding the global receptive field of the network. This method could not only locate the pod's target more accurately, but also saved a lot of computing overhead.

3.2.3 Improved border regression loss function

Loss function was one of the important criteria to judge whether a model was applicable to the current data set. This function was used to characterize the fitting degree between the predicted value and the real value. When the loss function curve gradually converged, the model had achieved a relatively ideal prediction effect (Lv and Lu, 2021). Yolov5 used GIOU Loss as the loss function of the bounding box (Wang H. et al., 2022), and used binary cross entropy and Logits loss function to calculate the loss of class probability and target score. The calculation Equations were as follows:

$$Loss_{coord} = \sum_{i=0}^{S^2}\sum_{j=0}^{B} 1_{ij}^{obj}(1 - GIOU_{ij}) \tag{7}$$

$$GIOU_{ij} = \frac{J}{U} - \frac{A - U}{A} \tag{8}$$

$$U = \widehat{\omega_i} \cdot \widehat{h_i} + \omega_i \cdot h_i - J \tag{9}$$

$$IOU = \frac{A \cap B}{A \cup B} \tag{10}$$

Where, $Loss_{coord}$ represented the loss function of the target location, S represented the grid of S ×S each containing the prediction results, B represented two prediction boxes, 1_{ij}^{obj} represented the target contained in the prior box j generated by cell i, J represented the intersection area of the border, U represented the union area of the border, A represented the minimum circumscribed rectangular area of the border, ω_i and h_i represented the length and width of the prediction box, respectively, IOU represented the ratio of intersection and union between prediction frame and real frame. When the prediction frame coincided with the real frame, the maximum GIOU was 1. On the contrary, as the distance between the prediction frame and the real frame increased, GIOU tended to be -1, that was, the farther the distance between the prediction frame and the real frame was, the greater the loss value was. When the distance between the prediction frame and the real frame was inclusive or the width and height were aligned, and the difference set was 0, the loss function could not be derived and could not converge, which was easy to cause false detection and missed detection for the pods covered by leaves or stems. Therefore, in this study, the EIOU Loss frame regression loss function was used instead of

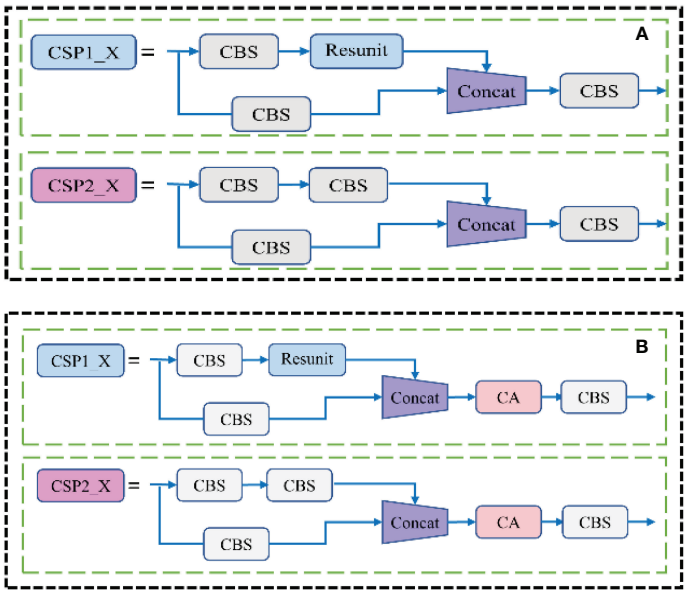


FIGURE 6
Model comparison before and after the CA module was embedded in the CSP structure **(A)** Before improvement; **(B)** After improvement.

the GIOU Loss frame loss function as the deviation index of the prediction frame deviation (Fan et al., 2022). EIOU Loss function included three parts: overlap loss, center distance loss and width height loss. The width height loss directly minimized the difference between the width and height of the target box and the anchor box. EIOU Loss was shown in Equation (11):

$$L_{EIOU} = L_{IOU} + L_{asp} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{c_w^2} + \frac{\rho^2(h, h^{gt})}{c_h^2} \quad (11)$$

Where, b was the center point of the prediction frame, b^{gt} was the center point of the real frame, ρ was the Euclidean distance between the two center points, c was the diagonal distance of the smallest closure area that can contain both the prediction frame and the real frame, c_w and c_h were the width and height of the minimum circumscribed rectangle box covering the prediction frame, respectively, w and w^{gt} represented the width of the prediction box and the real box, respectively.

In the boundary box regression loss function, EIOU Loss avoided the non- convergence when the real box and the prediction box were in the inclusion relationship, and could improve the recognition accuracy of pods in the case of occlusion effectively. Its width and height loss made the convergence speed faster and the accuracy higher, and its performance was better than that of the traditional YOLOv5 boundary loss function.

3.3 Estimation of soybean yield based on multi-dimensional pod's phenotypic traits

3.3.1 Calculation method of pod's phenotypic traits

The yield of soybean crops was closely related to the number of pods per plant, the phenotypic characters of pods and the degree of plumpness of pods. Therefore, it was of great significance to quickly and accurately obtain the pod's physiological and ecological indicators that determined the grain yield and quality for cultivating soybean varieties with high yield and quality traits and estimating the yield. In this paper, when the soybean plant grew to full maturity, all the pods on the mature soybean plants of the two varieties were picked, and the phenotypic traits including length, width, area, chord length and arc length were automatically calculated (He et al., 2022). Figure 7 showed pods picked from selected soybean plants randomly.

In order to calculate the phenotypic traits of soybean pods more accurately, it was necessary to carry out geometric distortion correction, Gaussian filter noise reduction (Wang G. et al., 2022), Canny edge detection operator (Luo et al., 2021) and morphological close operation to extract the pod's contour and other preprocessing on the pod's images. It provided a reliable data base for the calculation of phenotypic traits of soybean pods. Figure 8 was the schematic diagram for calculating the phenotypic traits of a single pod.

3.3.1.1 (1) Calculation of the pod length

In this study, the minimum circumscribed rectangle algorithm was used to determine the circumscribed rectangle of a single pod. The length of the long side of the circumscribed rectangle was taken as the length of the pods. The calculation ratio of the pods in the image was defined by the black marker blocks with length and width of two centimeters, and the actual size of the pods in the image was calculated by combining the Euclidean distance Equation. The specific definitions were as follows:

$$U = \frac{K_\alpha}{K_\beta} \quad (12)$$

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (13)$$

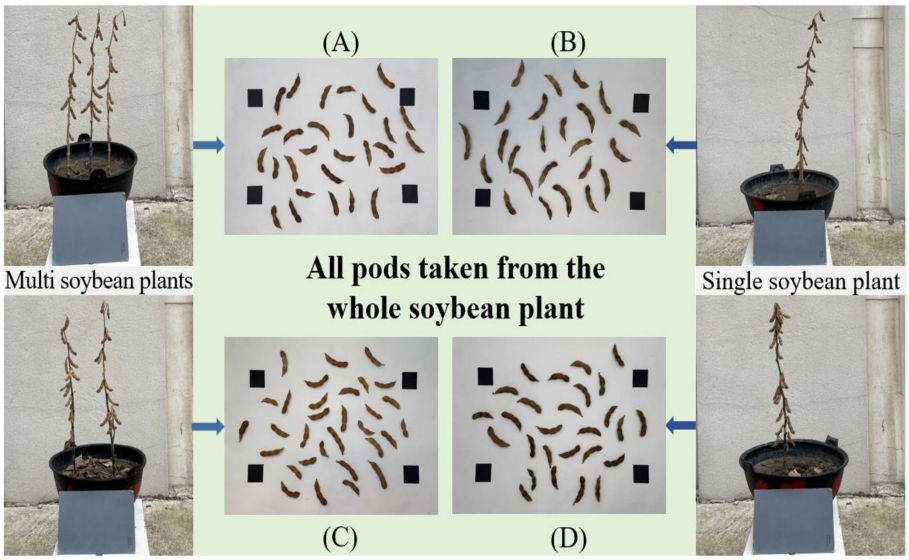


FIGURE 7
All mature soybean pods picked **(A)**, **(C)** Pods on multiple soybean plants; **(B)**, **(D)** Pods on single soybean plant.

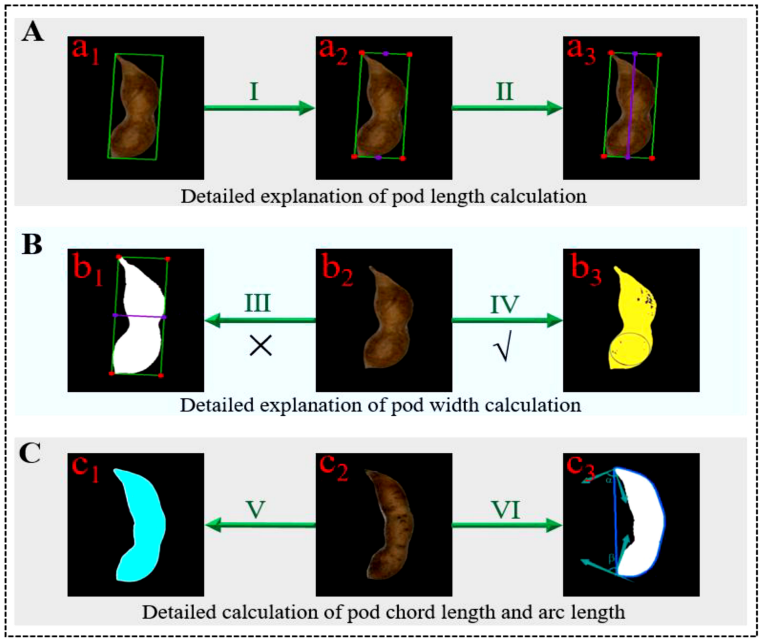


FIGURE 8
Steps for calculating phenotypic traits of a single pod **(A)** Calculation of pod length; **(B)** Calculation of the pod width; **(C)** Calculation of pod area, chord length and arc length.

Where, U represented the calculation ratio of the object size, K_α represented the pixel length of the object, and K_β represented the real length of the object. D represented the Euclidean distance between points (x_1,y_1) and (x_2,y_2). In Figure 8, A represented the decomposition diagram of pod length calculation, a_1 represented the schematic diagram of the minimum circumscribed rectangle of the pods, I represented the calculation of the vertex and the center points of the upper and lower sides of the rectangle, a_2 represented the drawing of the endpoint and midpoint of the rectangle, II represented the calculation of the Euclidean distance between the upper and lower center points of the rectangle, and a_3 represented the drawn centerline of the rectangle.

3.3.1.2 (2) Calculation of the pod width

According to the definition standard of pod width in the specification for the description of soybean germplasm resources, the widest part of the pods was taken as the pod width. Therefore, the maximum inscribed circle algorithm (Huang et al., 2021) was used to determine the maximum width of the pods by finding and calculating the maximum inscribed circle diameter of the pods contour. The Equations of maximum inscribed circle center and radius were as follows:

$$S = \frac{\sum_i x_i + x_{i-1})(y_i - y_{i-1})}{2} \tag{14}$$

$$m = \frac{\sum_i (x_i + x_{i-1})^2 (y_i - y_{i-1})}{S} \tag{15}$$

$$n = \frac{\sum_i (y_i + y_{i-1})^2 (x_i - x_{i-1})}{S} \tag{16}$$

$$D = 2R = 2\sqrt{(x_i - m)^2 + (y_i - n)^2} \tag{17}$$

Where, s was the area of the closed boundary contour, x_i was the abscissa on the closed boundary contour, y_i was the ordinate of the closed boundary contour, m was the abscissa of the center of the circle, n was the ordinate of the center of the circle, R was the radius of the maximum inscribed circle of the pods, and D was the maximum width of the pods. In Figure 8, B represented the schematic diagram of pod width calculation, in which III represented the common wrong calculation method of pod width, that was, it was considered that the width of the minimum circumscribed rectangle of the pods was the pod's width, and b_1 was its schematic diagram. IV represented the correct calculation method of pod width used in this paper, b_2 was the raw diagram of pod, and b_3 was the schematic diagram of the maximum width of pod.

3.3.1.3 (3) Calculation of the pod area

In this study, the extracted pod's contour was expanded and filled smoothly on the premise of clear image boundary. Then the pod area was calculated by combining the regional pixel counting and the template calibration method. Finally, the actual area of the standard reference block in the image was used to calculate the area of the pods. The calculation Equation was as follows:

$$S_d = \frac{W_d \times S_k}{W_k} \tag{18}$$

Where, S_d represented the actual area of the pods, W_d represented the number of pixels of the pods, S_k represented the actual area of the standard marker block, and W_k represented the number of pixels of the standard marker block. In Figure 8C, V represented the

calculation of pod area, c_1 represented the calculation diagram of pod area, and c_2 represented the raw diagram of pod.

3.3.1.4 (4) Calculation of the chord and arc length

At present, the ratio of chord to arc was an index used by some breeders to describe the bending degree of the pods. The greater the ratio of chord to arc, the straighter the pod's shape, the smaller the ratio, the more curved the pods. In this paper, first, the perimeter of the soybean pods contour was counted, and the polygon contour of the soybean pods was obtained by using the convex hull algorithm (Liu et al., 2018). Then, starting from the middle point on the left side of the contour, the vector angles of the two points close to the contour were calculated, and the two angles with the largest difference were obtained by dislocation subtraction. The two corresponding points were the two endpoints of the pod contour. In combination with Equation (13), the Euclidean distance between the two endpoints was the pod's chord length. By making a difference between the perimeter of the external polygon of the pods and the chord length of the pods contour, the convex arc length of the pods was obtained. The vector angles of the two adjacent points of the pods were as follows:

$$\cos\alpha = \frac{(x_1 y_1 + x_2 y_2)}{\sqrt{(x_1^2 + y_1^2)(x_2^2 + y_2^2)}} \tag{19}$$

Where, cosα was the cosine of the tangent vector of the two adjacent points of the pods contour, x_1 and x_2 were the abscissa of the two adjacent points of the pods contour respectively, y_1 and y_2 were the ordinates of the two adjacent points of the pods contour respectively. Figure 8 VI showed the calculation of pod chord length and arc length, and c3 was its schematic diagram, α, β represented the included angle between the two ends of the pod and the adjacent points.

3.3.2 Estimation method of soybean yield

Because the soybean yield had the characteristics of randomness and nonlinearity, the accurate mathematical model could not predict the soybean yield, effectively. Therefore, in this study, a three-layer nonlinear BP neural network (Jiang et al., 2021) system was used to predict soybean yield. Due to the strong self-adaptive resolution performance and fault tolerance, BP network could approximate continuous nonlinear functions with arbitrary accuracy and had significant local approximation characteristics (Guan et al., 2013; Roopali and Toran, 2020), which provided a technical guarantee for accurate and efficient prediction of soybean yield. The pod's phenotypic traits were the dimension of the feature space, which determined the number of nodes in the input layer. In this study, five traits including pod length, pod width, pod area, chord length and arc length were selected as the input values of the neural network, so the number of nodes in the network input layer was 5. The number of nodes in the output layer was determined according to the dimension of the mode space. Because the prediction result of pod weight was a specific value, the number of output nodes was 1. According to the theorem *Kolmogorov* (Liu M. et al., 2020) and practical experience, the number of neuron nodes in the hidden layer was determined to 120. Finally, the BP network topology for predicting the weight of the pod was 5-120-1 type, and each layer was fully interconnected (Figure 9). In order to reduce the training time and complete the training task efficiently, the iteration accuracy of training target was defined as 0.01, which was the termination condition of model training. The initial learning rate was set to 0.01 and gradually increased. The optimal learning rate was finally determined to 0.8 based on the value of training cost.

4 Results and analysis

4.1 Analysis of pods recognition results

4.1.1 Training of pods recognition model

4.1.1.1 (1) Setting environment parameters

This study used pytorch 1.9.0 machine learning framework, and the graphics processing unit (GPU) used NVIDIA GeForce GTX 1050 Ti (4096MB). Soybean pods recognition models in different growth stages were trained using YOLOv5 and improved YOLOv5 models on Windows 10 64-bit operating system. The input size of the images had a great impact on the performance of the detection model. Because a feature map ten times smaller than the raw images will be generated in the basic network, the details of smaller pods were not easy to capture. Thus, the input size of the image was adjusted to 640 × 640 (pixels) for training, which can improve the robustness of the detection model to the object size to a certain extent. In addition, the initial learning rate of the model, the size of the super parameter and the attenuation

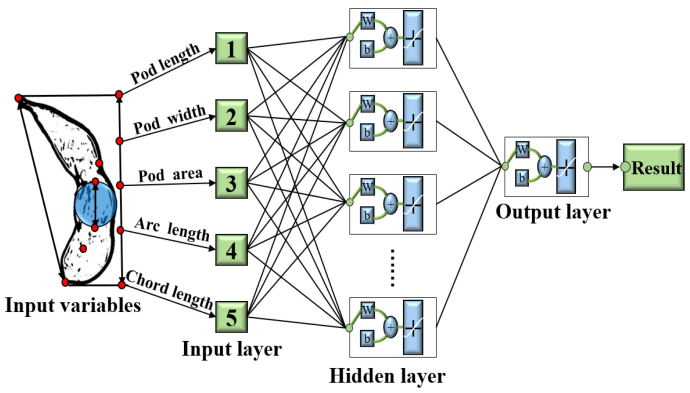


FIGURE 9
Structure diagram of neural network for estimation of single pod weight.

coefficient as well as the training round would affect the convergence of the loss function, thus affecting the accuracy of the model training. According to the parameter setting method proposed in the literature (Li. et al., 2022; Qiu et al., 2022), the initial learning rate was 0.001, the cosine annealing super parameter was 0.12, and the attenuation coefficient was 0.00036. A total of 300 epochs were trained.

4.1.1.2 (2) Model training

Loss function was one of the important criteria to measure the prediction effect of a model on the current data set. It mainly mapped the values of relevant random variables to non-negative real numbers to represent the gap between the prediction results and the measured data. When the loss function curve converged gradually, the model had achieved an ideal prediction effect. The comparison of the loss function changed between the training set and the validation set of the YOLOv5 model and the improved YOLOv5 model was shown in Figure 10.

Figure 10 showed that during the pod's recognition in the training set and the validation set, the early loss function of the model decreased rapidly. With the increase of the number of training rounds, the loss curve gradually decreased and tended to be stable. The loss value of the first 60 training groups in the training set decreased rapidly. When the epoch reached about 140, the loss value of the algorithm decreased to be stable, and the loss function value stabilized at about 0.02. The loss value of the first 40 training groups in the validation set decreased rapidly. When the epoch reached about 60, the loss value of the algorithm decreased to be stable, and the stable value of the loss function was also about 0.02. It could also be seen from the image that the loss curve of the improved YOLOv5 model of the training set and the validation set was always below the YOLOv5 model. It also showed that the loss value of the improved YOLOv5 model was always smaller than that of the YOLOv5 model, that was, the positioning accuracy was higher, the convergence speed was faster, and the prediction effect was better when identifying soybean pods in different growth stages.

4.1.2 Prediction effect analysis of the recognition model

4.1.2.1 (1) Evaluation of recognition effect of pods before and after improvement

The prediction result of the model was the most intuitive way to evaluate the quality of a model. The pods in the testing set were recognized using the original model and the improved yolov5 model, which were shown in Figure 11.

Figure 11showed that for different varieties, Heihe 49 had better overall detection effect on soybean pod's than Suinong 26. For the same variety, the recognition effect of single plant pods was better than that of multiple plants pods. For example, A-R5 and B-R5, there were 23 pods in A-R5, 22 of which were recognized correctly, with a recognition accuracy rate of 96%. There were 23 pods in B-R5, 20 of which were identified correctly, with a recognition accuracy rate of 87%. For different stages of the same variety, the recognition performance of the model was gradually enhanced with the maturity of soybean pods. In D-R7 and D-R8, there were 17 pods in D-R7, 14 of which were correctly recognized, with an accuracy rate of 82.4%. There were 20 pods in D-R8, 17 of which were correctly recognized, with an accuracy rate of 85%. This was because the learning performance of the model was affected by plants growth, leaves shading and pod's maturity. The denser the leaves of soybean plants, the more serious the shielding between stems and pods, pods and pods, and between leaves and pods, and the worse the recognition effect of the model will be. As the pods mature, the greater the color contrast between pods and plant leaves and stems, the better the recognition effect of the model will be.

The prediction effect of different models was analyzed by comparing the number of Ture Positive (TP) and False Positive (FP). The calculation method for the number of TP and FP was to first obtain the real box and the prediction box after recognition of the pods. The prediction box contained the detection category, confidence score and coordinate information of the detection box. If the retention confidence score of the prediction box was greater than 0.3, the maximum matching IOU value was calculated between the prediction box and the real box. If the IOU value was greater than 0.5 and the two boxes were matched for the first time, the result was recorded as TP, otherwise it was recorded as FP. The more the number of TP, the higher the detection accuracy of the model, and the stronger the performance will become. The more the number of FP, the lower the detection accuracy of the model, and the worse the performance will become. The number of TP and FP of the original model and the improved YOLOv5 model in the test set was counted, which was shown in Table 2.

Table 2 showed that, for the number of TP, the improved YOLOv5 model had significantly more TP detected for soybean

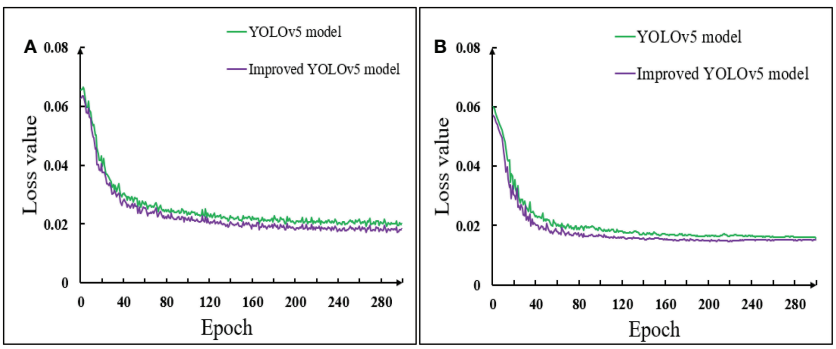


FIGURE 10
Comparison of loss function of model before and after improvement **(A)** Training set; **(B)** Validation set.

10.3389/fpls.2022.1096619

FIGURE 11
Recognition results of pods by using improved YOLOv5 model. **(A)** Single soybean plant of Suinong 26; **(B)** Multi soybean plants of Suinong 26; **(C)** Single soybean plant of Heihe 49; **(D)** Multi soybean plants of Heihe 49.

pods in different growth stages than the YOLOv5 model. The number of TP at R5-R8 was 87, 126, 22 and 7 more than that in the original model, respectively. For the number of FP, the number of FP detected by the improved YOLOv5 model for soybean pods in different growth stages was significantly less than that of the YOLOv5 model, in which the maximum difference in the number of FP in R5 was 168, and the minimum difference in the number of FP in R8 was 21. It could also be seen from the table that the number of FP was the largest in R5. With the growth of soybean plants, the number of FP detected by the model was gradually decreasing. The main reason for this situation was that the soybean plants had dense leaves at R5 and R6 stages, and the leaves and stems had a serious shelter against the pods, and the morphology of some new leaves was very similar to that of the pods, resulting in the error of model recognition. At R7 and R8 stages of soybean plants, with the falling off of leaves and stems, the blocking

TABLE 2 Comparison of TP and FP between YOLOv5 model and improved YOLOv5model.

Model	TP/FP	Growth stage of soybean plants			
		R5	R6	R7	R8
YOLOv5	TP	1708	1725	1707	1697
	FP	504	378	336	227
Improved YOLOv5	TP	1795	1851	1729	1704
	FP	336	297	266	206

10.3389/fpls.2022.1096619

phenomenon gradually disappeared, so the error rate of pod's recognition decreased.

In order to more intuitively compare the difference between the detection results before and after the model improvement, a pot of soybeans in each of the four stages identified and output by YOLOv5 and the improved YOLOv5 algorithm was randomly selected in the testing set. The difference between the recognition results of the two models was shown in Figure 12.

Figure 12 showed that all the pods not detected in the YOLOv5 model of soybean plants at R5 and R8 stages were recognized in the improved YOLOv5 model. At R6 stage, 8 pods were not recognized in the YOLOv5 model and 6 more were recognized in the improved YOLOv5 model, but 2 pods were still not recognized because the leaves were too dense and covered by the stems. At R7 stage, 4 pods were not recognized in the YOLOv5 model, 3 more were recognized in the improved YOLOv5 model, and 1 pod was not recognized

FIGURE 12
Comparison of partial recognition results before and after model improvement **(A)** YOLOv5 model; **(B)** Improved YOLOv5 model.

because its shape and color were very similar to soybean leaves. For the situation that the YOLOv5 model had missed detection, the improved YOLOv5 model had been significantly improved. This result showed that the improvement of YOLOv5 model in this study was effective for the recognition of soybean pods in different growth stages.

4.1.2.2 (2) P-R curve analysis of model prediction before and after improvement

The precision (P), recall (R) and P-R curve (Wang Z. et al., 2022) of pods recognition were compared between YOLOv5 and the improved YOLOv5 model on the testing set (Figure 13). The precision and recall were a pair of contradictory variables, the higher the precision, the lower the recall. In order to balance the relationship between the two indexes, was the area under the P-R curve was used as the average precision value to evaluate the performance of the model. The specific Equation was as follows:

$$P = \frac{TP}{TP + FP} \times 100\,\% \tag{20}$$

$$R = \frac{TP}{TP + FN} \times 100\,\% \tag{21}$$

Where, P and R represented the precision rate and recall rate respectively, TP(True Positive) represented that the positive sample was judged as a positive sample, FP(False Positive) represented that the negative sample was judged as a positive sample, and FN(False Negative) represented that the positive sample was judged as a negative sample. The average precision AP value of pods recognition could be obtained by calculating the area of the lower part in the P-R curve coordinate system. Since this article only contained one class of identification targets, the AP value was the mAP value of all classes. The Equation was as follows:

$$AP = \int_0^1 PRds \tag{22}$$

$$mAP = \frac{1}{N}\sum_{m=1}^{N} AP_m \tag{23}$$

Where, AP represented the area below the P-R curve, N represented the total number of categories. mAP was the result of averaging the AP values of all prediction categories. The larger the mAP value was, the better the prediction effect of the model was.

Figure 13A showed that the precision rate of the model before and after the improvement was relatively close, and both of them began to converge in about 80 groups. However, it was impossible to judge the prediction effect of a model only from the precision rate. In Figure 13B, the models before and after the improvement began to converge in about 60 groups, but the recall of the improved YOLOv5 model was always greater than that of the YOLOv5 model. Figurse 13C, D showed that, in the YOLOv5 model, when the recall rate was less than 0.45, the precision rate remained at 1. In the improved YOLOv5 model, when the recall rate was less than 0.55, the precision rate remained at 1. For the YOLOv5 model, the inflection

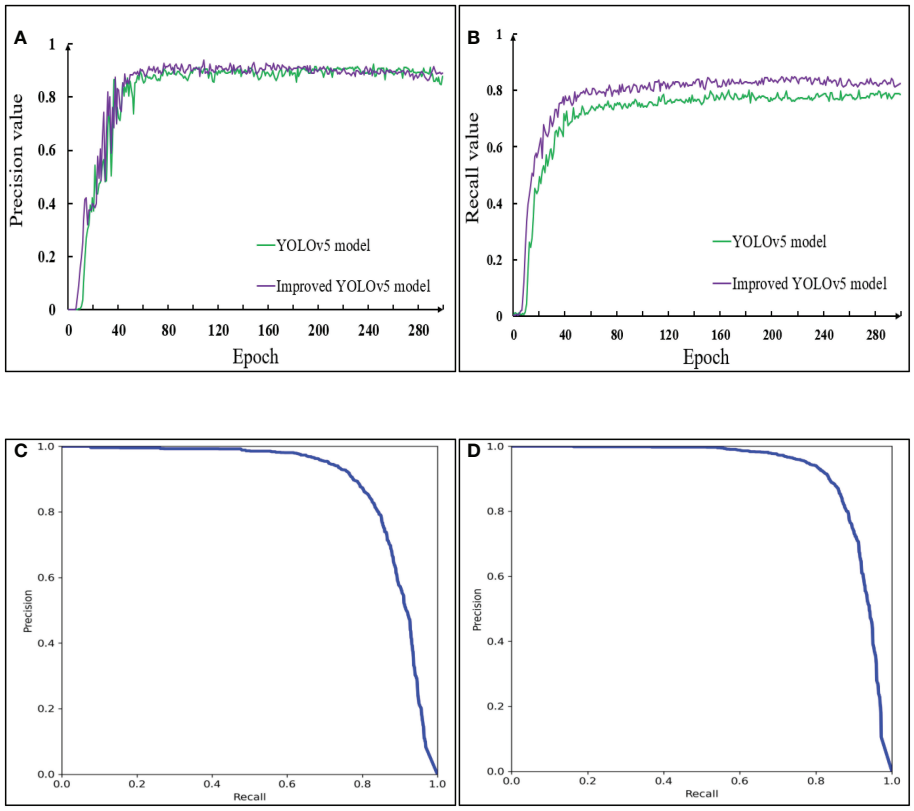


FIGURE 13
Comparison of P-R curve between YOLOv5 model and improved YOLOv5 model in the testing set. **(A)** Precision; **(B)** Recall; **(C)** P-R curve of the YOLOv5; **(D)** P-R curve of the Improved YOLOv5.

point of P-R curve appears before the recall rate was 0.8, while for the improved YOLOv5 model, the inflection point of P-R curve appears after the recall rate was 0.8, and the P-R curve was smoother. This was because when the recall rate became larger, the advantage of the model with CA attention mechanism module became obvious and the learning ability became stronger. The original model thought that the contribution of each region in the images was evenly distributed, but in the actual detection process, affected by the size, color and occlusion of the pods, the model had different and complex regions of interest for different images. The model after embedding the attention mechanism focused on the information useful for the detection category, and the use of EIOU Loss frame regression loss function took into account the real difference between the width and height of the pods and the confidence, so that the occluded pods were not easy to be incorrectly detected or missed. In addition, the average precision of pods recognition of the YOLOv5 model was 88.7%, and that of the improved YOLOv5 model was 91.7%. The mAP value of the improved model was increased by 3%. The results showed that the improved model had strong generalization ability and higher recognition accuracy for pods in different growth stages.

4.1.3 Performance evaluation of pod recognition model

In this paper, the performance of the traditional YOLOv5 and the improved YOLOv5 models for pod's recognition was compared and analyzed by using four indicators: detection rate (Xu Z. et al., 2022), test time, calculation amount (FLOPs) required for processing an image, and parameter amount (Hsia et al., 2021). Table 3 showed the comparison of pods recognition performance of different models in the data set.

FPS referred to the number of images that the model could process per millisecond. The larger the FPS, the higher the rate of the model and the better the performance. Test time referred to the time taken to process all images in the testing set, including pre-processing time, network pre-processing time and post-processing time. The calculation quantity and parameters quantity of the model reflected the complexity of the model. Table 3 showed that the improved YOLOv5 model could process 24.39 pictures per second, the test time of the model was 32.759s, and the calculation amount and parameter amount of the model were 4.79×10^9 and 2.085×10^7, respectively. Compared with the original YOLOv5 model, the detection rate was increased by 34.16%, the test time was saved by 11.499s. The calculation amount and parameter amount of the model were reduced by 17% and 7.6%, respectively. From the above analysis, it could be seen that the improved YOLOv5 model, which embedded CA module and EIOU regression loss function, could not only reduce the amount of calculation and parameters of the model, but also improved the accuracy of pods recognition.

4.2 Results analysis of soybean yield estimation

4.2.1 Result analysis of pod's phenotype calculation

The actual manual measurement value of electronic vernier caliper was used as the standard value of pod's length, width and chord length. The number of pixels counted by Photoshop tool (PhotoshopCC2017, San Jose, US) was used as the standard value of the pod's area and convex arc length. In order to verify the effectiveness of the calculation method of phenotypic traits for soybean pods, Equations (24-26) were used to calculate the relative error, average absolute error and average relative error between the calculated and measured values. The correlation between the calculated and measured values of pod length, pod width, pod area, chord length and arc length were shown in Figure 14.

$$P_i = \frac{|x_i - \overline{x_i}|}{\overline{x_i}} \times 100\,\% \tag{24}$$

$$AE = \frac{1}{m}\sum_{i=1}^{m}|x_i - \overline{x_i}| \tag{25}$$

$$RE = \frac{1}{m}\sum_{i=1}^{m}p_i \times 100\,\% \tag{26}$$

Where, p_i represented the relative error in the calculation of phenotypic traits of a single pod; AE represented the mean absolute error in the calculation of phenotypic traits of a single pod; RE represented the average relative error in the calculation of phenotypic traits of a single pod; x_i represented the calculated value of phenotypic traits of a single pod; $\overline{x_i}$ represented the measured value of phenotypic traits of a single pod; m represented the number of pods.

Figures 14A–E showed the linear correlation between the calculated and measured values of pod length, pod width, pod area, chord length and arc length at maturity, with coefficients of determination of 0.962, 0.939, 0.976, 0.930 and 0.929, respectively. It could be seen from the figure that the length of pods at maturity was mainly distributed between 4.5~5.5cm, the width was mainly distributed between 0.95~1.2cm, the area was mainly distributed between 2.5~5cm^2, and the chord length and arc length were mainly distributed between 3.5~4.5cm and 3.8~5.3cm. According to Equation (25), the average absolute errors of pod length, width, area,

TABLE 3 Comparison of performance evaluation indexes of different models.

Model	Model performance evaluation index			
	FPS (frame /ms)	Test time/ms	FLOPs	Parameter quantity
YOLOv5	18.18	44.258	4.79×10^{10}	2.845×10^7
Improved YOLOv5	24.39	32.759	3.95×10^{10}	2.085×10^7

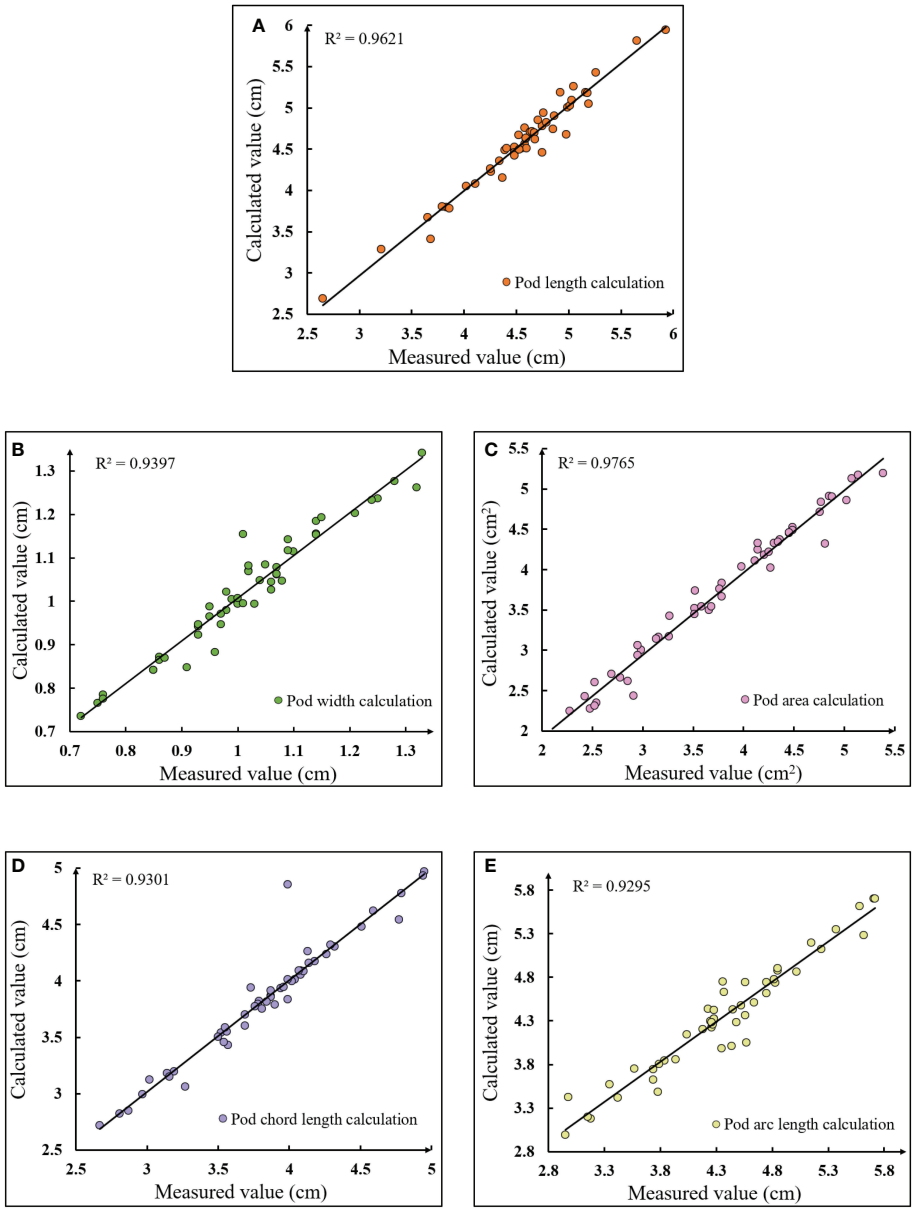


FIGURE 14
Linear correlation between calculated and measured values of pods phenotypic traits **(A)** Pod length calculation; **(B)** Pod width calculation; **(C)** Pod area calculation; **(D)** Pod chord length calculation; **(E)** Pod arc length calculation.

chord length and arc length were 0.084cm, 0.025cm, 0.096cm^2, 0.063cm and 0.12cm, respectively; According to Equation (26), the average relative errors were 1.85%, 2.46%, 2.83%, 1.65% and 2.86%, respectively. It could be seen that there was an obvious linear correlation between the calculated value and the measured value obtained by the pod's phenotypic traits calculation method. The average coefficient of determination R^2 was 0.947, and the average relative error was 2.33%. Thus, the phenotypic traits of pods could be calculated quickly and accurately by using the proposed methods.

4.2.2 Result analysis of soybean yield estimation

4.2.2.1 (1) Result analysis of estimating the weight of a single pod

In this study, 100 soybean pods of two varieties at R8 stage were trained using the proposed single pod weight estimation model. The training results were shown in Table 4. The pod's weight measured by the laboratory electronic weighing scale with an accuracy of 0.0001g was used as the standard value, and the pod's weight obtained by model training was used as the prediction value. The linear correlation between the predicted value of a single pod and the measured value was shown in Figure 15.

Table 4 showed that the estimated average weight of single pod of Heihe 49 and Suinong 26 was 0.726g and 0.409g, respectively, which was 0.041g and 0.049g lower than the actual value measured by the electronic libra with an accuracy of 0.0001g. The MSE reflected the difference between the actual value and the estimated value. The MSE of the two varieties of model training was 0.0089 and 0.0084, respectively. Figure 15 showed that the coefficient of determination R^2 between the estimated and measured single pod's weight of Heihe 49 and Suinong 26 was 0.9596 and 0.9311, respectively, and the single

TABLE 4 Statistics of pods weight estimation model at maturity stage.

Varieties	Average value of pods phenotypic traits (cm)					Estimated weight (g)	Actual weight (g)	MSE
	Pod length	Pod width	Pod area	Chord length	Arc length			
Heihe 49	4.56	1.02	4.11	3.85	4.34	0.726	0.767	0.0089
Suinong 26	4.20	0.98	3.70	4.14	4.43	0.409	0.458	0.0084

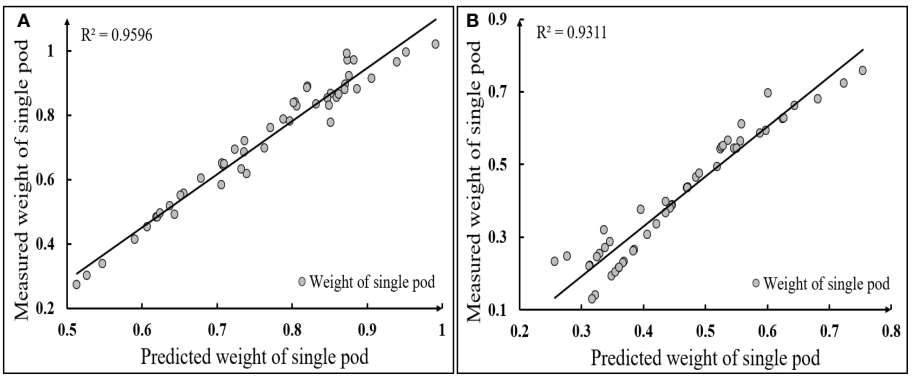


FIGURE 15
Linear correlation between estimated and measured values of single pod weight **(A)** Heihe 49; **(B)** Suinong 26.

pod's weight of Heihe 49 was mainly distributed between 0.6 and 0.9g, while that of Suinong 26 was mainly distributed between 0.3 and 0.7g. From the above analysis, it could be seen that the average of the MSE of the estimation of the single pod weight of the two varieties was 0.00865, and the average coefficient of determination between the estimated weight and the actual weight was 0.9453. Thus, the type of 5-120-1 BP neural network constructed in this study was effective and stable for the prediction of the pods' weight.

4.2.2.2 (2) Result analysis of estimating the weight of the whole plant pods

Three pots of soybean were randomly selected from Heihe 49 and Suinong 26. By combining the average number of pods per plant recognized by the improved YOLOv5 algorithm with the average weight of a single pod estimated in result analysis of pod's phenotype calculation, and the pods weight of all the potted soybeans plants of this variety could be predicted. Table 5 showed the prediction information of potted soybean yield.

Table 5 showed that combining the pod's number per plant estimated by the proposed method with the weight of a single pod. The predicted total weight of all potted soybeans of Heihe 49 and Suinong 26 were 165.58g and 1076.49g, respectively. According to Equation (24), the relative errors of pods weight prediction of Heihe 49 and Suinong 26 were 0.106 and 0.138 respectively. The relative error of pods weight prediction of Heihe 49 was 0.032, lower than that of Suinong 26, which mainly because Heihe 49 had shorter stems than Suinong 26 at maturity stage, and the degree of shielding between pods, pods and stems was smaller, so the accuracy of pods number detection was higher. According to the above analysis, the average relative error of pods weight prediction of the two varieties of soybean was 0.122. It could be seen that the average value calculated by the phenotypic traits of a single plant was effective for the yield prediction of all potted soybean plants which could provide a technical reference for the yield prediction of soybean plants in the field and the breeding of excellent soybean varieties.

5 Discussion

(1) Comparative analysis of related studies

The counting of pods and the estimation of pods weight per plant are of great significance for the breeding, cultivation and field management of soybean varieties. Due to the successful application of machine vision technology in the field of agricultural phenotype

TABLE 5 Prediction information of potted soybean yield.

Varieties	Weight of single pod (g)	Number of pods per plant	Estimation of pods weight per plant (g)	Predicted weight of all pods (g)	Measured weight of all pods (g)	Relative error
Hehei 49	0.726	19	13.794	165.58	149.60	0.106
Suinong 26	0.409	28	11.452	1076.49	945.70	0.138

exploration, researchers began to use these technologies to obtain soybean phenotypes in high-throughput and high-precision, so as to speed up crop improvement and breeding of new varieties to achieve high yield of soybean plants. Guo, et al improved the YOLOv4 target detection algorithm (Guo et al., 2021). For mature soybean plants without leaves, the number of pods per plant and the number of seeds in pods were detected. In this study, the coordinate attention mechanism was combined with the traditional YOLOv5 model to achieve the accurate recognition of pods in complex images with leaf occlusion under the outdoor growth state. Compared with the literature (Ning et al., 2021), the recognition accuracy of the number of pods in this study has been improved by 5.46%, and the prediction model of soybean yield has been constructed by combining BP neural network. Compared with the calculation methods of soybean pod length, width, area and other phenotypic traits proposed in the literature (Uzal et al., 2018; Yan et al., 2020; Li et al., 2021; Yang et al., 2022), the intelligent calculation methods used in this study not only had higher accuracy and less calculation, but also do not require manual marking, avoiding the error caused by human subjective judgment effectively.

(2) Establishment of pods recognition model

In this paper, an improved YOLOv5 model was proposed, which realized the accurate recognition of pods in different growth stages, and the recognition accuracy rate reached 91.7%. It solved the problems of mutual occlusion, unclear edges and difficult detection of small target pods in the process of soybean pods recognition. However, from the actual test results, it was found that a few pods were not correctly detected due to serious occlusion or similar shape and color to the leaves. In view of this phenomenon, the acquisition equipment with higher resolution should be used to obtain the soybean plant images, and the soybean plants should be rotated for several times to shoot, and the images with the least occlusion should be selected. At the same time, we should further improve the model feature extraction network from the internal structure of the network, so that the model paid more attention to the color, shape and texture features of pods, and improved the accuracy of pods detection.

(3) Error analysis

The average errors of soybean pod's recognition and yield prediction methods based on the improved depth learning model proposed in this paper were 8.3% and 12.2%, respectively. On the one hand, the error was caused by the algorithm itself. In the detection process, a few pods will be missed or wrongly detected. On the other hand, it was the geometric distortion caused by the machine vision system and shooting angle when acquiring images, and the cumulative error caused by the precision limitation of the instrument itself when measuring. In this regard, the acquisition equipment with higher resolution should be used and the camera position should be fixed to make it perpendicular to the object to be measured and maintain a fixed distance during the shooting process to reduce the geometric distortion caused by the raw image acquisition. For the accuracy limitation of the measuring instrument itself and the error in measurement, the average value of multiple measurements should be taken as the measured value to reduce the error.

(4) Application and promotion

In this study, CA attention mechanism was combined with YOLOv5 algorithm to construct a lightweight depth learning method, which was applied to the detection of pods in different growth stages successfully. The model had the advantages of high precision, small scale, less parameters, and greatly reduced the calculation amount of the model, which can meet the deployment requirements of portable mobile terminal devices. The future work should be based on the research content of this paper, combine the computer vision technology with the research of crop phenotypic parameters, and achieve accurate prediction of soybean yield in the field environment. For the method of narrow row and dense planting, high resolution color image sensor carried by Unmanned Aerial Vehicle (UAV) can be used to obtain multi angle soybean plant images. For the method of three ridge planting, the movable platform can be directly used to take the soybean plant images at the same time interval from different angles, and different degree of overlapping of pods, etc. Finally, recognition algorithms will be optimized to select images that best reflects the actual situation of the pods from the acquired images for this research.

6 Conclusion

In this study, a soybean pods recognition method based on improved YOLOv5 algorithm in different growth stages was proposed, and the prediction of soybean yield was realized by combining the pod phenotypic traits obtained by intelligent calculation methods. The experimental results showed that, the average accuracy of the proposed model reached 91.7%, increasing by 3% compared with traditional YOLOv5 model. The coefficients of determination R^2 between the calculated value and the measured value of pod's length, width, area, chord length and convex edge arc length were 0.962, 0.939, 0.976, 0.930 and 0.929, respectively. The MSE of single pod weight prediction was 0.00865, and the average coefficients of determination between the predicted value and the actual value was 0.945. Combined with the detection of pods per plant by the improved model, the MRE of all potted soybean yield predictions was 0.122. The proposed methods not only achieved high-precision recognition of pods and calculation of phenotypic traits, but also provided quantitative basis and technical support for estimation of soybean yield and cultivation of excellent soybean varieties in agronomy.

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

XM and HG conceived and designed the experiments. HH, PS and FW performed the experiments and acquired the digital image data of

soybean plants in different growth stages. HH and HG analyzed and processed the data. HH and XM wrote the paper. All authors contributed to the article and approved the submitted version.

Funding

Funding details, this study was funded jointly by Natural Science Foundation of Heilongjiang Province, China (LH2021C062), National Natural Science Foundation of China (31601220), Heilongjiang Bayi Agricultural University Support Program for San Heng San Zong, China (TDJH202101 and ZRCQC202006), Postdoctoral Scientific Research Developmental Fund of Heilongjiang Province, China (LBH-Q20053).

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

References

Anam, I., Redoun, I., S.M, R., S.M, M., and Mohammad, A. (2021). Rice leaf disease recognition using local threshold based segmentation and deep CNN. *Int. J. Intelligent Syst. Appl. (IJISA)* 13 (5), 35–45. doi: 10.5815/IJISA.2021.05.04

Chen, Z., Wu, R., Lin, Y., Li, C., Chen, S., Yuan, Z., et al. (2022). Plant disease recognition model based on improved YOLOv5. *Agronomy* 12 (2), 365. doi: 10.3390/AGRONOMY12020365

Fan, Y., Qiu, Q., Hou, S., Li, Y., Xie, J., and Qin, M. (2022). Application of improved YOLOv5 in aerial photographing infrared vehicle detection. *Electronics* 11 (15), 2344. doi: 10.3390/ELECTRONICS11152344

Fu, L., Wu, F., Zou, X., Jiang, Y., Lin, J., Yang, Z., et al. (2022a). Fast detection of banana bunches and stalks in the natural environment based on deep learning. *Comput. Electron. Agricult.* 194, 106800. doi: 10.1016/J.COMPAG.2022.106800

Fu, L., Zhou, W., Feng, Y., Zou, X., Lin, J., Cao, Y., et al. (2022b). YOLO-banana: A lightweight neural network for rapid detection of banana bunches and stalks in the natural environment. *Agronomy* 12 (2), 391. doi: 10.3390/AGRONOMY12020391

Gao, M., Du, Y., Yang, Y., and Zhang, J. (2019). Adaptive anchor box mechanism to improve the accuracy in the object detection system. *Multimedia Tools Applications* 78 (19), 27383–27402. doi: 10.1007/s11042-019-07858-w

Gao, F., Fang, W., Sun, X., Wu, Z., Zhao, G., Li, G., et al. (2022). A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard. *Comput. Electron. Agricult.* 197, 107000. doi: 10.1016/J.COMPAG.2022.107000

Guan, H., Du, S., Li, C., Su, J., Liang, Y., Wu, Z., et al. (2013). Electric shock signal recognition based on finite impulse response and radial basis function neural network. *J. Agric. Engineering* 29 (08), 187–194. doi: 10.3969/j.issn.1002-6819.2013.08.000

Guo, R., Yu, C., He, H., Zhao, Y., and Feng, X. (2021). Detection method of soybean pods number per plant using improved YoLOv4 algorithm. *J. Agric. Engineering* 18), 179–187. doi: 10.11975/j.issn.1002-6819.18.021

Gu, Y., Wang, S., Yan, Y., Tang, S., and Zhao, S. (2022). Identification and analysis of emergency behavior of cage-reared laying ducks based on YOLOv5. *Agriculture* 12 (4), 485. doi: 10.3390/AGRICULTURE12040485

He, H., Ma, X., and Guan, H. (2022). A calculation method of phenotypic traits of soybean pods based on image processing technology. *Ecol. Inform* 69, 101676. doi: 10.1016/J.ECOINF.2022.101676

Hsia, S., Wang, S., and Chang, C. (2021). Convolution neural network with low operation FLOPS and high accuracy for image recognition. *J. Real-Time Image Process.* 18 (4), 1309–1319. doi: 10.1007/S11554-021-01140-9

Huang, J., Yang, R., Ge, H., and Tan, J. (2021). An effective determination of the minimum circumscribed circle and maximum inscribed circle using the subzone division approach. *Measurement Sci. Technol.* 32 (7), 075014. doi: 10.1088/1361-6501/ABF803

Jia, W., Meng, H., Ma, X., Zhao, Y., Ji, Z., and Zheng, Y. (2021). Efficient detection model of green target fruit based on optimized transformer network. *J. Agric. Engineering* 37 (14), 163–170. doi: 10.11975/j.issn.1002-6819.2021.14.018

Jiang, X., Luo, S., Fang, S., Cai, B., Xiong, Q., Wang, Y., et al. (2021). Remotely sensed estimation of total iron content in soil with harmonic analysis and BP neural network. *Plant Methods* 17 (1), 1–12. doi: 10.1186/S13007-021-00812-8

Kong, D., Li, J., Zheng, J., Xu, J., and Zhang, Q. (2021). Research on fruit recognition and positioning based on you only look once version4 (YOLOv4). *J. Physics: Conf. Series* 2005 (1), 12020. doi: 10.1088/1742-6596/2005/1/012020

Lee, Y., and Shin, B. (2021). Development of potato yield monitoring system using machine vision. *J. Biosyst. Engineering* 45 (4), 282–290. doi: 10.1007/S42853-020-00069-4

Li, Y., Bao, Z., and Qi, J. (2022). Seedling maize counting method in complex backgrounds based on YOLOV5 and kalman filter tracking algorithm. *Front. Plant Sci.* 13. doi: 10.3389/FPLS.2022.1030962

Liu, M., Guan, H., Ma, X., Yu, S., and Liu, G. (2020). Recognition method of thermal infrared images of plant canopies based on the characteristic registration of heterogeneous images. *Comput. Electron. Agric.* 177, 105678. doi: 10.1016/j.compag.2020.105678

Liu, H., Jia, H., Wang, G., Glatzel, S., Yuan, H., and Huang, D. (2020). Corn stalk recognition method and experiment based on deep learning and image processing. *J. Agric. Machinery* 51 (04), 207–215. doi: 10.6041/j.issn.1000-1298.2020.04.024

Liu, R., Tang, Y., and Chan, P. (2018). A fast convex hull algorithm inspired by human visual perception. *Multimedia Tools Applications* 77 (23), 31221–31237. doi: 10.1007/s11042-018-6185-0

Li, S., Yan, Z., Guo, Y., Su, X., Cao, Y., Jiang, B., et al. (2021). SPM-IS: An auto-algorithm to acquire a mature soybean phenotype based on instance segmentation. *Crop J.* 10 (5), 1412–1423. doi: 10.1016/j.cj.2021.05.014

Lu, W., Du, R., Niu, P., Xing, G., Luo, H., Deng, Y., et al. (2022). Soybean yield preharvest prediction based on bean pods and leaves image recognition using deep learning neural network combined with GRNN. *Front. Plant Sci.* 12. doi: 10.3389/FPLS.2021.791256

Luo, C., Sun, X., Sun, X., and Song, J. (2021). Improved Harris corner detection algorithm based on canny edge detection and Gray difference preprocessing. *J. Physics: Conf. Series* 1971 (1), 12088. doi: 10.1088/1742-6596/1971/1/012088

Lv, H., and Lu, H. (2021). Research on traffic sign recognition technology based on YOLOv5 algorithm. *J. Electronic Measurement Instrumentation* 35 (10), 137–144. doi: 10.13382/j.jemi.B2104449

Mathew, M., and Mahesh, T. (2022). Leaf-based disease detection in bell pepper plant using YOLOv5. *Signal Image Video Processing* 16 (3), 841–847. doi: 10.1007/S11760-021-02024-Y

Ma, Y., Zhang, X., and Yang, G. (2021). Study on impurity segmentation of rice stalk based on improved mask r-CNN. *China J. Agric. Machinery Chem.* 42 (06), 145–150. doi: 10.13733/j.jcam.issn.2095-5553.2021.06.23

Mirhaji, H., Soleymani, M., Asakereh, A., and Abdanan, M. (2021). Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions. *Comput. Electron. Agricult.* 191, 106533. doi: 10.1016/J.COMPAG.2021.106533

Ning, S., Chen, H., Zhao, Q., and Wang, Y. (2021). Extraction of whole soybean phenotype information based on IM-SSD + ACO algorithm. *J. Agric. Machinery* 52 (12), 182–190. doi: 10.6041/j.issn.1000-1298.2021.12.019

Qi, J., Liu, X., Liu, K., Xu, F., Guo, H., Tian, X., et al. (2022). An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease. *Comput. Electron. Agricult.* 194, 106780. doi: 10.1016/J.COMPAG.2022.106780

Qiu, T., Wang, L., Wang, P., and Bai, Y. (2022). Research on target detection algorithm based on improved YOLOv5. *Comput. Eng. Application* 58 (13), 63–73. doi: 10.3778/j.issn.1002-8331.2202-0093

Roopali, G., and Toran, V. (2020). Tomato leaf disease detection using back propagation neural network. *Int. J. Innovative Technol. Exploring Engineering* 9 (8), 529–538. doi: 10.35940/ijitee.h6531.069820

Sharma, M., Kumar, C., and Deka, A. (2022). Early diagnosis of rice plant disease using machine learning techniques. *Arch. Phytopathol. Plant Protection* 55 (3), 259–283. doi: 10.1080/03235408.2021.2015866

Uzal, L. ,. C., Grinblat, G. ,. L., Namías, R., Larese, M. ,. G., Bianchi, J. ,. S., Morandi, E. ,. N., et al. (2018). Seed-per-pod estimation for plant breeding using deep learning. *Comput. Electron. Agricult.* 150, 196–204. doi: 10.1016/j.compag.2018.04.024

Wang, G., Wang, K., Ren, J., Ma, S., and Li, Z. (2022). A novel doublet-based surface plasmon resonance biosensor via a digital Gaussian filter method. *Sens. Actuators: B. Chem.* 360, 131680. doi: 10.1016/J.SNB.2022.131680

Wang, Z., Wu, L., Li, T., and Shi, P. (2022). A smoke detection model based on improved YOLOv5. *Mathematics* 10 (7), 1190. doi: 10.3390/MATH10071190

Wang, H., Zhang, S., Zhao, S., Wang, Q., Li, D., and Zhao, R. (2022). Real-time detection and tracking of fish abnormal behavior based on improved YOLOV5 and SiamRPN++. *Comput. Electron. Agricult.* 192, 106512. doi: 10.1016/J.COMPAG.2021.106512

Xu, Z., Huang, X., Huang, Y., Sun, H., and Wan, F. (2022). A real-time zanthoxylum target detection method for an intelligent picking robot under a complex background, based on an improved YOLOv5s architecture. *Sensors* 22 (2), 682. doi: 10.3390/S22020682

Xu, X., Tian, X., and Yang, F. (2022). A retrieval and ranking method of mathematical documents based on CA-YOLOv5 and HFS. *Math. Biosci. Engineering: MBE* 19 (5), 4976–4990. doi: 10.3934/MBE.2022233

Yang, S., Zheng, L., Yang, H., Zhang, M., Wu., T., Sun, S., et al. (2022). A synthetic datasets based instance segmentation network for high-throughput soybean pods phenotype investigation. *Expert Syst. With Applications* 192, 116403. doi: 10.1016/J.ESWA.2021.116403

Yan, Z., Yan, X., Shi, J., Sun, K., Yu, J., Zhang, Z., et al. (2020). Classification of soybean pods using deep learning. *J. Crops* 46 (11), 1771–1779. doi: 10.3724/SP.J.1006.2020.94187

Yu, H., Zhang, H., and Yuan, S. (2022). Research status of genetic regulation of soybean grain size. *Soil Crops* 11 (01), 18–20. doi: 10.11689/j.issn.2095-2961.2022.01.003

Zhang, P., Chen, Z., Ma, S., Yi, D., and Jiang, H. (2021). Prediction of soybean yield by canopy leaf color skew distribution model and RGB model. *J. Agric. Eng.* 37 (09), 120–126. doi: 10.11975/j.issn.1002-6819.2021.09.014

Zhao, B., Wu, Y., Guan, X., Gao, L., and Zhang, B. (2021). An improved aggregated-mosaic method for the sparse object detection of remote sensing imagery. *Remote Sensing* 13 (13), 2602. doi: 10.3390/RS13132602

Zhao, C., Zhao, X., Sun, L., Li, S., Guo, B., and Wang, R. (2021). Field identification and screening of soybean germplasm resources from different sources. *Northwest Agric. J.* 30 (11), 1638–1647. doi: 10.7606/j.issn.1004-1389.2021.11.006

Zhou, J., Zhou, J., Ye, H., Ali, M., Chen, P., and Nguyen, H. (2021). Yield estimation of soybean breeding lines under drought stress using unmanned aerial vehicle-based imagery and convolutional neural network. *Biosyst. Engineering* 204, 90–103. doi: 10.1016/J.BIOSYSTEMSENG.2021.01.017

Zhu, K., Ma, X., and Guan, H. (2021). A method of calculating the leafstalk angle of the soybean canopy based on 3D point clouds. *Int. J. Remote Sensing* 42 (7), 2463–2484. doi: 10.1080/01431161.2020.1854889

frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 18 January 2023
DOI 10.3389/fpls.2022.1098864

OPEN ACCESS

EDITED BY
Yijun Yan,
Robert Gordon University,
United Kingdom

REVIEWED BY
Yingjun Pu,
Southwest University, China
Marcin Wozniak,
Silesian University of Technology,
Poland
Yongliang Qiao,
The University of Sydney, Australia

*CORRESPONDENCE
Linsheng Huang
linsheng0808@ahu.edu.cn

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

RECEIVED 15 November 2022
ACCEPTED 19 December 2022
PUBLISHED 18 January 2023

CITATION
Zheng L, Zhao M, Zhu J, Huang L, Zhao J, Liang D and Zhang D (2023) Fusion of hyperspectral imaging (HSI) and RGB for identification of soybean kernel damages using ShuffleNet with convolutional optimization and cross stage partial architecture.
Front. Plant Sci. 13:1098864.
doi: 10.3389/fpls.2022.1098864

Fusion of hyperspectral imaging (HSI) and RGB for identification of soybean kernel damages using ShuffleNet with convolutional optimization and cross stage partial architecture

Ling Zheng, Mingyue Zhao, Jinchen Zhu, Linsheng Huang*, Jinling Zhao, Dong Liang and Dongyan Zhang

National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, Anhui University, Hefei, China

Identification of soybean kernel damages is significant to prevent further disoperation. Hyperspectral imaging (HSI) has shown great potential in cereal kernel identification, but its low spatial resolution leads to external feature infidelity and limits the analysis accuracy. In this study, the fusion of HSI and RGB images and improved ShuffleNet were combined to develop an identification method for soybean kernel damages. First, the HSI-RGB fusion network (HRFN) was designed based on super-resolution and spectral modification modules to process the registered HSI and RGB image pairs and generate super-resolution HSI (SR-HSI) images. ShuffleNet improved with convolution optimization and cross-stage partial architecture (ShuffleNet_COCSP) was used to build classification models with the optimal image set of effective wavelengths (OISEW) of SR-HSI images obtained by support vector machine and ShuffleNet. High-quality fusion of HSI and RGB with the obvious spatial promotion and satisfactory spectral conservation was gained by HRFN. ShuffleNet_COCSP and OISEW obtained the optimal recognition performance of ACC_p=98.36%, Params=0.805 M, and FLOPs=0.097 G, outperforming other classification methods and other types of images. Overall, the proposed method provides an accurate and reliable identification of soybean kernel damages and would be extended to analysis of other quality indicators of various crop kernels.

KEYWORDS

soybean damages, hyperspectral imaging, super resolution, image fusion, lightweight deep learning

1 Introduction

Soybean is one of the most important legume crops used as human food and animal feed in the world; it has 18%–22% oil and 38%–56% vegetable protein in its seeds (Arslan et al., 2018). Soybean has a protective effect against many diseases, such as high cholesterol, osteoporosis, cardiovascular, chronic diseases, and cancers (Xiao, 2008). The shell of soybean kernels is easily broken during transportation and storage because of its weak protective morphological arrangement; as such, soybeans are susceptible to mildew in the high-temperature and muggy atmosphere due to the post-maturation effect (Rani et al., 2013; Bessa et al., 2021). In broken and moldy soybean kernels, proteins and lipids undergo degradation more readily during storage, leading to quality deterioration (Yousif, 2014). Identification of damaged soybean kernels is prerequisite and conducive to reduce the infection of healthy kernels to ensure the quality of subsequent product and avoid economic loss.

Commonly used methods for soybean damage detection include morphological analysis, chemical analysis, and imaging techniques (Zhao et al., 2011; Yang et al., 2015; Adão et al., 2017). Morphological analysis requires the operator to be experienced and is susceptible to subjective interference. Chemical analysis, such as chromatography and enzyme-linked immunosorbent assay, owns high accuracy and excellent reproducibility but is a destructive, time-consuming, and labor-intensive process. Imaging techniques, such as red–green–blue (RGB) imaging with high spatial resolution and hyperspectral imaging (HSI) with high spectral resolution, have been popularized in image classification, object detection, and semantic segmentation. However, subtle changes in the internal composition of the kernels are difficult to be perceived by RGB due to insufficient spectral information (Steinbrener et al., 2019).

HSI can simultaneously provide spectral responses and spatial images of hundreds of continuous wavelengths to obtain spectral and external features, thereby enriching the description of soybean kernels (Lu et al., 2020). HSI hardware typically sacrifices spatial resolution to ensure premium spectral resolution due to limited incident energy (Dian et al., 2021). The low spatial resolution leads to weak fidelity of appearance-based features especially when discriminating small objects, such as soybean kernels (Fabiyi et al., 2020). This problem can be solved by multi-modal image fusion, which extracts and combines the most meaningful information from images of different sources to generate a single image that is more informative and beneficial for subsequent applications (Zhang et al., 2021). Thus far, the methods for fusing HSI and RGB images can be broadly divided into multi-scale transformation based on coefficients (Wei et al., 2021), saliency (Muddamsetty et al., 2013), sparse representation (Wei et al., 2015), and deep learning (Wei et al., 2017). Fusion rules in the first three categories are specifically designed in the transform or spatial domain in virtue of transform bases. However, applying the same transformation basis such as wavelet basis (Starck et al., 2001) and ridgelet basis (Chen et al., 2005) to HSI and RGB images may lead to confined fusion performance (Yang and Li, 2012).

In recent years, deep learning-based fusion methods can extract diverse and multi-scale features to achieve adaptive fusion. Notably, the absence of ground truth (GT) images in real scenes leads to the inapplicability of widely used supervised learning models (Zhang et al., 2021). Therefore, an unsupervised method has practical significance (Wang et al., 2020). Proposed an unsupervised and coupled autoencoder (AE) framework implemented by CNNs for super-resolution HSI. However, continuous convolution leads to the loss of information from shallow layers containing low-level features at high spatial resolution, which is unbeneficial for fusion. Dense connections enhance feature propagation and improve information flow by interconnect layers and bypass settings, thereby providing continuous attention of features and preserving the detailed information of HSI and RGB images (Dolz et al., 2018).

Spectral preservation plays an crucial role in fusing HSI and RGB images due to the skewed spectral information that affects the quality of the fused image (Hu et al., 2021). Channel attention is commonly used to assign feature importance by dynamically adjusting the weight of each channel to assist the performance improvement of various task; they can also be used to correct the spectral information in image fusion (Hu et al., 2017).

Although a number of approaches are available for constructing super-resolution hyperspectral image (SR-HSI), few researchers focus on the identification effect of SR-HSI in real environment. End-to-end neural networks use translation invariance and rotation invariance to automatically extract key features without manual feature engineering in image recognition applications (Dhaka et al., 2021; Kundu et al., 2021). However, the high computing and memory requirements hinder the application of complex networks. Lightweight networks, such as MobileNetV2 (Sandler et al., 2018), GhostNet (Han et al., 2020), and ShuffleNet (Ma et al., 2018), which have small parameter and low computation, can achieve good accuracy on resource-constrained devices. In particular, the efficient architecture of ShuffleNet solves the boundary effect problem caused by depth-wise (DW) separable convolution. Convolution optimization including pruning the redundant convolution layer and enlarging the convolution kernel can accelerate the network inference speed and extract richer global features (Luo et al., 2016). The CSP architecture with switching concatenation and transition steps as shortcut operation allows the gradient flow to propagate through different paths of network to enrich the gradient combination and quicken the rate of reasoning (Wang et al., 2020). ShuffleNet can be combined with convolution optimization (CO) and CSP architecture to identify damages to soybean kernels.

Herein, fusion of HSI and RGB images and improved ShuffleNet were proposed to identify soybean kernel damages (Figure 1). First, a super-resolution module based on AE and

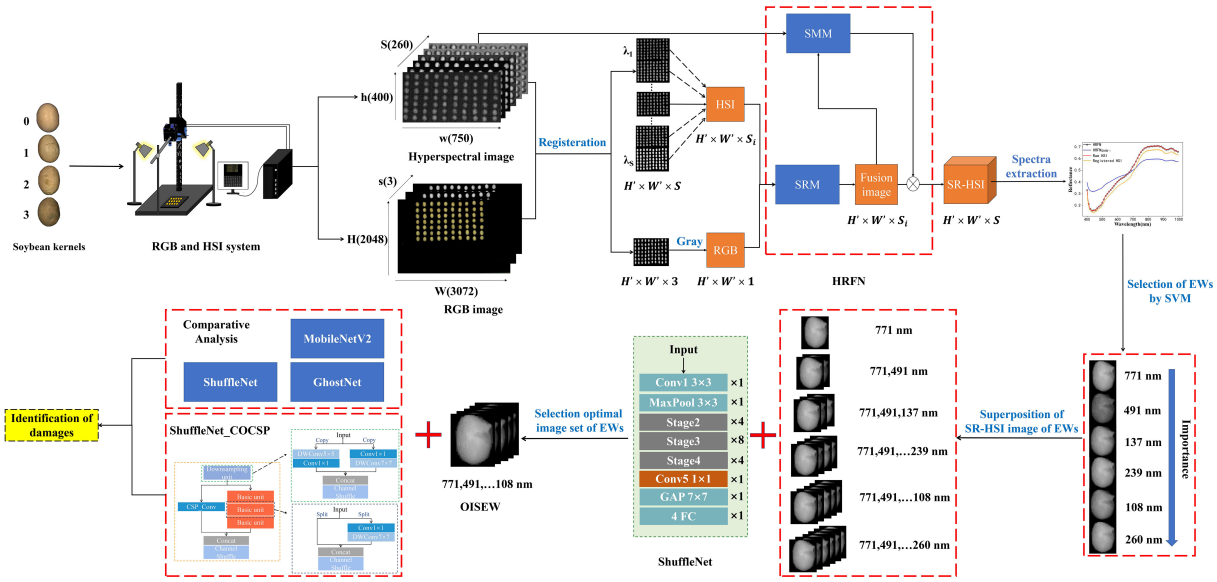


FIGURE 1
Flowchart of identification of soybean kernel damages; Class 0-3: healthy, broken, mildly moldy and severely moldy soybean kernels; SRM is designed by Autoencoder and Dense block.

dense connection and spectroscopy-modification module from the idea of channel attention were designed and integrated to construct a HSI-RGB fusion network (HRFN) and generate SR-HSI images. SVM and ShuffleNet were used to select the SR-HSI monochromatic images of effective wavelengths for rapid identification of soybean kernel damages. Finally, a new identification network architecture, namely, ShuffleNet _COCSP, was developed by combining CO and CSP architectures with ShuffleNet to identify soybean damages with SR-HSI. The main contribution of this study can be summarized as follows.

1. To the best of our knowledge, this study is the first to fuse the HSI and RGB images of small kernels and develop the lightweight network ShuffleNet _COCSP for practical identification of damages.
2. The proposed novel network for HSI and RGB image fusion consists of parallel super-resolution module (SRM) and spectral correction module (SMM).
3. An improved efficient ShuffleNet with convolution optimization and cross-stage partial is proposed for accurate identification of soybean kernel damages.

2 Materials and methods

2.1 Sample preparation

Samples of He13 soybean kernels (Figure S1 in Supplementary Material) were obtained from Shu County Agricultural Management Company. Soybeans with smooth surface were selected and considered as healthy ones. Healthy soybeans were soaked in warm water for 10 min and dried using a drying oven at 100 °C for 2 h to obtain broken soybeans. Moldy soybeans were prepared as follows. 1) Healthy soybeans were soaked in warm water for 10 min and placed in a glass Petri dish. 2) The dish was placed in an incubator with constant temperature of 34°C and humidity of 80% to obtain different degrees of moldy soybeans. 3) Moldy soybeans were collected daily. Mildly moldy soybeans have few spots on the epidermis, whereas severely moldy soybeans have mycelia on the epidermis.

Aflatoxin B1 test strip was used to determine the toxin of moldy soybeans. The lower limit of AFB1 toxin detection was 10 ppb, mildly moldy was in the range of 10–20 ppb, and severely moldy was greater than 20 ppb. For HSI and RGB image acquisition, 2,160 samples were collected, including 560 healthy kernels, 560 broken kernels, 560 mildly moldy kernels, and 480 severely moldy kernels.

2.2 Acquisition and calibration of HSI and RGB images

The image acquisition system of soybean kernels was composed of high-resolution RGB camera and HSI (Figure S2 in Supplementary Material). The industrial camera HIKVISION MV-CA060-11GM with a 12 mm/F2.0 lens was used to collect RGB images at 3072×2048 pixels and save them in BMP format. HSI images were obtained by a visible/NIR HSI system consisting of a Headwall Nano-Hyperspec (Headwall

Photonics Inc., Bolton, MA, USA) push-broom sensor that offers 272 spectral bands, two halogen neodymium lamps (75 W), and a computing unit. For imaging, 70 soybean kernels were placed on a black plate, and the distance between the kernels and lens of the HSI sensor and RGB camera was adjusted to 40 cm. Two halogen lamps were placed on both sides of the lens for illumination. During data acquisition, the RGB industrial camera was set to operate in manual mode with an ISO of 400 and a shutter speed of 16 ms. The parameters of the HSI system were set as follows: exposure time, 70 ms; frame period, 70 ms; and scanning speed, 0.45 deg/s. For calibrating the image, white and dark reference images were acquired by scanning a standard white board with 98% reflectance and covering the lens before collecting HSI images. The correction formula is as follows:

$$I_c = \frac{I_r - I_d}{I_w - I_d} \quad (1)$$

where I_c is the corrected image, $I_r$is a measured raw image of soybean kernels, and $I_w$and I_d are the white and dark reference images, respectively.

2.3 Image preprocessing

The spatial misalignment of source images was caused by the difference between image sensors. In fusion tasks, operations along the spatial pixel positions in deep learning methods are unavailable for real source images due to spatial dislocation (Jiang et al., 2021). As a result, high-precision registration is a key issue in image fusion for constructing SR-HSI datasets. Transformation, rotation, and translation parameters were obtained by perspective deformation to align HSI and RGB images (Arsigny et al., 2005). Specifically, three band images were extracted from the HSI image to form a pseudo-RGB image, which was used as the image to be aligned with the RGB image as the reference image. The region of interest (ROI, rectangle) is selected from the pseudo-RGB image for perspective deformation. Transformation, rotation, and translation parameters were accurately calculated from the ROI vertices. The HSI image was transformed using these parameters to align with the RGB image. The designed registration visualization formula is as follows:

$$h_1(x,y) = \sum_{i=1}^{R} \alpha_i(x,y) \quad (2)$$

$$h_2(x,y) = \sum_{i=1}^{r} \beta_i(x,y) \quad (3)$$

$$f(x,y) = \frac{e^{h_2(x,y)}}{e^{h_1(x,y)} + e^{h_2(x,y)}} \quad (4)$$

where f(x,y) is the registration visualization map, x and y are pixel coordinates, is HSI image, is RGB image, and R and r are the spectral band number of HSI and RGB images, respectively.

Canny operator detects the contour of kernels in RGB. The Otsu's algorithm was used in threshold segmentation to obtain a binary image. The background noise of the RGB image was removed, and the mask obtained from RGB segmentation was transformed to HSI space to remove the background noise of the HSI image. Using the same mask, the samples of HSI and RGB data sets had one-to-one correspondence in the subsequent recognition, and inconsistent phenomenon of sample division did not exist.

2.4 Fusion of HSI and RGB images

In this study, a HSI-RGB fusion network (HRFN) was developed using parallel super-resolution module (SRM) and spectroscopy-modification module (SMM) to solve the problem of low spatial resolution of HSI images. In HRFN (Figure 2), the RGB grayscale image and the monochromatic image of 272 bands of HSI were fused to generate the SR-HSI monochromatic image of corresponding band.

SRM was designed based on AE, a widely-used super-resolution deep learning architecture, and dense block. AE is an unsupervised neural network composed of encoder and decoder and has excellent generalization (Liu et al., 2022). The potential representation of images obtained by encoder has valuable attributes, and the concatenated potential representations of multi-modal images can be reconstructed to a high-quality image by the decoder. For super-resolution in our study, the multi-modal images are the registered HSI and RGB image pair. The encoder is composed of four convolutional layers with the kernel size of 3 × 3 and channel of 16; the decoder contains four convolutional layers with the kernel of 3 × 3. However, the successive convolutions make AE suffer from gradient disappearance and inability to maintain shallow and detailed features, which are critical to obtain excellent super-resolution.

Dense connection is introduced to the encoder of AE. In the encoder, the first convolution is a common convolution, and the last three convolution layers are set as a dense block. Dense block can preserve as much information as possible in encoding by the multi re-utilization of features obtained in the former layers.

The super-resolution operation is performed on the registered image pairs, but the registration may lose some important spectral information. With the idea of attention mechanism, SMM, which is in parallel with SRM, is designed to preserve spectral information of the raw HSI. SMM consists of two global average pooling (GAP) layers and two convolution layers of 1×1. The GAP results of the raw HSI and super-resolution image are cascaded and input to the convolutional layers to obtain the weights that describe the correlation between

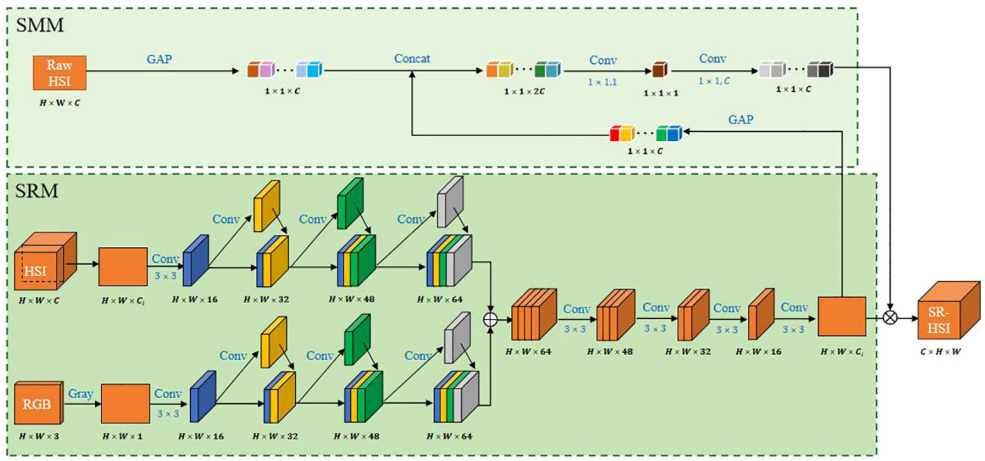


FIGURE 2
Architecture of HRFN.

channels. The super-resolution image in SRM is multiplied by the weights to obtain the final SR-HSI image.

2.5 SR-HSI images of effective wavelengths

The direct use of SR-HSI images containing the images of 272 wavelengths to identify damages to soybean kernels would result in low processing efficiency and high hardware and time costs. The Images of EWs have been proved as a feasible approach to alleviate the limitations in the previous works (Weng et al., 2021). The selection of EWs from the reflectance spectra was based on the performance of SVM models that describe the reflectance of wavelengths and classes of soybean kernels. Specifically, the reflectance of each wavelength and class for the soybean kernels were employed to develop classification models by using SVM. The higher the classification accuracy is, the more important the wavelength will be. The first six wavelengths were selected as EWs, and the SR-HSI images of EWs were sequentially overlaid on ShuffleNet to select the most suitable wavelength combination for determination of damages to soybean kernels.

Before the above operations, spectra were acquired by the following steps. The SR-HSI image with removed background noise was converted into a binary image by graying and converting to color space HSV. The ROI of the sample was extracted, and the reflectance values of all pixels within the ROI were averaged as the reflectance spectra of the soybean samples.

2.6 Recognition model

A deep network with deep architectures possesses powerful feature extraction capability and generally perform well in image tasks. Nevertheless, the high computing and memory requirements of the network hinder its wide application. One approach to solve the problem is the use of a lightweight network. ShuffleNet, which is a powerful lightweight network, can reduce parameters and computation costs by the operation as channels shuffle in the stage layer (Ma et al., 2018).

Specifically, ShuffleNet is composed of convolutional layers, pooling layers, stage layers, and fully connected layers, where the stage is consists of a downsampling unit and a basic unit. These units include DW convolutional layers and 1×1 convolutional layers. However, ShuffleNet replaces a large number of 1×1 point-wise convolutions with channel shuffle to induce the lack of representation ability and slight loss of accuracy. In the stage architecture, convolutional optimization (CO) and cross-stage partial (CSP) architecture were adopted to alleviate the above challenges in this study. The removal of the last convolutional layer and the substitution of the DW convolution kernel size of 3×3 with 7×7 reduce the model parameters, expand the perceptual field, and obtain rich global features (Ding et al., 2022). By replacing all DW convolution 3×3 with 7×7, the padding needs to be changed from 1 to 3, so the resolution of the output feature map remains the same as the original. The CSP architecture firstly divides the feature maps of the downsampling unit into two parts, make them pass through different paths, and concatenate them together in the end of the stage layer. One part passes through the original path, and the other shortcuts directly to the end of the stage. Through the operation, CSP enables richer gradient sets and reduces computation by splitting gradient streams to propagate through different network paths (Wang et al., 2020).

In this study, ShuffleNet_CO was first constructed by removing convolution and expanding the DW kernel in stage layers based on the ShuffleNet framework. ShuffleNet_COCSP (Figure 3) was then developed by introducing the CSP

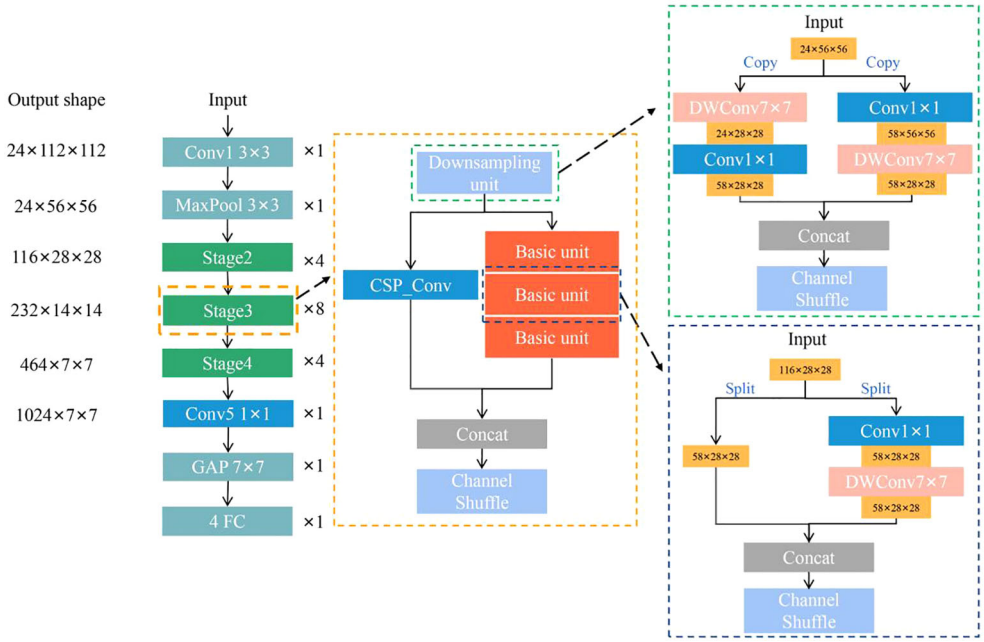


FIGURE 3
Architecture of ShuffleNet_COCSP.

architecture in ShuffleNet_CO. The detailed parameter settings of ShuffleNet_COCSP are shown in Table S1 in Supplementary Material.

2.7 Performance evaluation

Mutual information(MI) (Wells et al., 1996), structural similarity (SSIM) (Wang et al., 2004), peak signal to noise ratio (PSNR) (Feng et al., 2012), and mean absolute differences (MAD) (Cheng et al., 2016)are used in registration. The registration performance increases with increasing values of MI, SSIM, and PSNR and decreasing values of MAD, where the ideal values for MI and SSIM are 1. Higher PSNR and lower MAD indicate better quality of registration. Pixel feature mutual information (FMI_{pixel}) (Haghighat and Razian, 2014), multi-scale structural similarity (MS-SSIM) (Ma et al., 2015), and Nabf (Xydeas and Petrovic, 2000) are used in fusion. The fusion performance increases with increasing values of FMI_{pixel} and MS-SSIM and decreasing values of Nabf.

Healthy, broken, mildly moldy, and severely moldy soybean kernels were divided into a calibration set, a validation set, and a prediction set according to the ratio of 3:1:1. The calibration and validation sets were used for parameter adjustment and preliminary evaluation of the recognition model. Model performance was quantitatively evaluated using accuracy of calibration set (ACC_C), validation set (ACC_V), and prediction set (ACC_P) as well as precision, recall, and F1-score of the prediction set. The evaluation index of the network was the number of floating point operations (FLOPs) and the number of model parameters (Params). Deep networks were constructed based on the PyTorch framework in Python. All the methods were conducted on a computer with an NVIDIA GeForce RTX 3090 GPU.

3 Results and discussion

3.1 Image registration

The registration of HSI and RGB image pairs was performed by perspective deformation, and the performance for healthy soybean kernels is shown in Figure 4. From (Figures 4A, B), the MI and SSIM of the registered image pairs were higher than 0.46 and 0.71 while those of the unregistered image pairs were lower 0.15 and 0.2. The registration operation greatly improves the structural similarity between HSI and RGB images. The PSNR trend of the registered image pairs increased first and then decreased, while the MAD trend was opposite. From the registration visualization (Figures 4E, F), the raw image pairs of HSI and RGB are almost not spatially aligned, but the image pairs are almost perfectly aligned with only minor misalignment at the edge after the registration. Hence, the proposed registration operation satisfactorily solves the spatial dislocation of image pairs.

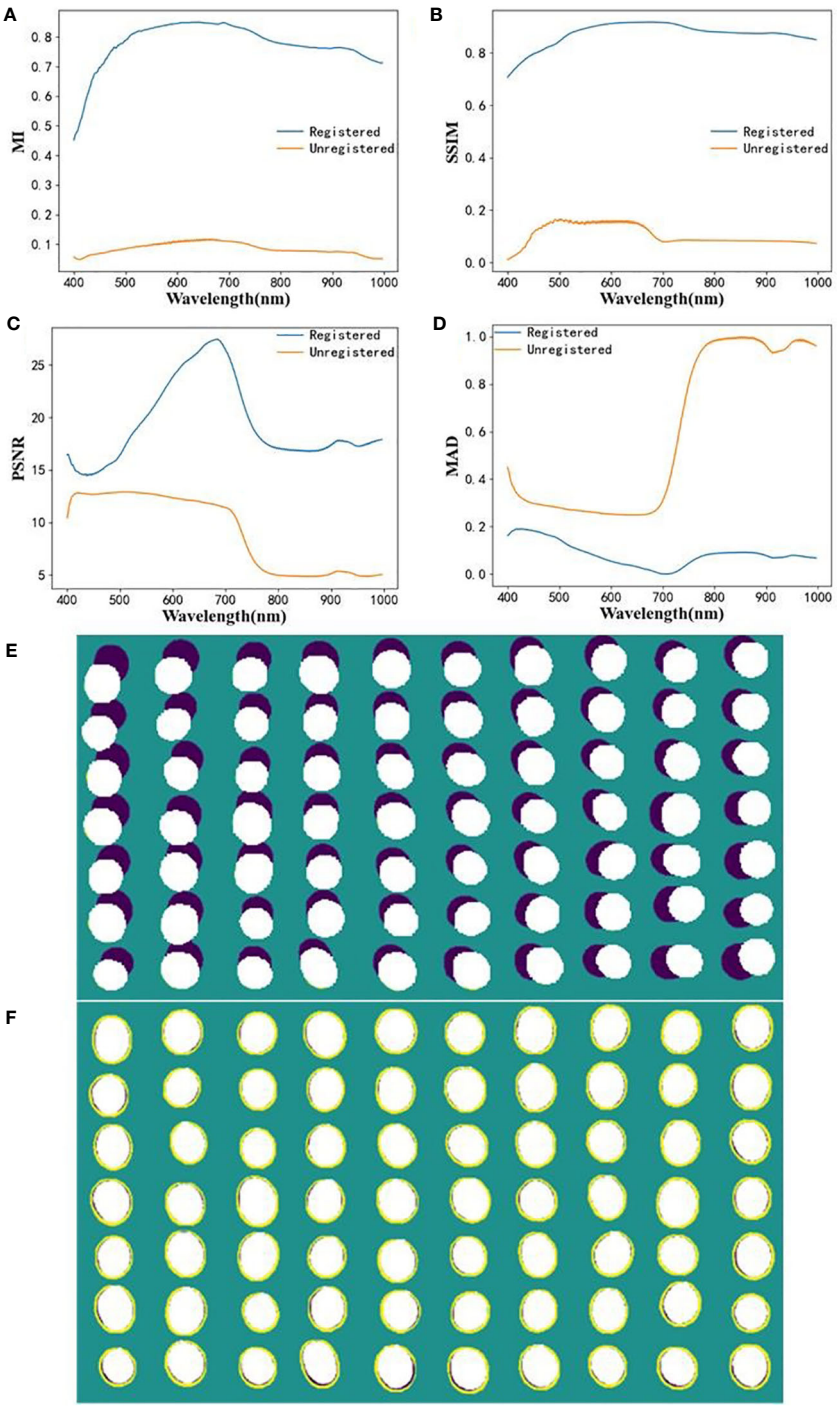


FIGURE 4
Evaluation results of HSI and RGB registration. **(A)** MI; **(B)** SSIM; **(C)** PSNR; **(D)** MAD; **(E, F)** visualization of image pairs before and after registration.

3.2 Fusion of HSI and RGB

HRFN was adopted to fuse the registered HSI and RGB pair to generate SR-HSI (Figure 5). The texture and color of soybean kernels can be clearly observed in the RGB image (Figure 5A), but their details are very blurred in the registered HSI image (Figure 5B). SR-HSI has better spatial resolution than HSI (Figure 5C) and more spectral bands (272) than RGB (3). In simple terms, SR-HSI can be regarded as the spectral resolution improvement of RGB or the spatial resolution enhancement of

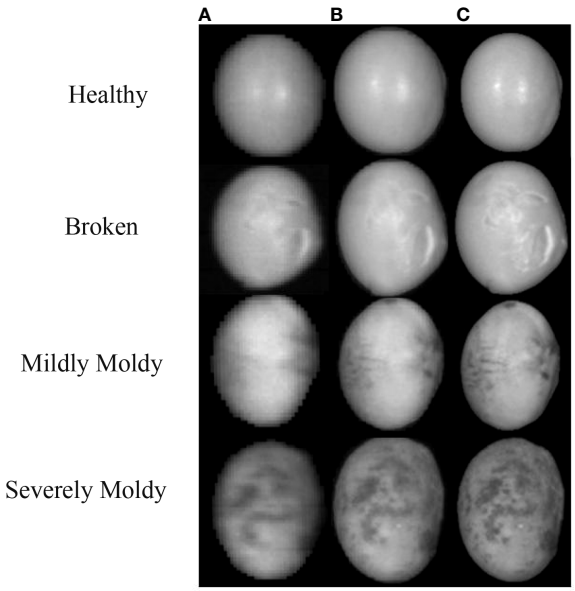


FIGURE 5
Fused results for HSI and RGB images with HRFN. **(A)** The registered HSI image of the 618 nm, **(B)** SR-HSI image of the 618 nm, **(C)** RGB Gray image.

HSI. From the quantitative results of soybean kernels in Table 1, HRFN was highly effective in the fusion for the four classes of soybean kernels, with the FMI_pixel of 0.9415–0.9614, MS_SSIM of 0.9678–0.9880, and Qabf below 0.0508. The FMI_pixel and MS_SSIM of healthy soybean kernels were 0.9488 and 0.9842 higher than those of the three other classes because of the low values of statistical contrast characteristics of structural information for broken or moldy areas of soybean.

The reflectance spectra of SR-HSI almost perfectly overlapped with those of raw HSI, indicating that the SMM module learned the mapping relationship between the HSI and super-resolution image (Figure 5). Thus, HRFN achieves good fusion of HSI and RGB, and the SMM module retains the spectral information from HSI to improve the quality of SR-HSI.

3.3 Selection of image set of EWs

Selecting the key variables of HSI data cube can avoid dimensional disasters and improve the interpretability and generalization ability. Here, the SR-HSI monochromatic images of EWs were extracted, and the optimal image set of EWs (OISEW) was selected according to the classification results of soybean kernel damages. The six EWs selected by SVM are 771, 491, 700, 927, 635, and 973 nm. The SR-HSI monochromatic images of the six EWs were stacked in the above order, and the most remarkable image set was screened based on the performance of ShuffleNet (Table 2). The ShuffleNet parameters are presented in Table S1 in Supplementary Material. For the SR-HSI image of 771 nm, poor results were obtained with ACC_T= 99.12%, ACC_V= 88.94%, and ACC_P= 86.97% mainly because of insufficient information of the image of one wavelength. With the addition of images of other wavelengths, the accuracy first increased and then decreased. The best results of ACC_T= 99.89%, ACC_V= 95.65%, and ACC_P=92.27% were obtained using the SR-HSI image sets of 771, 491, 700, 927, and 635 nm, named as OISEW. The classification performance for severely moldy kernels was improved significantly with precision of 86.14%–97.75%, recall of 91.43%–92.55%, and F1-score of 89.23%–95.08%. The results on mildly moldy kernels were

TABLE 1 Fusion performance of HSI and RGB using HRFN.

Method	Classes	FMI_pixel	MS_SSIM	Nabf
HRFN	Healthy	0.9614	0.9880	0.0422
	Broken	0.9437	0.9678	0.0349
	Mildly moldy	0.9494	0.9781	0.0375
	Severely moldy	0.9415	0.9778	0.0508

TABLE 2 Classification results of soybean kernels using ShuffleNet with SR-HSI images of different combinations of EWs.

SR-HSI image set	Classes	Accuracy (%)	Prediction dataset		
			Precision (%)	Recall (%)	F1-score (%)
771nm	Healthy Broken Mildly moldy Severely moldy	ACC_T=99.12 ACC_V=88.94 ACC_P=86.97	80.81 91.58 89.47 88.14	88.89 87.00 80.19 91.43	84.66 89.23 84.58 89.23
771 and 491 nm	Healthy Broken Mildly moldy Severely moldy	ACC_T=99.38 ACC_V=91.54 ACC_P=90.31	85.57 94.70 90.00 95.56	92.22 89.00 93.40 91.49	88.77 91.53 91.67 93.48
771, 491 and 700 nm	Healthy Broken Mildly moldy Severely moldy	ACC_T=99.67 ACC_V=90.84 ACC_P=90.44	86.46 91.78 96.70 85.85	92.22 89.51 83.02 96.81	89.25 90.37 89.34 91.00
771, 491, 700 and 927nm	Healthy Broken Mildly moldy Severely moldy	ACC_T=99.86 ACC_V=94.12 ACC_P=91.35	86.81 90.82 95.10 91.92	87.78 89.00 91.51 96.81	87.29 89.90 93.27 94.30
771, 491, 700, 927 and 635 nm	Healthy Broken Mildly moldy Severely moldy	ACC_T=99.89 ACC_V=95.65 ACC_P=92.27	87.69 94.44 92.04 97.75	92.22 86.00 98.11 92.55	90.30 90.17 94.98 95.08
771, 491, 700, 927, 635 and 973 nm	Healthy Broken Mildly moldy Severely moldy	ACC_T=99.82 ACC_V=94.65 ACC_P=91.04	80.30 98.85 94.64 94.06	94.64 76.79 94.64 98.96	86.89 86.43 94.64 96.45

similar to those on severely moldy kernels. The classification results of healthy and broken classes were superior to the other image set of EWs, with F1-scores of 90.30% and 90.17%, respectively. Thus, OISEW was adopted to classify the various damages of soybean kernels in subsequent analysis.

3.4 Identification of soybean kernel damages using ShuffleNet_COCSP

The ACC_P=92.27% is insufficient for identification of kernel damages. Thus, ShuffleNet needs to be further optimized considering its efficiency and accuracy. Taking ShuffleNet as backbone, CO and CSP were combined to construct ShuffleNet_COCSP, which was also compared with two widespread lightweight networks, namely, MobileNetV2 and GhostNet. The parameters of each model are shown in Table S1 in Supplementary Material. The identification results of each model are shown in Table 3. MobileNetV2 achieved ACC_P of 95.95%, Params of 2.231 M, and FLOPs of 0.326 G, and GhostNet achieved ACC_P of 95.17%, Params of 4.207 M, and FLOPs of 0.197 G. The identification was satisfactory; however, the point-wise convolutions consume vast and expensive computing resources. ShuffleNet_COCSP obtained the best result with ACC_P of 98.36%, Params of 0.805 M, and FLOP of 0.097 G. The ACC_P of ShuffleNet_COCSP increased by 5.58%, and the params and FLOPs decreased by 36.01% and 37.42%, respectively, compared with those of ShuffleNet. The F1-scores of ShuffleNet_COCSP for mildly and severely moldy kernels were both 100.00%, and all samples of the two classes were accurately classified. The precision levels of healthy and broken classes were 93.81% and 99.03%, and the recall rates were 98.91% and 94.44%, respectively. Most samples of the 2 classes were identified correctly, with only a small part of broken kernels classified mistakenly as healthy.

ShuffleNet_COCSP improved the identification accuracy and vastly reduced the computational effort by enlarging the receptive field and removing the redundant convolution layer and CSP shunting techniques. The curves of accuracy and loss (Figures 6A, B) showed that ShuffleNet_COCSP was better than MobileNetV2, GhostNet, and ShuffleNet, and the fluctuations of the learning curves gradually decreased. In summary, ShuffleNet_COCSP performed well in the identification of soybean kernel damages with excellent accuracy and efficiency.

4 Ablation experiment

In previous studies in the field of fusion of HSI and RGB, the spectral reflectance of a fused super-resolution image was rarely

TABLE 3 Classification results of soybean kernels using MobileNetV2, GhostNet, ShuffleNet and ShuffleNet_COCSP.

Model	Classes	Accuracy (%)	Prediction dataset			Params (M)	FLOPs (G)
			Precision (%)	Recall (%)	F1-score (%)		
MobileNetV2	Healthy Broken Mildly moldy Severely moldy	ACC_T=100.00 ACC_V=97.42 ACC_P=96.82	93.68 97.17 99.08 96.94	96.74 95.37 96.43 98.96	95.19 96.26 97.74 97.94	2.231	0.326
GhostNet	Healthy Broken Mildly moldy Severely moldy	ACC_T=99.78 ACC_V=96.46 ACC_P=95.37	88.35 100.00 96.46 96.88	98.91 88.89 97.32 96.88	93.33 94.12 96.89 96.88	4.207	0.197
ShuffleNet	Healthy Broken Mildly moldy Severely moldy	ACC_T=99.88 ACC_V=94.74 ACC_P=93.16	86.96 94.06 92.50 98.95	86.96 87.96 99.11 97.92	86.96 90.91 95.69 98.43	1.258	0.155
ShuffleNet_COCSP	Healthy Broken Mildly moldy Severely moldy	ACC_T= 99.87 ACC_V= 98.64 ACC_P= 98.36	95.70 97.20 100.00 100.00	96.74 96.30 100.00 100.00	96.22 96.74 100.00 100.00	0.805	0.097

concerned and generally different from that of a raw HSI image. SMM in HRFN was constructed to correct the spectral information and realize good fusion.

HRFN without the SMM module ($HRFN_{SMM-}$) was adopted to fuse the HSI and RGB image pairs and investigate the effect of SMM. The spectral reflectance data of the raw HSI, HRFN, and $HRFN_{SMM-}$ are shown in Figure 5. Compared with those of HRFN, the spectra of $HRFN_{SMM-}$ is nonoverlapping with the raw HSI. Thus, SMM can preserve the spectral information of raw HSI because it can learn the mapping relationship between HSI and hyperspectral super-resolution images to obtain the missing spectral information of each band. Based on the fusion results of $HRFN_{SMM-}$ in Table 4, its FMI_pixel, MS_SSIM, and Nabf are worse than those of HRFN. As a result, SMM does help HRFN focus on missing spectral details to improve the quality of the SR-HSI image (Hu et al., 2021).

ShuffleNet_COCSP combined with CO and CSP achieved the ideal identification and was ultra-lightweight. To further corroborate its effectiveness, we employed ShuffleNet with CO (ShuffleNet_CO) and ShuffleNet with CSP (ShuffleNet_CSP) for developing identification models of soybean kernel damages in SR-HSI images (Table 5). The ACC_P of ShuffleNet_CO and ShuffleNet_CSP increased by 2.79% and 3.46%, the Params decreased by 23.93% and 27.03%, and the FLOPs decreased by 23.26% and 29.68%, respectively, compared with those of ShuffleNet. The results of ShuffleNet_CSP and ShuffleNet_CO were better than that of ShuffleNet and worse than ShuffleNet_COCSP, confirming that CO and CSP played a positive role in recognition. The performance of ShuffleNet_CO is mainly because CO has a large effective receptive field to increase the sensing area of feature maps and extract richer global features. Meanwhile, the redundant 1×1 convolution layer is removed to improve the network efficiency. The Params and FLOPs of ShuffleNet_CSP were greatly reduced because the strategy of truncating the gradient flow was adopted in the CSP architecture; as such, the gradient information will not be reused. Surprisingly, ShuffleNet_CSP has a good accuracy in identifying soybean kernel damages. The reason may be that CSP architecture enhance the variability of the learned features within different layers, thereby greatly improving the learning ability of the network. The advantages of CO and CSP are perfectly combined to make ShuffleNet more efficient and ensure the accuracy of recognition.

5 Discussion

In the application of recognizing agricultural product damages, HSI has been widely used as a mainstream, rapid, and non-destructive measurement method that can provide morphological and compositional information. However, as for the crop kernel of small sizes, the low spatial resolution of HSI leads to weak recognition accuracy (Fabiyi et al., 2020). High spatial resolution is easily obtained from RGB images. In this study, SR-HSI images were generated by fusion of HSI and RGB images to identify soybean kernel damages.

In most image fusion studies based on public datasets, the images from different sources are pre-registered; however, the HSI and RGB image pairs of kernels have the obvious nonlinear appearance differences in our experiments (Zhang et al., 2021). Perspective deformation was used to register image pairs and eliminate spatial dislocation. Thus far, the image fusion methods based on deep learning networks are advantageous

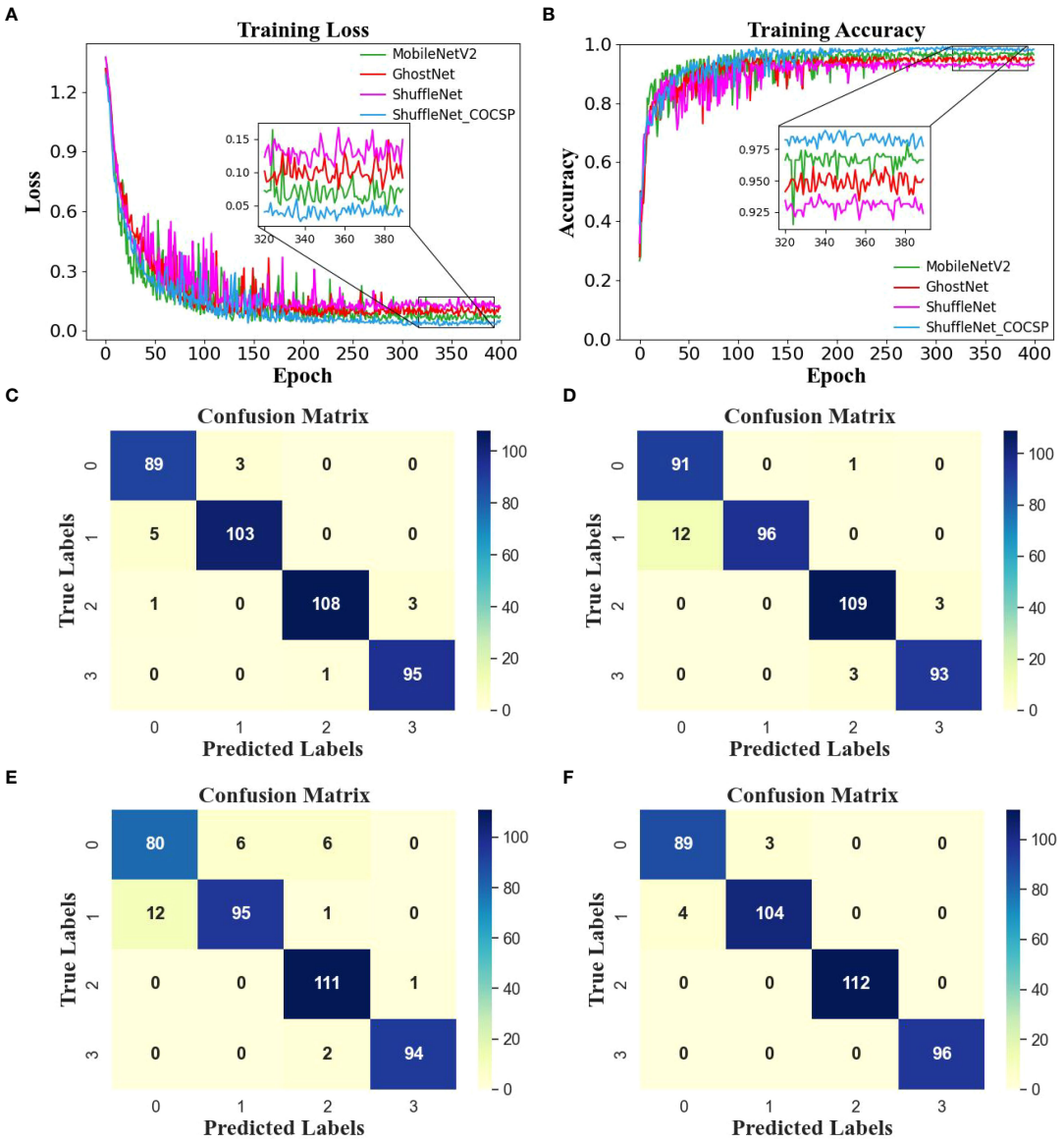


FIGURE 6
(A, B) Learning curves and **(C–F)** confusion matrix of MobileNetV2, GhostNet, ShuffleNet, and ShuffleNet_COCSP in the prediction dataset.

TABLE 4 Quantitative results of fusion HSI and RGB using $HRFN_{SMM-}$.

Method	Classes	FMI_pixel	MS_SSIM	Nabf
$HRFN_{SMM-}$	Healthy	0.9614	0.9865	0.0424
	Broken	0.9435	0.9537	0.0353
	Mildly moldy	0.9491	0.9777	0.0378
	Severely moldy	0.9410	0.9770	0.0509

because the networks can extract the targeted features and achieve adaptive feature fusion. The source image set has no real GT image, so the networks of supervised learning are unapplicable. A network based on unsupervised AE architecture and dense blocks, called as SRM, was constructed to fuse image pairs from HSI and RGB. The above fusion can ensure the spatial quality of images well, but the spectral information of raw HSI is difficult to guarantee. Based on the

TABLE 5 Recognition results based on SHUFFLENET adding CO and CSP .

Model	Classes	Accuracy (%)	Prediction dataset			Params (M)	FLOPs (G)
			Precision (%)	Recall (%)	F1-score (%)		
ShuffleNet_CO	Healthy Broken Mildly moldy Severely moldy	ACC_T=99.93 ACC_V=96.47 ACC_P=95.95	88.89 97.80 98.13 98.92	97.78 89.00 99.06 97.87	93.12 93.19 98.59 98.40	0.957	0.119
ShuffleNet_CSP	Healthy Broken Mildly moldy Severely moldy	ACC_T=99.74 ACC_V=96.47 ACC_P=96.62	87.50 98.96 100.00 100.00	98.91 87.96 100.00 100.00	92.86 93.14 100.00 100.00	0.918	0.109

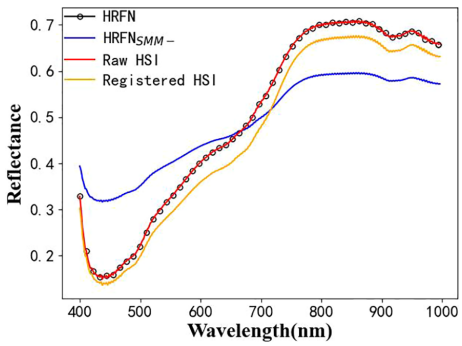


FIGURE 7
Spectral reflectance for the broken class of soybean as obtainedby the raw HSI, registered HSI, and the fused results of HRFN and $HRFN_{SMM-}$.

experiment, the spectral trend of the super-resolution images obtained by SRM was distorted (Figure 7). Referring to the channel attention mechanism, the new branch network was developed and called SMM to extract spectral details by learning the mapping relationship between the HSI and super-resolution images and accomplish spectral correction. By integrating SRM and SMM, HRFN achieved a good fusion of HSI and RGB images, that is, the FMI_pixel, MS_SSIM, and Qabf of the four classes of soybean kernels were 0.9415–0.9614, 0.9678–0.9880, and below 0.0508. The SR-HSI image with high spatial and excellent spectral resolution are expected to provide more accurate results for analysis of soybean damages.

TABLE 6 Classification results of ShuffleNet_COCSP based on HSI and RGB images.

Data	Classes	Accuracy (%)	Prediction dataset		
			Precision (%)	Recall (%)	F1-score (%)
HSI	Healthy Broken Mildly moldy Severely moldy	ACC_T= 99.99 ACC_V= 96.71 ACC_P= 95.82	90.43 93.40 100.00 98.97	92.39 91.67 99.11 100.00	0.91.40 0.92.52 0.99.55 0.99.48
RGB	Healthy Broken Mildly moldy Severely moldy	ACC_T=100.00 ACC_V= 97.64 ACC_P= 96.12	100.00 98.18 92.24 94.90	91.30 100.00 95.54 96.88	95.45 99.08 93.86 95.88

An SR-HSI image of soybean kernels contains images of 272 wavelengths with redundant information, resulting in low processing efficiency and huge modeling cost. Thus, selecting the SR-HSI image of many significant wavelengths is essential to damage identification (Weng et al., 2021). Here, candidate EWs were first selected by SVM models developed with spectral reflectance of each wavelength, and OISEW was finalized by ShuffleNet and the successive superposition of monochromatic images of each EW. The ACC_P=92.27%, Params=1.258 M, and FLOPs=0.155 G of ShuffleNet with OISEW are insufficient to identify damaged kernels in the real world within the limited computational budget. ShuffleNet_COCSP was constructed by adding the CO operation and the CSP architecture into ShuffleNet and obtained ACCp=98.36%, Params=0.805 M, and FLOPs=0.097 G, outperforming ShuffleNet, MobileNetV2, and GhostNet. The increase in the depth and kernel size of convolution extended the effective receptive field and led to enhanced promotion of the networks (Luo et al., 2016). However, the former brings optimization problems. Thus, the CO operation increased the kernel size from 3 × 3 to 7 × 7 and removed the last convolution of 1 × 1 to increase the perceived area of the feature map and extract rich global features. Meanwhile, shortcut is especially vital for networks with large convolution kernels (Ding et al., 2022). CSP with ingenious shortcut operation was induced to reduce the possibility of duplication in information integration and alleviate oversmoothing (Wang et al., 2020), thereby improving the learning ability of the network.

Based on the ablation experiments, the combination of the CO operation and the CSP architecture is better than the single optimization. That is, ShuffleNet_COCSP had ACCp that increased by 2.41% and 1.74%, parameters that decreased by 15.88% and 12.31%, and FLOP that decreased by 18.49% and 11.01% compared with ShuffleNet_CO and ShuffleNet_CSP, respectively. Further, ShuffleNet_COCSP and HSI and RGB images were used to identify soybean kernel damages (Table 6). The ACCp=95.82% of HSI and the ACCp=96.12% of RGB were worse than those of SR-HSI. The SR-HSI images are more discriminative than the HSI images in the subtle information of external features, such as texture and edge, and have wider wavelength perception and stronger diffraction ability than RGB to better identify the internal tissue characteristics of soybean kernels (Sharma et al., 2016).

The damaged soybean kernels identified were accurately analyzed by fusion of HSI and RGB and ShuffleNet_COCSP. However, some aspects need further optimization to obtain better application prospects. The acquisition method was unpractical owning to source images from HSI and RGB cameras in our work. In the future, customized and simplified imaging equipment should be developed to easily obtain EW and RGB images. ShuffleNet_COCSP with small network size and fast recognition speed will be embedded in mobile devices to provide a wide range of application scenarios for intelligent soybean sorting. In contrast to the orderly arrangement of soybeans in this experiment, soybeans in actual sorting equipment have overlapping and adhesion phenomenon. Therefore, image segmentation and image correction in complex background should be considered.

6 Conclusion

In this work, damages to soybean kernels were identified using the improved ShuffleNet and fusion of HSI and RGB. The HSI and RGB image pairs of healthy, broken, and moldy soybean kernels were collected and registered by perspective deformation to eliminate spatial misalignment. HRFN, an unsupervised fusion network, was designed using SRM and SMM in parallel to generate SR-HSI with high spatial resolution and excellent spectral resolution. HRFN achieves a good fusion of HSI and RGB images for the four classes of soybean kernels, with FMI_pixel of 0.9415–0.9614, MS_SSIM of 0.9678–0.9880, Qabf below 0.0508, and perfectly preserved spectral information. Six EWs were selected by SVM, and the OISEW composed of the monochromatic images in 771, 491, 700, 927, and 635 nm was further screened by ShuffleNet. ShuffleNet_COCSP was constructed by adding the CO operation and the CSP architecture into ShuffleNet, and the best result was obtained with ACCp=98.36%, Params=0.805 M, and FLOPs=0.097 G, outperforming MobileNetV2, GhostNet, and the cases of HSI and RGB images. The high-quality SR-HSI images obtained by fusing HSI and RGB images can quickly and accurately identify small kernels, and a customized simplified imaging device can be designed to acquire SR-HSI images with scattered wavelength to meet the practical requirement of damaged kernel identification in the future. The lightweight ShuffleNet_COCSP will be deployed in mobile devices for large-scale detection of damaged kernels and real-time management in the future. In addition, advanced image correction is indispensable due to environmental factors, such as position of imaging devices and motion of samples, causing kernels to overlap one another in the sorting equipment.

Data availability statement

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

Author contributions

LZ: Conceptualization, Methodology, Validation, Funding acquisition, Writing – review & editing. MZ: Conceptualization, Methodology, Writing original draft, Software, Datacuration, Writing – review & editing. JZhu: Conceptualization, review & editing. LH: Visualization, Investigation. JZhao: Supervision.

DL: Supervision. DZ: Supervision. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Natural Science Foundation of China (31701323), The National Key Research and Development Program of China(2019YFE0115200).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1098864/full#supplementary-material

References

Adão, T., Hruška, J., Pádua, L., Bessa, J., Peres, E., Morais, R., et al. (2017). Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry. *Remote Sens.* 9 (11), 1110. doi: 10.3390/rs9111110

Arsigny, V., Pennec, X., and Ayache, N. (2005). Polyrigid and polyaffine transformations: a novel geometrical tool to deal with non-rigid deformations–application to the registration of histological slices. *Med. image Anal.* 9 (6), 507–523. doi: 10.1016/j.media.2005.04.001

Arslan, H., Karakus, M., Hatipoglu, H., Arslan, D., and Bayraktar, O. V. (2018). Assessment of performances of yield and factors affecting the yield in some soybean varieties/lines grown under semi-arid climate conditions. *Appl. Ecol. Environ. Res.* 16 (4), 4289–4298. doi: 10.15666/aeer/1604_42894298

Bessa, J. F. V., Resende, O., Lima, R. R. D., Lopes, M. A. D. S., Rosa, E. S., and Siqueira, V. C. (2021). Storage of soybean with high percentage of grains damaged by bugs in the crop. *Rev. Ceres* 68, 185–193. doi: 10.1590/0034-737x202168030004

Cheng, J., Qiu, W., Jing, Y., Fenster, A., and Chiu, B. (2016). Accurate quantification of local changes for carotid arteries in 3D ultrasound images using convex optimization-based deformable registration. *Med. Imaging 2016: Image Process.* 9784), 1115–1120. doi: 10.1117/12.2217342

Dhaka, V. S., Meena, S. V., Rani, G., Sinwar, D., Kavita,, Ijaz, M. F., et al. (2021). A survey of deep convolutional neural networks applied for prediction of plant leaf diseases. *Sensors* 21, 4749. doi: 10.3390/s21144749

Dian, R., Li, S., Sun, B., and Guo, A. (2021). Recent advances and new guidelines on hyperspectral and multispectral image fusion, inf. *Fusionvol* 69, 40–51. doi: 10.1016/j.inffus.2020.11.001

Ding, X., Zhang, X., Zhou, Y., Han, J., Ding, G., and Sun, J. (2022). "Scaling up your kernels to 31x31: Revisiting large kernel design in cnns. arXiv e-prints. doi: 10.48550/arXiv.2203.06717

Dolz, J., Gopinath, K., Yuan, J., Lombaert, H., Desrosiers, C., and Ayed, I. B. (2018). HyperDense-net: a hyper-densely connected CNN for multi-modal image segmentation. *IEEE Trans. Med. Imaging* 38 (5), 1116–1126. doi: 10.1109/TMI.2018.2878669

Fabiyi, S. D., Vu, H., Tachtatzis, C., Murray, P., Harle, D., Dao, T. K., et al. (2020). Varietal classification of rice seeds using RGB and hyperspectral images. *IEEE Access* 8, 22493–22505. doi: 10.1109/ACCESS.2020.2969847

Haghighat, M., and Razian, M. A. (2014). Fast-FMI: Non-reference image fusion metric. IEEE International Conference on Application of Information & Communication Technologies. IEEE 1–3. doi: 10.1109/ICAICT.2014.7036000

Han, K., Wang, Y., Tian, Q., Guo, J., and Xu, C. (2020). GhostNet: More Features From Cheap Operations. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. doi: 10.1109/CVPR42600.2020.00165

Hu, J. F., Huang, T. Z., Deng, L. J., Jiang, T. X., Vivone, G., and Chanussot, J. (2021). Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks. *IEEE Trans. Neural Networks Learn. Syst.* doi: 10.48550/arXiv.2005.14400

Hu, J., Shen, L., and Sun, G. (2017). Squeeze-and-excitation networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 99. doi: 10.1109/TPAMI.2019.2913372

Jiang, X., Ma, J., Xiao, G., Shao, Z., and Guo, X. (2021). A review of multimodal image matching: Methods and applications. *Inf. Fusion* 73, 22–71. doi: 10.1016/j.inffus.2021.02.012

Kundu, N., Rani, G., Dhaka, V. S., Gupta, K., Nayak, S. C., Verma, S., et al. (2021). IoT and interpretable machine learning based framework for disease prediction in pearl millet. *Sensors* 21, 5386. doi: 10.3390/s21165386

Liu, J., Wu, Z., Xiao, L., and Wu, X. J. (2022). Model inspired autoencoder for unsupervised hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* 60, 1–12. doi: 10.1109/TGRS.2022.3227938

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 (16), 2659. doi: 10.3390/rs12162659

Luo, W., Li, Y., Urtasun, R., and Zemel, R. (2017). Understanding the effective receptive field in deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 29. doi: 10.48550/arXiv.1701.04128

Ma, K., Zeng, K., and Wang, Z. (2015). Perceptual quality assessment for multi-exposure image fusion. *IEEE Trans. Image Process.* 24 (11), 3345–3356. doi: 10.1109/TIP.2015.2442920

Ma, N., Zhang, X., Zheng, H. T., and Sun, J. (2018). ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In: V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, vol 11218. Springer, Cham. doi: 10.1007/978-3-030-01264-9_8

Muddamsetty, S. M., Sidibé, D., Trémeau, A., and Mériaudeau, F. (2013). A performance evaluation of fusion techniques for spatio-temporal saliency detection in dynamic scenes. *IEEE International Conference on Image Processing*. IEEE 3924–3928. doi: 10.1109/ICIP.2013.6738808

Rani, P. R., Chelladurai, V., Jayas, D. S., White, N. D. G., and Kavitha-Abirami, C. V. (2013). Storage studies on pinto beans under different moisture contents and temperature regimes. *J. Stored Products Res.* 52, 78–85. doi: 10.1016/j.jspr.2012.11.003

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 4510–4520. doi: 10.1109/CVPR.2018.00474

Sharma, V., Diba, A., Tuytelaars, T., and Van Gool, L. (2016). Hyperspectral CNN for image classification & band selection, with application to face recognition. *Tech. Rep. KUL/ESAT/PSI/1604 KU Leuven ESAT Leuven Belgium*.

Song, F., Deng, L., Shu, G., Feng, W., and Ji, K. (2012). "A subpixel registration algorithm for low PSNR images. IEEE Fifth International Conference on Advanced Computational Intelligence. IEEE. doi: 10.1109/ICACI.2012.6463241

Starck, J. L., Donoho, D. L., and Candès, E. J. (2001). Very high quality image restoration by combining wavelets and curvelets. *Proc. SPIE* 4478, 9–19. doi: 10.1117/12.449693

Steinbrener, J., Posch, K., and Leitner, R. (2019). Hyperspectral fruit and vegetable classification using convolutional neural networks. *Comput. Electron. Agric.* 162, 364–372. doi: 10.1016/j.compag.2019.04.019

Tao, C., Zhang, J., and Ye, Z. (2005). Remote sensing image fusion based on ridgelet transform. Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International. IEEE. doi: 10.1109/IGARSS.2005.1525320

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. image Process.* 13 (4), 600–612. doi: 10.1109/TIP.2003.819861

Wang, Z., Chen, B., Lu, R., Zhang, H., Liu, H., and Arshney, P. K. V. (2020). FusionNet an unsupervised convolutional variational network for hyperspectral and multispectral image fusion. *IEEE Trans. ImageProcess* 29, 7565–7577. doi: 10.1109/TIP.2020.3004261

Wang, C. Y., Liao, H., Wu, Y. H., Chen, P. Y., and Yeh, I. H. (2020). A New Backbone that can Enhance Learning Capability of CNN. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE. 390–391. doi: 10.1109/CVPRW50498.2020.00203

Wei, Q., Bioucas-Dias, J., Dobigeon, N., and Tourneret, J. Y. (2015). Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Trans. Geosci. Remote Sens.* 53 (7), 3658–3668. doi: 10.1109/TGRS.2014.2381272

Wei, B., Feng, X., and Wang, W. (2021). 3M: A multi-scale and multi-directional method for multi-focus image fusion. *IEEE Access* 9, 48531–48543. doi: 10.1109/ACCESS.2021.3068770

Wei, Y., Yuan, Q., Shen, H., and Zhang, L. (2017). Boosting the accuracy of multispectral image pansharpening by learning a deep residual network. *IEEE Geosci. Remote Sens. Lett.* 14 (10), 1795–1799. doi: 10.1109/LGRS.2017.2736020

Wells, W. M.III, Viola, P., Atsumi, H., Nakajima, S., and Kikinis, R. (1996). Multi-modal volumeregistration by maximization of mutual information, med. *Image Anal.* 1 (1), 35–51. doi: 10.1016/S1361-8415(01)80004-9

Weng, S., Han, K., Chu, Z., Zhu, G., Liu, C., Zhu, Z., 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

Xiao, C. W. (2008). Health effects of soy protein and isoflavones in humans. *J. Nutr.* 138 (6), 1244S–1249S. doi: 10.1093/jn/138.6.1244S

Xydeas, C. S., and Petrovic, V. (2000). Objective image fusion performance measure. *Electron. Lett.* 36 (4), 308–309. doi: 10.1049/el:20000267

Yang, H. C., Haudenshield, J. S., and Hartman, G. L. (2015). Multiplex real-time PCR detection and differentiation of colletotrichum species infecting soybean. *Plant Dis.* 99 (11), 1559–1568. doi: 10.1094/PDIS-11-14-1189-RE

Yang, B., and Li, S. (2012). Pixel-level image fusion with simultaneous orthogonal matching pursuit. *Inf. fusion* 13 (1), 10–19. doi: 10.1016/j.inffus.2010.04.001

Yousif, A. M. (2014). Soybean grain storage adversely affects grain testa color, texture and cooking quality. *J. Food Qual.* 37 (1), 18–28. doi: 10.1111/jfq.12064

Zhang, H., Xu, H., Tian, X., Jiang, J., and Ma, J. (2021). Image fusion meets deep learning: A survey and perspective. *Inf. Fusion* 76, 323–336. doi: 10.1016/j.inffus.2021.06.008

Zhao, T., Wang, Z. T., Branford-White, C. J., Xu, H., and Wang, C. H. (2011). Classification and differentiation of the genus peganum indigenous to China based on chloroplast trnL-f and psbA-trnH sequences and seed coat morphology. *Plant Biol.* 13 (6), 940–947. doi: 10.1111/j.1438-8677.2011.00455.x

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 25 January 2023
DOI 10.3389/fpls.2022.1088499

OPEN ACCESS

EDITED BY
Baohua Zhang,
Nanjing Agricultural University, China

REVIEWED BY
Chu Zhang,
Huzhou University, China
Goutam Kumar Dash,
Centurion University of Technology and Management, India

*CORRESPONDENCE
Guangcheng Shao
sgcln@126.com
Xiyun Jiao
xyjiao@hhu.edu.cn

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

RECEIVED 03 November 2022
ACCEPTED 29 December 2022
PUBLISHED 25 January 2023

CITATION
Wu T, Zhang W, Wu S, Cheng M, Qi L, Shao G and Jiao X (2023) Retrieving rice (*Oryza sativa* L.) net photosynthetic rate from UAV multispectral images based on machine learning methods. *Front. Plant Sci.* 13:1088499. doi: 10.3389/fpls.2022.1088499

Retrieving rice (*Oryza sativa* L.) net photosynthetic rate from UAV multispectral images based on machine learning methods

Tianao Wu[1,2,3], Wei Zhang[1], Shuyu Wu[1,2,3], Minghan Cheng[4], Lushang Qi[1], Guangcheng Shao[1]* and Xiyun Jiao[1,2,3]*

[1]College of Agricultural Science and Engineering, Hohai University, Nanjing, China, [2]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, China, [3]Cooperative Innovation Center for Water Safety and Hydro Science, Hohai University, Nanjing, China, [4]Jiangsu Key Laboratory of Crop Genetics and Physiology/Jiangsu Key Laboratory of Crop Cultivation and Physiology, Agricultural College, Yangzhou University, Yangzhou, China

Photosynthesis is the key physiological activity in the process of crop growth and plays an irreplaceable role in carbon assimilation and yield formation. This study extracted rice (*Oryza sativa* L.) canopy reflectance based on the UAV multispectral images and analyzed the correlation between 25 vegetation indices (VIs), three textural indices (TIs), and net photosynthetic rate (Pn) at different growth stages. Linear regression (LR), support vector regression (SVR), gradient boosting decision tree (GBDT), random forest (RF), and multilayer perceptron neural network (MLP) models were employed for Pn estimation, and the modeling accuracy was compared under the input condition of VIs, VIs combined with TIs, and fusion of VIs and TIs with plant height (PH) and SPAD. The results showed that VIs and TIs generally had the relatively best correlation with Pn at the jointing–booting stage and the number of VIs with significant correlation ($p< 0.05$) was the largest. Therefore, the employed models could achieve the highest overall accuracy [coefficient of determination (R^2) of 0.383–0.938]. However, as the growth stage progressed, the correlation gradually weakened and resulted in accuracy decrease (R^2 of 0.258–0.928 and 0.125–0.863 at the heading–flowering and ripening stages, respectively). Among the tested models, GBDT and RF models could attain the best performance based on only VIs input (with R^2 ranging from 0.863 to 0.938 and from 0.815 to 0.872, respectively). Furthermore, the fusion input of VIs, TIs with PH, and SPAD could more effectively improve the model accuracy (R^2 increased by 0.049–0.249, 0.063–0.470, and 0.113–0.471, respectively, for three growth stages) compared with the input combination of VIs and TIs (R^2 increased by 0.015–0.090, 0.001–0.139, and 0.023–0.114). Therefore, the GBDT and RF model with fused input could be highly recommended for rice Pn estimation and the methods could also provide reference for Pn monitoring and further yield prediction at field scale.

KEYWORDS

UAV multispectral remote sensing, rice canopy, net photosynthetic rate, vegetation index, textural index, machine learning

1 Introduction

Photosynthesis is one of the most crucial parts of the global carbon and energy cycle (Reichstein et al., 2013; A Ivlev, 2017). The crop photosynthesis activities assimilate carbon dioxide (CO_2) and water (H_2O) by using light energy to form organic matter and, therefore, are a key determinant of food production and security (Reich and Amundson, 1985; Long et al., 2006). Net photosynthetic rate (Pn) is the value of the total photosynthetic rate minus the respiration rate, which directly refers to the organic matter accumulated. Although researchers have gradually deepened the understanding of photosynthesis based on cell-scale gas exchange, current methods and equipment developed based on these theories are still mainly focused on the leaf level, which is time-consuming and has a poor representation (Stinziano et al., 2019). It is scientific to use large canopy photosynthesis and transpiration measurement system (CAPTS) (Song et al., 2016) to observe photosynthesis at the canopy scale, but the investment is too expensive to be popularized in regional-scale monitoring.

The mobile high-throughput phenotyping platforms (HTPPs) (Deery et al., 2014; Li et al., 2014) with RGB, fluorescence, hyperspectral, thermal, 3D laser, and computed tomography (CT) imaging sensors provide a non-destructive method for rapid crop phenotypic acquisition. In particular, a high-spectral-resolution spectroradiometer (Aguirre-Gomez et al., 2001; Meroni and Colombo, 2006) (most Fieldspec 4 or 4pro, Analytical Spectral Devices, ASD, Boulder, CO, USA) is the most physical and effective equipment for photosynthesis monitoring on the ground. The sensitive band reflectance or vegetation indices (VIs), generally including 2 or more band reflectance, was commonly used to establish a linear or nonlinear relationship with crop physiological and biochemical parameters. Qiu et al. (2015) comprehensively analyzed the correlation between main photosynthetic, fluorescence parameters and hyperspectral data in ear position leaves of maize and found that Dλ699 had the best correlation with Pn. Sun et al. (2016) introduced wavelet analysis (WA) to select the sensitive bands of hyperspectral for estimating Pn of winter wheat on the leaf scale and found that the models based on WA were more accurate than the VIs method. Fu et al. (2019) constructed a stacking framework for retrieving the maximum carboxylation rate of Rubisco ($V_{c,max}$) and the maximum electron transport rate supporting RuBP regeneration (J_{max}) in the photosynthesis parameters of tobacco based on canopy hyperspectral reflectance, which further improve model accuracy compared with the basic models. Based on the advantages of ground platform on high-resolution continuous spectrum and texture features, the above research could provide practical and accurate estimation of photosynthetic parameters. However, the photosynthetic monitoring in the actual production field could hardly be represented due to environmental factors and the use of various equipment requires expertise.

As a new near-ground remote sensing approach, unmanned aerial vehicles (UAVs) (de Castro et al., 2021) can flexibly provide higher-resolution and bigger-scale images by carrying different sensors (e.g., multispectral, hyperspectral, and thermal infrared cameras). It has already been widely used in the inversion of physiological and biochemical parameters such as plant height (PH) (Che et al., 2020), leaf area index (LAI) (Chen et al., 2022b), nutrient states (Xu et al., 2021), and aboveground biomass (Wang et al., 2022). Equipped with hyperspectral imaging (HSI) sensors, Liu and Peng. (2020) employed eight chlorophyll-related VIs for estimating maximum Pn, and proposed a model based on chlorophyll index (CI) and photosynthetically active radiation (PAR) for different rice varieties. However, the water vapor in the field (especially the paddy field) might have a great influence on the hyperspectral data. Additionally, high price and tedious data processing process (e.g., noise processing, dimension reduction, and spectral unmixing) have prevented the commercial application of this method. Therefore, multispectral sensors that can characterize key points (usually including blue, green, red, red edge, and near-infrared bands) in crop canopy spectra features are more commonly used for practical application on UAV. Chen et al. (2018) established linear inversion models of photosynthetic parameters at different time points in the cotton bud stage based on UAV six-band multispectral images. However, cross-growth stage comparison and regression model selection could be done more comprehensively. Based on the relationship between VIs constructed from UAV multispectral image and photosynthetic parameters, Zhang et al. (2020a) explored the inversion method of diurnal variation of photosynthesis in rice canopy combined with the light response curve model and provided a method with physical basis for gross primary productivity (GPP) inversion, while the scale effect between 100-m UAV multispectral data and PAR monitoring data from a single point on the ground should be further discussed. On the other hand, the image obtained by UAV remote sensing has a higher resolution than satellite remote sensing; thus, it has more detailed texture features that can better reflect the difference in the set window size. Therefore, textural indices (TIs) are commonly introduced with VIs to improve the model accuracy. According to previous studies, TIs have a good correlation with aboveground biomass (Sarker and Nichol, 2011; Liu et al., 2019) and thus also have a good relationship with the accumulated amount of canopy elements (Pimstein et al., 2011; Lu et al., 2019; Zhang et al., 2021) (e.g., nitrogen, potassium, and chlorophyll). Zheng et al. (2019) have found that the normalized difference texture index (NDTI) is in good relationship with rice biomass and the fusion of NDTI with VIs improved the accuracy of biomass estimation. Similarly, Lu et al. (2021) and Zheng et al. (2020) demonstrated that the fusion of TIs and multispectral VIs could effectively improve the estimation of potassium accumulation and nitrogen accumulation in rice. Since the accumulated organic matter of photosynthesis can directly affect the basic growth indicators of rice such as plant height, tiller number, and leaf area index, TIs could also have untapped potential in Pn estimation.

For models employed in the inversion studies, linear regression or nonlinear regression were commonly used to construct inverse functions with definite expressions, but the accuracy is relatively low and poor in portability (Wan et al., 2021). Machine learning methods have been widely used in the regression and classification issues and have been proven to be fast, accurate, and good at generalization. WA (Bruce et al., 2002), partial least square regression (PLSR) (Fu et al., 2022), and least absolute shrinkage and selection operator (LASSO) (Yang and Bao, 2017) are usually used in HSI studies to reduce the high-dimension hyperspectral data to a few important components that are sensitive to the target parameters.

Artificial neural network (ANN) (Liang et al., 2015), kernel-based support vector machine regression (SVR) (Liang et al., 2016), and random forest (RF) (Cheng et al., 2022) regression methods have been most wildly employed to explore and fit the nonlinear relationship between reflectance or VIs and inversion objects. Other machine learning methods based on ANN, kernel function, and tree also have great potential in this issue. Yang et al. (2022) built a Bayesian neural network (BNN) model to predict potential maximum quantum yield (Fv/Fm) and two other chlorophyll fluorometer parameters of grape by quantifying the HSI response indices of photosynthetic pigments and water status parameters. Yuan et al. (2022) simulated the maximum carboxylation rate at 25°C (V_{m25}) of crops over time based on the convolutional neural network (CNN) model combining flux and satellite remote sensing data to further improve the estimation accuracy of GPP. Based on the leaf phenotype data, Zhang et al. (2020b) established poplar Pn estimating models using the extreme gradient boosting model (XGBoost). (Fu et al., 2020, 2022) have also proven the good performance of machine learning models based on the rich feature input of the HSI data for photosynthetic parameter estimation. However, there is a lack of understanding of machine learning methods for photosynthesis parameters estimating with less reflectance features based on the multispectral data and less research on rice.

In this study, multispectral images of rice canopy were acquired by UAV, and the responses of multispectral reflectance features together with Pn, PH, and SPAD to the different nitrogen or leakage treatments were analyzed at different growth stages. The correlation between Pn, VIs, and TIs extracted from the multispectral reflectance was compared, and the VIs with relatively significant correlations were employed as input of the five machine learning models. Model performance comparison under different input combinations was performed, and the improvement of fusing TIs and basal growth index PH and SPAD was further analyzed. The final purpose is to explore an economical and accurate method at field scale for the estimation of Pn and photosynthesis stress detection during the whole growth season of rice.

2 Materials and methods

2.1 Study area

The experiment was conducted at the Jiangning Campus of Hohai University, Nanjing City, Jiangsu Province in China (31°54'57" N, 118° 46'37" E). A total of 22 plots were set in this study, with a length × width of 2.5 m × 2.0 m and a depth of 2.0 m, which were cultivated with a rice–wheat rotation for many years. Rice cultivar (Nanjing-9108) was transplanted on 4 July with a spacing of 20 cm × 15 cm and harvested on 25 October 2021 under the controlled irrigation and drainage scheme. In order to obtain various spectral characteristics and photosynthetic characteristics parameters of rice canopy at different stages, five nitrogen fertilizer levels (N1–N5: 0, 150, 225, 300 and 375 kg/ha total pure nitrogen) and two infiltration levels (W1 and W2: 3 and 5 mm/day) were applied. The above two-factor complete experimental scheme was used for randomized design within the 22 plots. N fertilizers were employed as the base fertilizer (5 July), tiller initiation fertilizer (14 July), and spikelet-developing fertilizer (14 August) with the proportion of 40%, 30%, and 30% of total pure nitrogen, respectively. Phosphate (P) and potassium (K) fertilizers were applied once as the base fertilizer. All plots were well managed with practices commonly adopted by local farmers. The basic properties of the test soils are listed in Table 1 (Chen et al., 2022a) and the location of the experimental area and the arrangement of experimental treatments are shown in Figure 1.

2.2 UAV based multispectral data acquisition and processing

A DJI Innovation's Phantom4-M (P4M) was employed as the phenotyping platform in this study. It is equipped with a multispectral camera with six CMOS, including one color sensor for visible light imaging (RGB) and five monochrome sensors for multispectral imaging. Each sensor has an effective pixel of 2.08 million, a lens field angle of 62.7°, and a focal length of 5.74 mm. Specific parameters of the sensor are shown in Table 2. The UAV-based multispectral image data were obtained under clear and cloudless weather conditions (10:00–14:00) at each rice growth stage. The UAV flew at an altitude of 15 m, with a heading overlap of 85% and a sideway overlap of 75%.

The multispectral original images of five bands acquired by each UAV flight sortie were exported into the PIE-UAV software (Piesat Information Technology Co., Ltd., China) to correct and splice into field orthophoto. The production steps of the orthophoto were as follows: (1) image matching: match the original images with 40,000 key and tie point limits by geographical location matching method; (2) image aligning: import ground control point (GCP) information and align the images with high adjustment accuracy, 0.05 pixel GCP measurement accuracy, and 0.5 pixel connection point matching accuracy; (3) DEM building: generate DEM data using a resolution of 1 GSD; (4) tessellation building: generate the tessellation line based on the Voronoi Geometry method; (5) orthophoto correction: correct

TABLE 1 Basic soil properties of different layers.

Soil layer (cm)	Soil particle fraction (%)			Bulk density (g cm^{-3})	Organic matter (%)	pH (H_2O)
	Sand	Silt	Clay			
0–20	40.21 ± 9.06	38.22 ± 6.43	21.57 ± 3.26	1.36 ± 0.23	1.24 ± 0.06	6.94 ± 0.06
20–40	39.12 ± 6.31	39.16 ± 4.71	21.72 ± 2.63	1.40 ± 0.19	1.35 ± 0.06	6.97 ± 0.05
40–60	38.87 ± 5.46	39.86 ± 4.06	21.27 ± 2.83	1.43 ± 0.20	1.20 ± 0.08	6.85 ± 0.07
60–160	40.25 ± 5.02	38.12 ± 3.72	21.63 ± 2.41	1.48 ± 0.21	/	6.80 ± 0.04

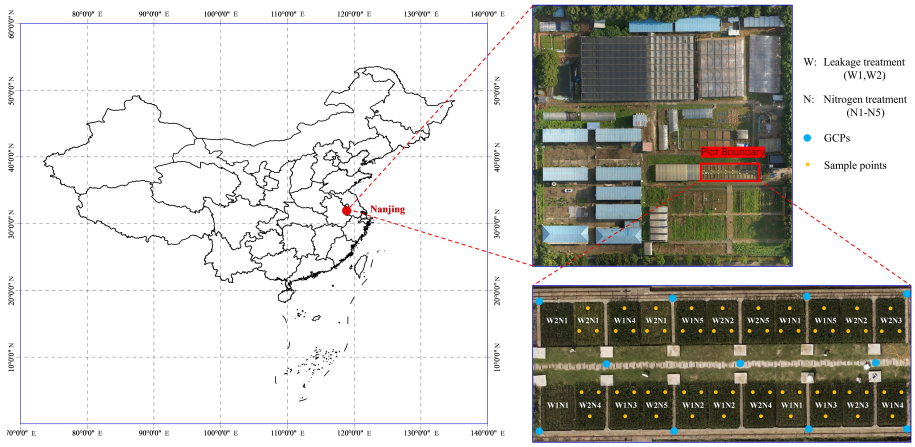


FIGURE 1
Study area and experiment treatments. W represents the leakage treatments (including W1: 3mm/day and W2: 5mm/day); N represents the nitrogen treatments (including N1-N5: 0, 150, 225, 300 and 375 kg/ha total pure nitrogen, respectively); GCP is abbreviation of gourd control points for geometric correction; Ground measurements in each sample point were averaged from 3 representative plants.

the orthophoto with automatically calculated image resolution and the mosaic line mask method; (6) color balancing: set the number of pyramid layers to 3 for color homogenization of mosaic images; and (7) image mosaicking: resample the orthophoto by the cubic convolution method and export into Geo-Tiff format. The final size and resolution of the orthophoto were 9,012 × 5,126 pixels and 7.25 mm/pixel, respectively. ENVI 5.3 was used to perform layer stacking on Tiff images of each band to obtain five-band multispectral images, and the digital numbers (DNs) were transformed into reflectance by radiometric correction.

2.3 Field data collection

Simultaneous field measurements were conducted within the same day of the UAV multispectral image data acquisition, including rice PH, SPAD, and photosynthetic parameters Pn, Tr, and Gs. The PH values were measured with a soft ruler from the soil ground to the leaf tip (cm). The SPAD values were measured by the chlorophyll meter model (SPAD-502, Spectrum Technologies, Inc., NE, USA) and averaged from the measurements at the tip, middle, and base of each leaf. The photosynthetic parameters were measured by the portable photosynthesis system (LI-6800, LI-COR Inc., NE, USA) at 10:00–11:30 a.m. The measured leaf position was the middle of the latest fully unfolded leaf at the jointing–booting stage, and the middle of the panicle leaf at heading–flowering and ripening stages. Each parameter was averaged from three representative plants within a 30 cm × 30 cm quadrat, and three quadrats were measured for each plot. Thus, 60 groups of field data were obtained for each growth stage and 180 groups in the total growth season. The details of ground measurements and UAV flights are listed in Table 3.

It should be noted that the weather conditions in early September 2021 were mainly cloudy and rainy; the measurement on 3 September was the only relatively ideal condition. Therefore, the ground SPAD and Pn measurement at the heading–flowering stage and photography of UAV were affected to a certain extent.

TABLE 2 Multispectral camera sensor parameters.

Band name	Abbreviations	Center wavelength (nm)	Band width (nm)	Resolution (pixels)
Blue	B	450	16	1,600 × 1,300
Green	G	560	16	1,600 × 1,300
Red	R	650	16	1,600 × 1,300
Red edge	RE	730	16	1,600 × 1,300
Near infrared	NIR	840	26	1,600 × 1,300

2.4 Vegetation index and textural index calculation

2.4.1 Vegetation index calculation

VI is established by the linear or nonlinear combination of different spectral band reflectances, which is a common method to retrieve physiological and biochemical indicators of crops (Zeng et al., 2022). A set of 25 commonly used VIs were employed in this study to investigate the relationship between VIs and rice photosynthetic parameters. Threshold processing was firstly performed on the stacked multispectral image to eliminate the influence of water on the reflectance. The canopy reflectance of each band within the 30 cm × 30 cm region of interest (ROI) was then averaged to calculate the VIs of each plot. The involved VIs and formulas are listed in Table 4.

2.4.2 Textural index calculation

Gray-level cooccurrence matrix (GLCM) (Haralick et al., 1973) was applied in this study to extract eight texture features from each band in the stacked image, including mean (MEAN), variance (VAR), homogeneity (HOM), contrast (CON), dissimilarity (DIS), entropy (ENT), second moment (SEC), and correlation (COR), and a total of 40 texture features (with a 3 × 3 pixel window size) were obtained.

TABLE 3 Ground and UAV data acquisition details.

Date	Growth stage	Temperature	Wind speed	Ground measurements	UAV data acquisition		
		(°C)	(m/s)		Time	Height (m)	Resolution (mm/pixel)
17 August 2021	Jointing–booting	30.38	1.90	PH, SPAD, Pn	10:45 a.m.	15.00	7.25
3 September 2021	Heading–flowering	31.08	4.30	PH, SPAD, Pn	11:10 a.m.	15.00	7.25
21 September 2021	Ripening	30.05	2.60	PH, SPAD, Pn	11:00 a.m.	15.00	7.25

The same ROI size with VIs was used to extract texture features and the average value was taken. The extracted GLCM texture features were numbered in the order of MEAN, VAR, HOMO, CON, DIS, ENT, SEC, and COR of each band (in the order of band 1 to band 5) from 1 to 40. The normalized difference textural index (NDTI), difference textural index (DTI), and renormalized difference textural index (RDTI) were selected to construct TI involving two different texture features. The TI formulas are as follows:

$$\mathrm{NDTI} = \frac{T_1 - T_2}{T_1 + T_2}$$

$$\mathrm{DTI} = T_1 - T_2$$

$$\mathrm{RDTI} = \frac{T_1 - \mathrm{T}_2}{\sqrt{T_1 + T_2}}.$$

where T_1 and T_2 represented two random different texture features.

2.5 Modeling and validation

2.5.1 Machine learning regression methods

Linear regression (LR), support vector regression (SVR), gradient boosting decision tree (GBDT), random forest (RF), and multilayer perceptron neural network (MLP) were employed in this study for Pn estimation. The *gridsearch* tuning results for the hyperparameters of each model are listed in Table 5, where the unmentioned hyperparameters were the default values.

(1) Linear regression: LR is a traditional algorithm based on classical statistics, which is the most commonly used model in the spectral inversion research because of its simple construction form and strong interpretation. Combined with the correlation analysis, the relationship between variables and target parameters can be directly reflected. In this study, the LR model with the ordinary least squares method was used for Pn multiple regression.

(2) Support vector regression: SVR is an important application branch of support vector machine (SVM), which seeks the optimal hyperplane by minimizing the total deviation of all sample points from the hyperplane (Cortes and Vapnik, 1995). Unlike ordinary least squares, the SVR model sets a threshold ϵ around the regression line such that all data points within ϵ are not penalized for their errors. Kernel function, gamma, and C are crucial parameters in the SVR model and have been tuned through the *gridsearch* method in the *sklearn* package.

(3) Gradient boosting decision tree: GBDT is an iterative decision tree algorithm with a "boosting" ensemble learning method (Friedman, 2001; Wu et al., 2020). The basic learners [usually classification and regression tree (CART)] in the GBDT model have strong dependencies between each other and are trained by progressive iterations based on the residuals. The results of all basic learners are added together as the final output, which grant GBDT great advantages in overfitting and computational cost fields and reduce bias at the same time.

(4) Random forest: RF is one of the most popular tree algorithms proposed by Breiman (2001) based on the bagging idea of ensemble learning. RF applies the "bootstrap" method to retrieve samples to train the N basic learners (usually CART) in parallel without dependence. The final output of the RF model is derived by combining results of the basic models with the "voting" method, which makes the RF model insensitive to outlier variable values.

(5) Multilayer perceptron neural network: MLP is generally composed of a fully connected input layer, a hidden layer, and an output layer, in which the hidden layer can be multiple (Khoshhal and Mokarram, 2012). As the most basic form of feed-forward neural network, the MLP model has been widely applied in the analysis of various complex problems and is also the foundation of CNN, deep neural network (DNN), and other complex neural networks. A typical three-layer MLP model was used in this study and parameters were well tuned.

2.5.2 Model validation and evaluation

Sixty groups of multispectral data and field measured data in each growth stage were divided into training and validation sets by the 10-fold cross-validation method. Each time, 90% and 10% of the data were employed as training and validation sets, respectively; this process continues 10 times until all the samples have been predicted once and only once. The model final performance was averaged from the evaluation criteria in the cross-validation. To comprehensively evaluate the model performance of Pn estimation, the mean square error (MSE), mean absolute error (MAE), explained variance score (EVS), and coefficient of determination (R^2) were considered in this study as the evaluation criteria. All model code

TABLE 4 Vegetation indices and formula examined in this study.

Vegetation index	Abbreviations and formula
NIRv(Zeng et al., 2019)	$NIRv = b_{NIR}\frac{b_{NIR}-b_R}{b_{NIR}+b_R}$
Chlorophyll Index Green(Gitelson et al., 2005)	$CI_{green} = \frac{b_{NIR}}{b_G} - 1$
Chlorophyll Index Red Edge (Gitelson et al., 2005)	$CI_{red\ edge} = \frac{b_{NIR}}{b_{RE}} - 1$
Chlorophyll Vegetation Index (Gitelson et al., 2003)	$CVI = \frac{b_{NIR}\times b_R}{b_G^2}$
Difference Vegetation Index (Vincini et al., 2008)	$DVI=b_{NIR}-b_R$
Enhanced Vegetation Index (Jordan, 1969)	$EVI = 2.5\frac{b_{NIR}-b_R}{b_{NIR}+6b_R-7.5b_B+1}$
Greenness Index (Huete et al., 2002)	$GI = \frac{b_G}{b_R}$
Green Normalized Difference Vegetation (Smith et al., 1995)	$GNDVI = \frac{b_G-b_R}{b_G+b_R}$
Modified Chlorophyll Absorption in Reflectance Index (Gitelson et al., 1996)	$MCARI = (b_{RE}-b_R) - 0.2\frac{(b_{RE}-b_G)}{b_{RE}/b_R}$
Modified Nonlinear Vegetation Index (Daughtry et al., 2000)	$MNVI = \frac{1.5(b_{NIR}^2-b_R)}{b_{NIR}^2+b_R+0.5}$
Modified Soil Adjusted Vegetation Index (Gong et al., 2003)	$MSAVI = \frac{b_{NIR}+1-}{0.5\sqrt{(2b_{NIR}+1)^2-8(b_{NIR}-b_R)}}$
Modified Simple Ratio (Goel and Qin, 1994)	$MSR = \frac{b_{NIR}/b_R-1}{\sqrt{b_{NIR}/b_R+1}}$
MERIS Terrestrial Chlorophyll Index (Chen, 1996)	$MTCI = \frac{b_{NIR}-b_{RE}}{b_{RE}-b_R}$
Modified Triangular Vegetation Index (Dash and Curran, 2004)	$MTVI = \frac{1.5(1.2(b_{NIR}-b_G)-2.5(b_R-b_G))}{\sqrt{(2b_{NIR}+1)^2-(b_{NIR}-5\sqrt{b_R})}-0.5}$
Nonlinear Vegetation Index (Haboudane et al., 2004)	$NLI = \frac{b_{NIR}^2-b_R}{b_{NIR}^2+b_R}$
Normalized Difference Vegetation Index (Tucker et al., 1979)	$NDVI = \frac{b_{NIR}-b_R}{b_{NIR}+b_R}$
Optimization of Soil-Adjusted Vegetation Index (Rondeaux et al., 1996)	$OSAVI = \frac{1.16(b_{NIR}-b_R)}{b_{NIR}+b_R+0.16}$
Renormalized Difference Vegetation Index (Roujean and Breon, 1995)	$RDVI = \frac{b_{NIR}-b_R}{\sqrt{b_{NIR}+b_R}}$
Ratio Vegetation Index 1 (Birth and McVey, 1968)	$RVI1 = \frac{b_{NIR}}{b_R}$
Ratio Vegetation Index 2 (Xue et al., 2004)	$RVI12 = \frac{b_{NIR}}{b_G}$
Structure Intensive Pigment Index (Blackburn, 1998)	$SIPI = \frac{b_{NIR}-b_B}{b_{NIR}+b_B}$
Transformed Chlorophyll Absorption in Reflectance Index (Haboudane et al., 2002)	$TCARI = 3((b_{RE}-b_R) - 0.2(b_{RE}-b_G)\frac{b_{RE}}{b_R})$
Triangular Vegetation Index (Broge and Leblanc, 2001)	$TVI=60(b_{RE}-b_G)-100(b_R-b_G)$
Visible Atmospherically Resistant Index (Gitelson et al., 2002)	$VARI = \frac{b_G-b_R}{b_G+b_R-b_B}$
Visible Difference Vegetation Index (Zhang et al., 2019)	$VDVI = \frac{2b_G-b_R-b_B}{2b_G+b_R+b_B}$

b_B, b_G, b_R, b_{RE}, and b_{NIR} represent blue (450 ± 16 nm), green (560 ± 16 nm), red (650 ± 16 nm), red edge (730 ± 16 nm), and near-infrared (840 ± 26 nm) band reflectance, respectively.

TABLE 5 Tuned hyperparameters of models employed in this study.

Model	Tuned hyperparameters
SVR	kernel='rbf'; gamma= 'auto'; C=1.0.
GBDT	n_estimators=100;max_features='none';max_depth='adaptive'.
RF	n_estimators=100;max_features='none';max_depth='adaptive'.
MLP	hidden_layer_sizes= (100),; activation='relu'; solver='lbfgs'; learning_rate=0.001.

LR, SVR, GBDT, RF, and MLP represent linear regression, support vector regression, gradient boosting decision tree, random forest, and multilayer perceptron neural network models, respectively.

and evaluation criteria calculations were written in Python3.2 and implemented in a laptop with Intel Core i7-9750H CPU @2.60 GHz, NVDIA GeForce GTX 1660 Ti GPU, and 16 GB of RAM.

$$\mathbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$\mathbf{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$\mathbf{EVS} = 1 - \frac{\mathbf{VAR}(y_i - \hat{y}_i)}{\mathbf{VAR}(y_i)}$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where y_i, $\hat{y}_i$ and $\bar{y}$epresent the measured value, the mean measured value, and the estimated value, respectively. n represents the number of the results. VAR represents the variance of the results. MSE and MAE are in the same unit with the measured value, ranging from 0 (optimum value) to +∞ (worst value). EVS and R^2 are dimensionless, ranging from 0 (worst value) to 1 (optimum value).

3 Results

3.1 Rice photosynthetic traits and canopy multispectral feature response to different treatments

3.1.1 Rice plant height, SPAD, and net photosynthetic rate

The PH, SPAD, and Pn of tested rice at the jointing–booting, heading–flowering, and ripening stage are shown in Figure 2, respectively. PH increased obviously with the increase of nitrogen (N) application and advancement of growth stage, while it decreased slightly with ear filling at the ripening stage (Figure 2A). When PH reached the highest at the heading–flowering stage, the average rice PH with N2–N5 level under W1 (low leakage) treatment was 10.11%, 14.51%, 18.53%, and 21.23% higher than that under the N1 level, respectively, and 18.48%, 15.59%, 15.47%, and 15.59% higher than the N1 level, respectively, for W2 treatment. PH under W2 (high leakage) treatment was significantly higher than those under W1 treatment at N1 and N2 levels in the jointing–booting and heading–flowering stage, respectively, but the difference was not obvious under other N applications.

As shown in Figure 2B, SPAD generally showed a trend of initially increasing (N1–N3) then decreasing (N3–N4) and finally ending up with a small increase (N4–N5) with the increase of nitrogen application under the same leakage conditions. The maximum value of SPAD in each growth stage almost appeared at the N3 level, while the lowest value was found at the N1 level without N application. The SPAD value reached the maximum at the heading–flowering stage, and the average SPAD values of N2–N5 levels were 12.27%, 8.30%, 1.62%, and 10.29% higher than the N1 level for W1 leakage treatment and 1.44%, 3.60%, 0.36%, and 1.08% higher than the N1 level for W2 leakage treatment, respectively. For the same N application level, W1 leakage treatment could increase the SPAD value of N1–N5 levels by −0.18%, 10.48%, 4.35%, 1.08%, and 8.91%, respectively, compared with W2.

It can be seen from Figure 2C that the Pn of rice decreased with the advancement of the rice growth stage. Pn under different N treatments showed a generally similar change trend with SPAD, increasing with the increase of N application at N1 to N3 levels and reaching the maximum at N3, decreasing at N4, and reverting at the N5 level [the Pn increase from N4 to N5 is not significant ($p > 0.05$)]. At the jointing–booting stage when photosynthesis was most vigorous, the average Pn values under N2–N5 levels with W1 treatment were 9.78%, 21.74%, 8.69%, and 13.04% higher than the N1 level, respectively, and 16.36%, 22.87%, 9.72%, and 17.16% higher than the N1 level under the conditions of W2 treatment, respectively, which indicated that excessive application of N fertilizer might inhibit photosynthesis. Under the same N application level, the Pn with W1 treatment was slightly higher than that with W2 treatment, indicating that low leakage intensity could promote leaf photosynthesis to a certain extent.

3.1.2 Rice canopy multispectral reflectance characteristics

Figure 3 illustrates the rice canopy multispectral reflectance of blue (band 1), green (band 2), red (band 3), red edge (band 4), and near infrared (band 5) with different treatments at three growth stages. In general, the average reflectance value of all five bands decreased as the growth stage progressed. As one of the most representative features in the crop spectral curve, the band 5 reflectance value ranged from 0.337 to 0.465 at the jointing–booting stage, while it decreased slightly to 0.327–0.449 at the heading–

10.3389/fpls.2022.1088499

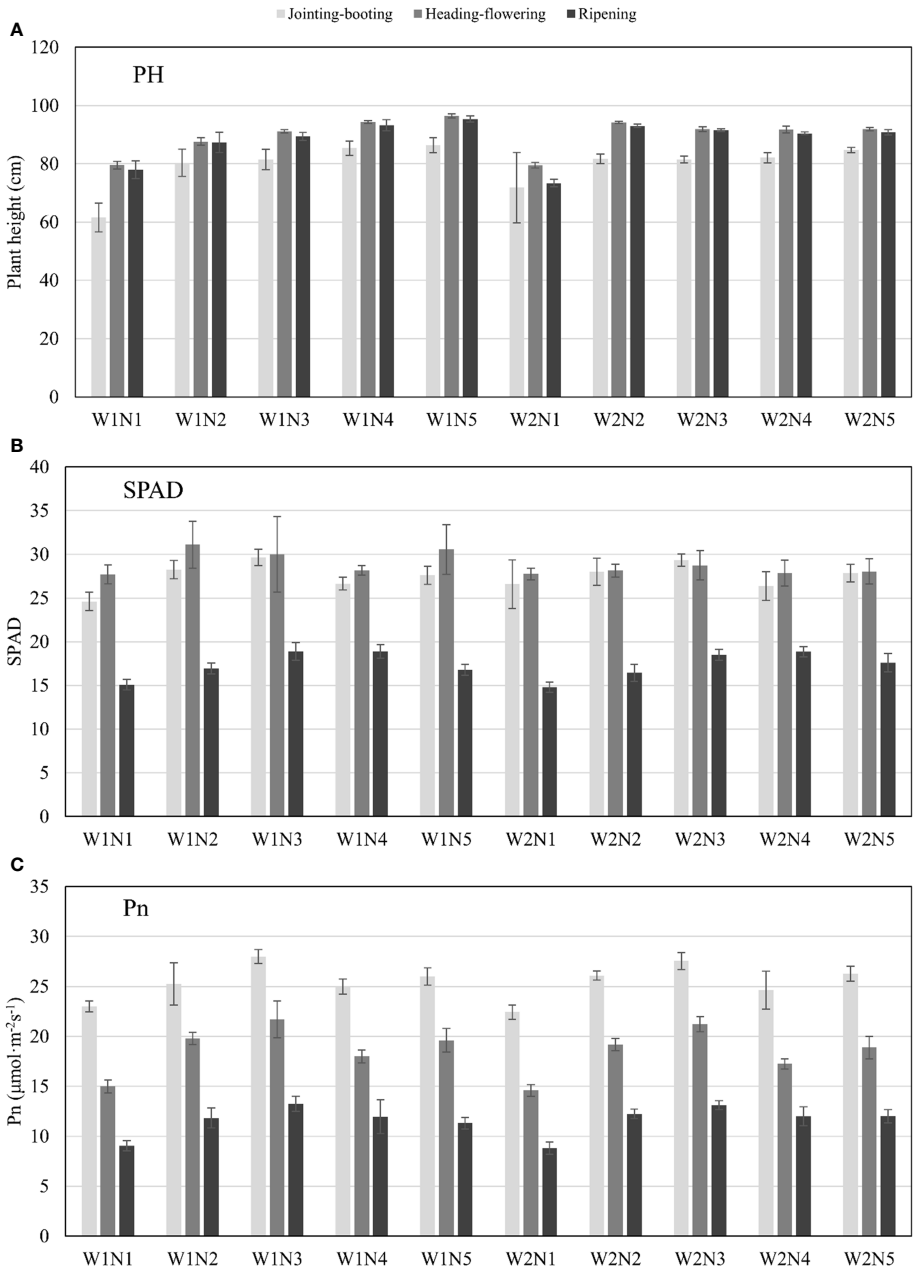


FIGURE 2
PH, SPAD and Pn of rice response to different treatments at different stages. **(A)** PH represents plant height (cm); **(B)** SPAD is relative chlorophyll content; **(C)** Pn represents the net photosynthetic rate (umol m-2s-1). W represents the leakage treatments (including W1: 3mm/day and W2: 5mm/day); N represents the nitrogen treatments (including N1-N5: 0, 75, 150, 225 and 300 kg/ha total pure nitrogen, respectively). The above ground measurements were conducted at the same time with UAV flight at 3 growth stages.

flowering stage, and finally dropped to approximately 0.030 at the ripening stage. Band 2 was the most intuitive band visible to the naked eye that could represent the nutrient statue and growth stage of crops, which reached a maximum of 0.160–0.221 at the jointing–booting stage, decreased to 0.139–0.174 with heading and flowering, and finally decreased to 0.027–0.038 with the yellowing of leaves and ears at the ripening stage.

For the same N application level at the jointing–booting stage, the average reflectance values of band 1, band 2, and band 3 under the W1 leakage treatment were generally lower than those under the W2 leakage treatment, except that W1N3 had higher band 1 and band 3 values than W2N3. However, the average reflectance values of band 4 and band 5 showed opposite trends; specifically, W1 leakage treatment could increase the reflectance of band 4 and band 5 compared with W2 treatment and the N3 level improved the most. It could also be found from Figure 3A that lower leakage treatment had a steeper increase from band 3 to band 4, which indicated the better growth status. Under the same leakage treatment, the average reflectance values of band 1 to band 4 generally presented a trend of initially decreasing (N1–N3), then increasing (N3–N4), and finally ending up with a small decrease (N4–N5), while the reflectance value of band 5 increased with the N application level, but the law was not

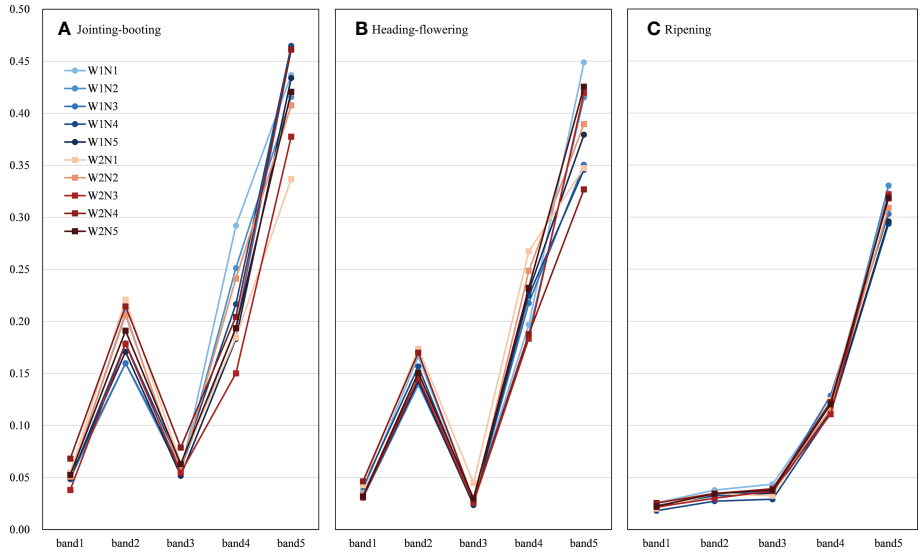


FIGURE 3
Average value of canopy band reflectance response to different treatments at different stages. **(A-C)** are the reflectance at jointing-booting, heading-flowering and ripening stage, respectively. Band1-band5 represent blue (450±16 nm), green (560±16 nm), red (650±16 nm), red edge (730±16 nm), near-infrared (840±26 nm) reflectance, respectively. W represents the leakage treatments (including W1: 3mm/day and W2: 5mm/day); N represents the nitrogen treatments (including N1-N5: 0, 75, 150, 225 and 300 kg/ha total pure nitrogen, respectively).

obvious. Moreover, the average reflectance value of these five bands did not show a certain rule in the heading–flowering stage and the ripening stage, which might be due to the influence of rice heading and flowering, here leaves and ear yellowing in the growth process on the multispectral characteristics of the canopy.

3.2 Correlation analysis between VIs, TIs, and Pn

Pearson's correlation coefficients (*r* values) between the above 25 VIs and Pn at three growth stages are listed in Table 6. The code for *r* value calculation and significance analysis was written in Python3.2 with the *scipy* package (1.20.3). Generally, the VIs had a better relationship with Pn at the jointing–booting stage, but the *r* value became worse as the growth stage advanced; however, it might be that the number of samples was relatively small and therefore no VIs passed the highly significant correlation test ($p< 0.01$). To be specific, CI_{green}, CVI, MNVI, NLI, OSAVI, RDVI, and RVI2 achieved a significant positive correlation ($p< 0.01$) with Pn at the jointing-boosting stage, with *r* value ranging from 0.3330 to 0.3893. NIRv, DVI, EVI, MSAVI, MSR, NDVI, RVI1, SIPI, and TCARI also had a satisfactory *r* value ($p< 0.05$) with absolute value between 0.2753 and 0.3262. At the heading–flowering stage, only CI_{green}, RVI2, and SIPI showed a significant relationship and NDVI, NLI, TCARI, and VDVI had a relatively higher *r* value. When the crop proceeded to the ripening stage, no VIs could achieve a satisfactory *r* value with Pn. The VIs employed for Pn estimation were thus selected based on the *r* value at different stages, and the selected VIs are shown in bold in Table 6.

The NDTI, DTI, and RDTI were calculated using any two texture features from NO. 1 to NO. 40 and the correlation thermal map in Figure 4 was thus drawn according to the *r* value between TIs and Pn. Because the order of the two selected texture features was different, the correlation *r* value in the figure presented positive and negative axis symmetry. Taking the thermal map at the jointing–booting stage as an example, it could be found that there were obviously deeper red and blue lines in the DTI and RDTI figure, indicating that the DTI and RDTI composed of any of the NO. 1 (MEAN1) and NO. 25 (MEAN4) features with the other feature had a better relationship with Pn (*r* value mainly approximately 0.35 and 0.39, respectively). Scattered hot spots with high *r* value could be seen in the RDTI figure, but no dominant texture feature could be found. For each TI, the feature combination with the highest correlation with Pn was selected and the results at different growth stages are listed in Table 7.

3.3 Estimation rice Pn from VIs at different growth stages

The accuracy comparison results between LR, SVR, GBDT, RF, and MLP models based on the selected VIs are listed in Table 8 and all criteria indices were calculated by the average of 10-fold cross-validation results. At the jointing–booting stage, most VIs showed good correlation with Pn and a total of 16 VIs were selected for modeling; therefore, the models achieved the relatively highest accuracy compared with those at other growth stages. Specifically, GBDT achieved the highest average accuracy (with an MSE of 0.253 μmol m^{-2} s^{-1}, an MAE of 0.414 μmol m^{-2} s^{-1}, an EVS of 0.938, and an R^2 of 0.938), while SVR models attained the worst performance (with an MSE of 2.512 μmol m^{-2} s^{-1}, an MAE of 1.155 μmol m^{-2} s^{-1}, an EVS of 0.390, and an R^2 of 0.383). RF, MLP, and LR models ranked second, third, and fourth, respectively. Figure 5A also shows that the Pn estimated value of GBDT was the closest to the measured Pn value in each validation set. The Pn estimated values of LR and SVR models were almost concentrated in the range of 24–27 μmol m^{-2} s^{-1}; thus,

TABLE 6 Correlation coefficient (*r* value) between selected VIs and Pn.

VIs	Jointing–booting (*n* = 60)	Heading–flowering (*n* = 60)	Ripening (*n* = 60)
NIRv	**0.3262***	0.2034	−0.0903
CI_{green}	**0.3893****	**0.3490****	**0.1693**
$CI_{red\ edge}$	0.1938	0.1046	0.0649
CVI	**0.3396****	0.1155	**0.1605**
DVI	**0.3186***	0.1737	−0.1083
EVI	**0.2803***	0.1333	−0.0949
GI	−0.1549	−0.0196	0.0137
GNDVI	−0.1473	0.1699	−0.0531
MCARI	−0.0091	−0.0642	−0.095
MNVI	**0.3342****	0.2119	−0.0798
MSAVI	**0.3281***	0.2194	−0.0820
MSR	**0.3157***	0.1905	0.0926
MTCI	0.1369	0.0886	0.0742
MTVI	0.1419	0.1113	−0.0846
NDVI	**0.3178***	**0.3166***	0.0370
NLI	**0.3386****	**0.2661***	−0.0056
OSAVI	**0.3330****	0.2622	−0.0487
RDVI	**0.3342****	0.2259	−0.0746
RVI1	**0.3136***	0.1038	0.1223
RVI2	**0.3893****	**0.3490****	**0.1693**
SIPI	**0.3258***	**0.3381****	0.0648
TCARI	**−0.2753***	**−0.3014***	**−0.2131**
TVI	−0.1262	−0.2094	**−0.1258**
VARI	−0.0984	0.0632	−0.0685
VDVI	−0.1881	**0.2746***	−0.0680

* and ** represent significant level p< 0.05 and p< 0.01, respectively. The number is the sample size of the data. VIs corresponding to the bolded value were selected as the inputs for machine learning modeling at different growth stages.

the estimation accuracy was not satisfactorily compared with GBDT, RF, and MLP models, where the Pn values were relatively lower and higher (inside the blue and red dashed circle). The correlation between VIs and Pn gradually weakened and inputted VIs thus decreased in number as the growth stage progressed; therefore, the model estimation accuracy decreased to a certain extent without ranking change. In detail, the LR model suffered the biggest loss in estimation accuracy with an R^2 value decreasing to 0.296 at the heading–flowering stage and then to 0.125 at the ripening stage. However, GBDT and RF models still showed good performance with R^2 values of 0.928 and 0.869 at the heading–flowering stage and 0.863 and 0.815 at the ripening stage, respectively. Figures 5B, C also demonstrated that the GBDT and RF models could better describe the relationship between VIs and Pn in the value full range at different growth stages. Although the performance of the MLP model ranked third, the performance was not ideal when compared with GBDT and RF models for both lower and higher PN value estimation. In conclusion, the accuracy of modes for estimating Pn value at the jointing–booting stage was relatively best and the GBDT model could be highly recommended for Pn estimation during the rice whole growth season.

3.4 Estimation rice Pn from fused VIs, TIs, and basal growth index

In order to further improve the Pn estimation accuracy, TIs (NDTI, DTI, and RDTI) and basal growth index (PH and SPAD) were introduced based on the input VIs. The accuracy comparison result of different models under different input combinations is shown in Table 9. After adding the TIs inputs for Pn estimation, all models performed higher accuracy at the jointing–booting stage with the MSE of LR, SVR, GBDT, RF, and MLP decreasing by 0.090, 0.370, 0.126, 0.113, and 0.057 μmol m^{-2} s^{-1}, respectively, and the R^2 increasing by 0.022, 0.090, 0.031, 0.028, and 0.015, respectively. As the basal growth index further increased, the final R^2 of the employed models increased to 0.792, 0.565, 0.987, 0.943, and 0.822, respectively. It is possible that the basal growth index improved the model accuracy slightly more than TIs because the difference in PH could reflect the stress of crops to a certain extent, and SPAD is directly related to chlorophyll content, which directly affects photosynthesis. The same improvement effect of model accuracy could also be found at the heading–flowering and ripening stages, and although the improvement of GBDT and RF models was relatively small (with R^2 increasing to 0.062 and 0.031 for VIs + TIs input and 0.113 and 0.132 for VIS + TIs + PH and SPAD input), they were still the top two models among the employed models. The greatest improvement could be found in the LR model, with an R^2 increase of 0.139 and 0.470 at the heading–flowering stage and 0.114 and 0.471 at the ripening stage under the VIs + TIs and VIS + TIs + PH and SPAD input combination, respectively. The accuracy of the SVR model also had been improved greatly, but it still is the lowest among the five models. In conclusion, both TIs and the basal growth index could obviously improve the model accuracy for Pn estimation and the PH and SPAD had a better effect compared with the TIs in this study, which significantly improved the model performance at the heading–flowering and ripening stages, especially for LR and MLP models.

4 Discussion

4.1 Relationship between rice growth and canopy multispectral feature

The ground sample results showed that the SPAD and Pn generally increased with nitrogen application (N1–N3 levels), then decreased at the N4 level, and finally recovered at the N5 level under the same leakage treatment at the jointing–booting stage, which also indicated that proper N application could improve the photosynthesis, while excessive N application not only had a poor effect on photosynthesis, but also affected plant growth and increased

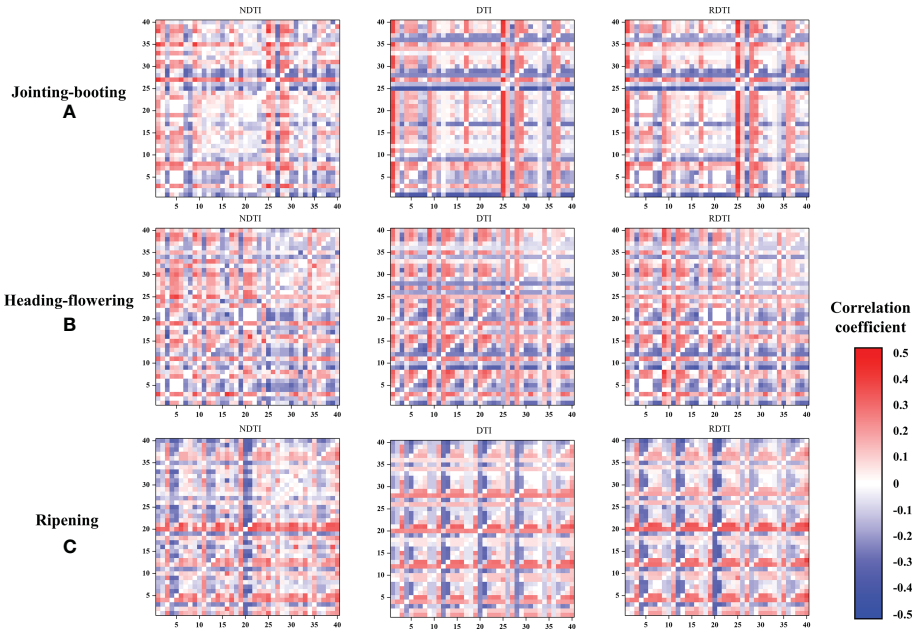


FIGURE 4
Correlation coefficient between Pn and TIs with different textural features combination. **(A–C)** are the correlation coefficient values at jointing-booting, heading-flowering and ripening stage, respectively. NDTI, DTI, RDTI represent normalized difference textural index, difference textural index and renormalized difference textural index, respectively. X-axis and Y-axis legends are the texture features in order of NO.1-40. The coloration in the thermal map is based on the correlation (r value) between TIs and Pn.

risk of contamination during leakage and drainage. This phenomenon was similar to that found by Cisse et al. (2020). It might due to the fact that there was no significant difference in Rubisco activity and non-photochemical quenching (NPQ) between the N4–N5 and the N3 level; thus, excessive energy could not be dissipated by NPQ, leading to oxidative stress, resulting in a decrease in Pn when excessive nitrogen was applied. An opposite trend could be found for canopy multispectral reflectance as the N application increased. Band 1 to band 4 generally decreased when N application increased from the N1–N3 level, while they slightly increased at the N4 level then decreased again at the N5 level; however, band 5 reflectance consistently increased with N level, which was consistent with previous studies on other crops Qiu et al. (2015). Generally, crops with good growth have a lower reflectance and a steeper increase from the red to the NIR band; thus, the variation trend of the canopy reflectance was consistent with SPAD and PN, which also provides a theoretical basis for the inversion of photosynthetic characteristic parameters using VIs.

4.2 Limitations and suggestions on Pn estimation using VIs and TIs in this study

Based on the result and analysis in *Section 3.1.1*, the relationship between VIs and Pn decreased with the growth stage, which might be due to the influence of heading and flowering, although panicle photosynthesis is also an important part of crop canopy photosynthesis and contributes significantly to grain formation. However, due to the limitation that the photosynthetic measurement equipment used in this paper could only be used to measure leaves, the canopy photosynthesis was thus approximately the photosynthetic capacity of the included leaves. Therefore, after rice heading and flowering, the spectral reflectance of the canopy was affected to a certain extent and the correlation based on the above data decreased significantly. In order to improve the estimation accuracy of Pn at the heading–flowering stage, the image segmentation should be carried out first to remove the panicle reflectance image. Meanwhile, the method of canopy photosynthesis measurement

TABLE 7 Compositions of selected TIs and its correlation with Pn.

TIs/stages	Jointing–booting		Heading–flowering		Ripening	
	Combination	*r*	Combination	*r*	Combination	*r*
NDTI	MEAN4, HOMO4	0.4361	COR2, HOMO1	0.4052	ENT2, DIS3	0.3526
DTI	MEAN4, DIS2	0.4042	MEAN2, COR1	0.3650	COR5, DIS3	0.3348
RDTI	MEAN4, COR5	0.4169	MEAN2, COR1	0.4016	COR5, DIS3	0.3808

NDTI, DTI, and RDTI represent normalized difference textural index, difference textural index, and renormalized difference textural index, respectively. The texture features mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation extracted by the GLCM method are abbreviated as MEAN, VAR, HOM, CON, DIS, ENT, SEC, and COR, respectively. The number after the abbreviation represents the band where the feature is extracted.

TABLE 8 Model performance for Pn estimation based on VIs.

Growth stage	Model	MSE (μmol m^{-2} s^{-1})	MAE (μmol m^{-2} s^{-1})	EVS	R^2
Jointing–booting	LR	1.859	1.128	0.543	0.543
	SVR	2.512	1.155	0.390	0.383
	GBDT	0.253	0.414	0.938	0.938
	RF	0.521	0.604	0.872	0.872
	MLP	0.942	0.777	0.768	0.768
Heading–flowering	LR	4.165	1.612	0.298	0.296
	SVR	4.388	1.589	0.261	0.258
	GBDT	0.425	0.505	0.928	0.928
	RF	0.774	0.691	0.869	0.869
	MLP	1.948	1.06	0.671	0.671
Ripening	LR	2.413	1.296	0.125	0.125
	SVR	2.293	1.061	0.205	0.169
	GBDT	0.377	0.507	0.863	0.863
	RF	0.511	0.610	0.815	0.815
	MLP	1.490	0.930	0.460	0.460

LR, SVR, GBDT, RF, and MLP represent linear regression, support vector regression, gradient boosting decision tree, random forest, and multilayer perceptron neural network models, respectively.

and the effect of panicle on photosynthetic contribution and reflectance should be revised and improved in future studies.

According to the correlation analysis results of TIs during the whole growth season in Figure 4, it could be concluded that most of the dominant texture features were extracted from band 1 to band 3, while the TIs constructed by the features extracted from band 4 and band 5 were not that satisfactory. Specifically, MEAN and COR texture features were more included in the optimal features for NDTI, DTI, and RDTI construction, because MEAN represents the average of moving windows containing targets and backgrounds, which can smooth the image and reduce the interference of background factors, while COR can reflect the local gray-level correlation in the image and distinguish the differences of image texture in all directions. In addition, window sizes of 6 × 6, 9 × 9, and 12 × 12 were also employed in the GLCM for textural feature extraction; however, the performance was not much different from that of the 3 × 3 window size, which might be due to the high resolution (4.5 mm/pixel) of each pixel in the image. Therefore, improving the resolution of visible images to extract higher-quality TIs might be an economical and practical approach to improve the accuracy of Pn estimation.

4.3 Influence of input combination on Pn estimation performance

The model performance under the different input combinations of VIs, VIs + TIs, and VIs + TIs + PH and SPAD concludes that the fusion of VIs and TIs could effectively improve the accuracy of Pn estimation because the VIs contain the canopy reflectance features and are more sensitive to the nutrition variations, while the TIs could better reflect the difference in canopy structure. The results were also consistent with previous studies (Liu et al., 2019; Zhang et al., 2021) in that this fusion combination could improve the estimation accuracy of biomass, LAI, nitrogen nutrition, and potassium nutrition and

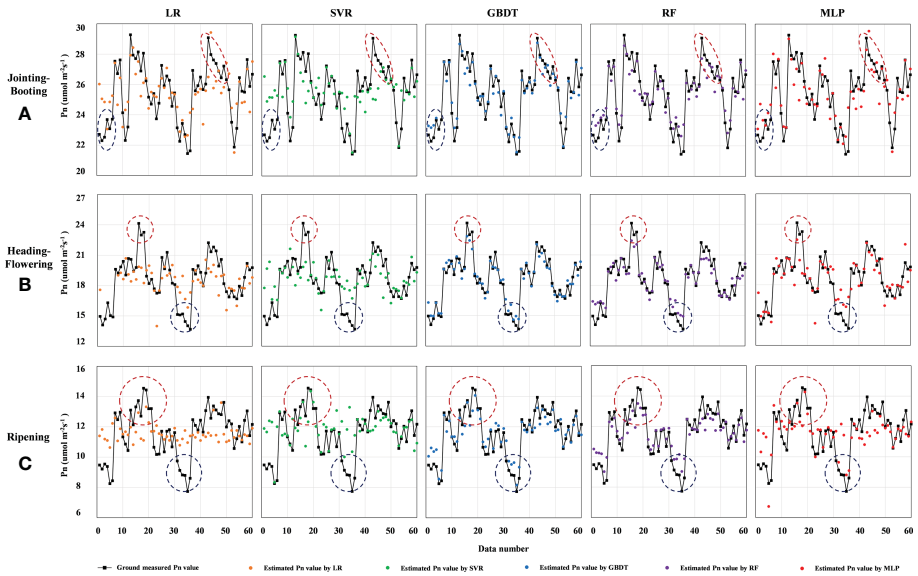


FIGURE 5
Comparison of estimated and measured Pn values of different models at certain stages. **(A-C)** are the comparison of estimated and measured Pn values at jointing-booting, heading-flowering and ripening stage, respectively. LR, SVR, GBDT, RF and MLP represent linear regression, support vector regression, gradient boosting decision tree, random forest and multilayer perceptron neural networks model, respectively. Pn represents the net photosynthetic rate (umol m-2s-1). The dashed blue and red circle in each figure are used to compare the fitting between the estimated and measured value when Pn value is low and high, respectively.

TABLE 9 Model performance for Pn estimation based on different input combination.

Growth stage	Model	VIs		VIs + TIs		VIs + TIs + PH and SPAD	
		MSE ($\mu mol\ m^{-2}\ s^{-1}$)	R^2	MSE ($\mu mol\ m^{-2}\ s^{-1}$)	R^2	MSE ($\mu mol\ m^{-2}\ s^{-1}$)	R^2
Jointing–booting	LR	1.859	0.543	1.769	0.565	0.846	0.792
	SVR	2.512	0.383	2.142	0.473	1.770	0.565
	GBDT	0.253	0.938	0.127	0.969	0.053	0.987
	RF	0.521	0.872	0.408	0.900	0.232	0.943
	MLP	0.942	0.768	0.885	0.783	0.723	0.822
Heading–flowering	LR	4.165	0.296	3.345	0.435	1.384	0.766
	SVR	4.388	0.258	4.052	0.315	3.471	0.413
	GBDT	0.425	0.928	0.321	0.946	0.052	0.991
	RF	0.774	0.869	0.768	0.870	0.174	0.971
	MLP	1.948	0.671	1.619	0.726	1.147	0.806
Ripening	LR	2.413	0.125	2.100	0.239	1.116	0.596
	SVR	2.293	0.169	2.229	0.192	1.834	0.335
	GBDT	0.377	0.863	0.207	0.925	0.066	0.976
	RF	0.511	0.815	0.426	0.846	0.146	0.947
	MLP	1.490	0.460	1.290	0.532	0.973	0.647

LR, SVR, GBDT, RF, and MLP represent linear regression, support vector regression, gradient boosting decision tree, random forest, and multilayer perceptron neural network models, respectively. VIs and TIs represent vegetation indices and textural indices, respectively. PH represents plant height (cm) and SPAD is relative chlorophyll content.

accumulation. Furthermore, the addition of PH and SPAD brought higher accuracy, which could attribute its success to the high correlation between PH and N nutrition status, SPAD and chlorophyll content, brought by the obvious difference of 5 N treatment. To sum up, more different kinds of data could introduce more direct or indirect related features, and it is also suggested that stacking and blending ensemble learning methods (Wu et al., 2021) could be employed to combine the model ability of feature extraction and analysis based on different principles in future research to improve the model accuracy for Pn estimation, which is also the purpose and significance of developing agricultural big data and agricultural intelligent models.

5 Conclusion

This paper studied and revealed the responses of canopy multispectral band reflectance and rice net photosynthetic rate (Pn) to different nitrogen applications and leakage treatments through different growth stages under controlled irrigation and drainage schemes. The relationship between VIs, TIs, and Pn based on the UAV multispectral image was comprehensively analyzed and focused on. The performance of LR, SVR, GBDT, RF, and MLP models for Pn estimation under different input combinations was evaluated and compared at the jointing–booting, heading–flowering and ripening stages. The results indicated that the selected VIs and TIs had a relatively better correlation relationship with Pn at the jointing–booting stage, while only a moderate correlation at the heading–flowering stage and an unsatisfactory correlation at the ripening stage could be found. Therefore, the employed models generally had a better performance during the jointing–booting stage and the accuracy decreased as the growth stage progressed. Among the five used models, GBDT and RF models achieved the highest and most stable accuracy in the whole growth season and could be highly recommended for Pn estimation in the paddy field. Meanwhile, the fusion of VIs with TIs and basal growth index could significantly improve the model accuracy, and the plant height (PH) and SPAD had a better effect on performance improvement compared with NDTI, DTI, and RDTI employed in this study. The techniques and results presented in this paper could be valuable for rice field-scale photosynthetic monitoring, which could assist further stress detection and yield prediction.

Data availability statement

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

Author contributions

Conceptualization, TW and WZ. Methodology, TW and MC. Software, TW. Validation, TW, ZW and MC. Formal analysis, SW. Data curation, LQ. Writing—original draft preparation, TW. Writing—review and editing, TW, and LQ. Visualization, SW. Supervision, GS and

XJ. Project administration, XJ. Funding acquisition, XJ. All authors contributed to the article and approved the submitted version.

Funding

This study is financially supported by the Jiangsu Provincial Key Research and Development Program (BE2022390) and the Postgraduate Research and Practice Innovation Program of Jiangsu Province (KYCX21_0537).

Acknowledgments

We sincerely thank Yi Wang and Luying Zhou from Hohai University for field management and data acquisition.

Conflict of interest

Author LQ was employed by China Design Group Co., Ltd.

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

Publisher's note

References

Aguirre-Gomez, R., Boxall, S. R., and Weeks, A. R. (2001). Detecting photosynthetic algal pigments in natural populations using a high-spectral-resolution spectroradiometer. *Int. J. Remote Sens.* 22 (15), 2867–2884. doi: 10.1080/01431160120387

Birth, G. S., and McVey, G. R. (1968). Measuring the color of growing turf with a reflectance spectrophotometer 1. *Agron. J* 60 (6), 640. doi: 10.2134/agronj1968.00021962006000060016x

Blackburn, G. A. (1998). Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves. *Int. J. Remote Sens.* 19 (4), 657–675.doi: 10.1080/014311698215919

Breiman, L. (2001). Random forests. *Mach. Learn* 45 (1), 5–32. doi: 10.1023/A:1010933404324

Broge, N. H., and Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sens. Environ* 76 (2), 156–172. doi: 10.1016/S0034-4257(00)00197-8

Bruce, L. M., Koger, C. H., and Li, J. (2002). Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Trans. Geosci. Remote Sens.* 40 (10), 2331–2338. doi: 10.1109/TGRS.2002.804721

Chen, J. M. (1996). Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Can. J. Remote Sens.* 22 (3), 229–242. doi: 10.1080/07038992.1996.10855178

Chen, J., Chen, S., Zhang, Z., Fu, Q., Bian, J., and Cui, T. (2018). Investigation on photosynthetic parameters of cotton during budding period by multi-spectral remote sensing of unmanned aerial vehicle. *Nongye Jixie Xuebao/Transactions Chin. Soc Agric. Mach* 49 (10), 230–239. doi: 10.6041/j.issn.1000-1298.2018.10.026

Cheng, M., Jiao, X., Liu, Y., Shao, M., Yu, X., Bai, Y., et al. (2022). Estimation of soil moisture content under high maize canopy coverage from UAV multimodal data and machine learning. *Agric. Water Manag.* 264. doi: 10.1016/j.agwat.2022.107530

Chen, K., Yu, S., Ma, T., Ding, J., He, P., Li, Y., et al. (2022a). Modeling the water and nitrogen management practices in paddy fields with HYDRUS-1D. *Agric*12 (7), 924. doi: 10.3390/agriculture12070924

Chen, Q., Zheng, B., Chenu, K., Hu, P., and Chapman, S. C. (2022b). Unsupervised plot-scale LAI phenotyping via UAV-based imaging, modelling, and machine learning. *Plant Phenomics* 2022. doi: 10.34133/2022/9768253

Che, Y., Wang, Q., Xie, Z., Zhou, L., Li, S., Hui, F., et al. (2020). Estimation of maize plant height and leaf area index dynamics using an unmanned aerial vehicle with oblique and nadir photography. *Ann. Bot* 126 (4), 765–773. doi: 10.1093/aob/mcaa097

Cisse, A., Zhao, X., Fu, W., Kim, R. E. R., Chen, T., Tao, L., et al. (2020). Non-photochemical quenching involved in the regulation of photosynthesis of rice leaves under high nitrogen conditions. *Int. J. Mol. Sci.* 21 (6), 215. doi: 10.3390/ijms21062115

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn* 20 (3), 273–297. doi: 10.1007/bf00994018

Dash, J., and Curran, P. J. (2004). The MERIS terrestrial chlorophyll index. *Int. J. Remote Sens.*25 (3), 5403–5413. doi: 10.1080/0143116042000274015

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., De Colstoun, E. B. , and McMurtrey, J. (2000) Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *E Remote Sens. Environ*74 (2), 229–239. doi: 10.1016/S0034-4257(00)00113-9

de Castro, A. I., Shi, Y., Maja, J. M., and Peña, J. M. (2021). Uavs for vegetation monitoring: Overview and recent scientific contributions. *Remote Sens.* 13 (11), 1–13. doi: 10.3390/rs13112139

Deery, D., Jimenez-Berni, J., Jones, H., Sirault, X., and Furbank, R. (2014). Proximal remote sensing buggies and potential applications for field-based phenotyping. *Agronomy* 4 (3), 349–379. doi: 10.3390/agronomy4030349

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Stat* 29 (5), 1189–1232. doi: 10.1214/aos/1013203451

Fu, P., Meacham-Hensold, K., Guan, K., and Bernacchi, C. J. (2019). Hyperspectral leaf reflectance as proxy for photosynthetic capacities: An ensemble approach based on multiple machine learning algorithms. *Front. Plant Sci.*10. doi: 10.3389/fpls.2019.00730

Fu, P., Meacham-Hensold, K., Guan, K., Wu, J., and Bernacchi, C. (2020). Estimating photosynthetic traits from reflectance spectra: A synthesis of spectral indices, numerical inversion, and partial least square regression. *Plant Cell Environ.*43 (5), 1241–1258. doi: 10.1111/pce.13718

Fu, P., Montes, C. M., Siebers, M. H., Gomez-Casanovas, N., McGrath, J. M., Ainsworth, E. A., et al. (2022). Advances in field-based high-throughput photosynthetic phenotyping. *J. Exp. Bot.* 73 (10), 3157–3172. doi: 10.1093/jxb/erac077

Gitelson, A. A., Gritz, Y., and Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol* 160 (3), 271–282. doi: 10.1078/0176-1617-00887

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

Gitelson, A. A., Kaufman, Y. J., Stark, R., and Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ* 80 (1), 76–87. doi: 10.1016/S0034-4257(01)00289-9

Gitelson, A. A., Viña, A., Ciganda, V., Rundquist, D. C., and Arkebauer, T. J. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett* 32 (8). doi: 10.1029/2005GL022688

Goel, N. S., and Qin, W. (1994). Influences of canopy architecture on relationships between various vegetation indices and LAI and FPAR: a computer simulation. *Remote Sens. Rev* 10 (4), 309–347. doi: 10.1080/02757259409532252

Gong, P., Pu, R., Biging, G. S., and Larrieu, M. R. (2003). Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* 41 (6), 1355–1362. doi: 10.1109/TGRS.2003.812910

Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., and Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ* 90 (3), 337–352. doi: 10.1016/j.rse.2003.12.013

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., and Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ* 81 (2-3), 416–426. doi: 10.1016/S0034-4257(02)00018-4

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

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ* 3 (6), 610–621. doi: 10.1016/S0034-4257(02)00096-2

Ivlev, A. (2017). Global photosynthesis and its regulatory role in natural carbon cycle. *J. Ecosyst. Ecography.* 7, 233. doi: 10.4172/2157-7625.1000233

Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology* 50 (4), 663. doi: 10.2307/1936256

Khoshhal, J., and Mokarram, M. (2012). Model for prediction of evapotranspiration using MLP neural network. *Int. J. Environ. Sci* 3, 1000–1009.

Liang, L., Di, L., Zhang, L., Deng, M., Qin, Z., Zhao, S., et al. (2015). Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method. *Remote Sens. Environ* 165, 123–134. doi: 10.1016/j.rse.2015.04.032

Liang, L., Qin, Z., Zhao, S., Di, L., Zhang, C., Deng, M., et al. (2016). Estimating crop chlorophyll content with hyperspectral vegetation indices and the hybrid inversion method. *Int. J. Remote Sens.* 37 (13), 2923–2949. doi: 10.1080/01431161.2016.1186850

Liu, Y., Liu, S., Li, J., Guo, X., Wang, S., and Lu, J. (2019). Estimating biomass of winter oilseed rape using vegetation indices and texture metrics derived from UAV multispectral images. *Comput. Electron. Agric* 166. doi: 10.1016/j.compag.2019.105026

Liu, C., and Peng, Yi (2020). Remote estimation of rice leaf net photosynthetic rate based on hyperspectral reflectance. *J. China Agric. Univ.* 25 (1), 56–65. doi: 10.11841/j.issn.1007-4333.2020.01.07

Li, L., Zhang, Q., and Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors (Switzerland).* 2020. doi: 10.3390/s141120078

Long, S. P., Zhu, X. G., Naidu, S. L., and Ort, D. R. (2006). Can improvement in photosynthesis increase crop yields? *Plant Cell Environ* 29 (3), 315–330. doi: 10.1111/j.1365-3040.2005.01493.x

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

Lu, N., Wang, W., Zhang, Q., Li, D., Yao, X., Tian, Y., et al. (2019). Estimation of nitrogen nutrition status in winter wheat from unmanned aerial vehicle based multi-angular multispectral imagery. *Front. Plant Sci* 10. doi: 10.3389/fpls.2019.01601

Meroni, M., and Colombo, R. (2006). Leaf level detection of solar induced chlorophyll fluorescence by means of a subnanometer resolution spectroradiometer. *Remote Sens. Environ* 103 (4), 438–448. doi: 10.1016/j.rse.2006.03.016

Pimstein, A., Karnieli, A., Bansal, S. K., and Bonfil, D. J. (2011). Exploring remotely sensed technologies for monitoring wheat potassium and phosphorus using field spectroscopy. *F. Crop Res* 121 (1), 125–135. doi: 10.1016/j.fcr.2010.12.001

Qiu, Y., Yang, F., Zhao, G., Wang, L., Zhang, Y., and Wang Rui, H. S. (2015). Analysis of correlation between photosynthetic fluorescence parameters and hyperspectral data in ear-leaf of maize. *Chin. Agric. Sci. Bull.* 31, 71–77.

Reich, P. B., and Amundson, R. G. (1985). Ambient levels of ozone reduce net photosynthesis in tree and crop species. *Science* 230 (4725), 566–570. doi: 10.1126/science.230.4725.566

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M. D., Seneviratne, S. I., et al. (2013). Climate extremes and the carbon cycle. *Nature* 500, 287–295. doi: 10.1038/nature12350

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

Roujean, J. L., and Breon, F. M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ* 51 (3), 375–384. doi: 10.1016/0034-4257(94)00114-3

Sarker, L. R., and Nichol, J. E. (2011). Improved forest biomass estimates using ALOS AVNIR-2 texture indices. *Remote Sens. Environ* 115 (4), 968–977. doi: 10.1016/j.rse.2010.11.010

Smith, R. C. G., Adams, J., Stephens, D. J., and Hick, P. T. (1995). Forecasting wheat yield in a Mediterranean-type environment from the NOAA satellite. *Aust. J. Agric. Res* 46 (1), 113–125. doi: 10.1071/AR9950113

Song, Q., Xiao, H., Xiao, X., and Zhu, X. G. (2016). A new canopy photosynthesis and transpiration measurement system (CAPTS) for canopy gas exchange research. *Agric. For. Meteorol* 217, 101–107. doi: 10.1016/j.agrformet.2015.11.020

Stinziano, J. R., McDermitt, D. K., Lynch, D. J., Saathoff, A. J., Morgan, P. B., and Hanson, D. T. (2019). The rapid A/Ci response: a guide to best practices. *New Phytol* 221 (2), 625–627. doi: 10.1111/nph.15383

Sun, S. B., Du, H. Q., Li, P. H., Zhou, G. M., Xu, X. J., Gao, G. L., et al. (2016). Retrieval of leaf net photosynthetic rate of moso bamboo forests using hyperspectral remote sensing based on wavelet transform. *Chin. J. Appl. Ecol* 27(1), 49–58. doi: 10.13287/j.1001-9332.201601.020

Tian, C., Huang, C., Guo, X., Liu, X., and Wang, D. (2018). Correlation analysis between cotton canopy reflectance spectra and leaf net photosynthetic rate. *Remote Sens. Inf.* 33 (1), 99–103. doi: 10.3969/j.issn.1000-3177.2018.01.015

Tucker, C. J., Elgin, J. H., McMurtrey, J. E., and Fan, C. J. (1979). Monitoring corn and soybean crop development with hand-held radiometer spectral data. *Remote Sens. Environ* 8 (3), 237–248. doi: 10.1016/0034-4257(79)90004-X

Vincini, M., Frazzi, E., and D'Alessio, P. (2008). A broad-band leaf chlorophyll vegetation index at the canopy scale. *Precis. Agric* 9, 303–319. doi: 10.1007/s11119-008-9075-z

Wang, Z., Ma, Y., Chen, P., Yang, Y., Fu, H., Yang, F., et al. (2022). Estimation of rice aboveground biomass by combining canopy spectral reflectance and unmanned aerial vehicle-based red green blue imagery data. *Front. Plant Sci.*13. doi: 10.3389/fpls.2022.903643

Wan, L., Zhang, J., Dong, X., Du, X., Zhu, J., Sun, D., et al. (2021). Unmanned aerial vehicle-based field phenotyping of crop biomass using growth traits retrieved from PROSAIL model. *Comput. Electron. Agric.*187, 106304. doi: 10.1016/j.compag.2021.106304

Wu, T., Zhang, W., Jiao, X., Guo, W., and Alhaj Hamoud, Y. (2021). Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration. *Comput. Electron. Agric* 184, 106039. doi: 10.1016/j.compag.2021.106039

Wu, T., Zhang, W., Jiao, X., Guo, W., and Hamoud, Y. A. (2020). Comparison of five boosting-based models for estimating daily reference evapotranspiration with limited meteorological variables. *PloS One* 15 (6). doi: 10.1371/journal.pone.0235324

Xue, L., Cao, W., Luo, W., Dai, T., and Zhu, Y. (2004). Monitoring leaf nitrogen status in rice with canopy spectral reflectance. *Agron. J* 26 (10), 3463–3469. doi: 10.2134/agronj2004.0135

Xu, H. C., Yao, B., Wang, Q., Chen, T. T., Zhu, T. Z., He, H. B., et al. (2021). Determination of suitable band width for estimating rice nitrogen nutrition index based on leaf reflectance spectra. *Sci. Agric. Sin.*54, 4525–4538. doi: 10.3864/j.issn.0578-1752.2021.21.004

Yang, D., and Bao, W. (2017). Group lasso-based band selection for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett* 14 (12), 2438–2442. doi: 10.1109/LGRS.2017.2768074

Yang, Z., Tian, J., Wang, Z., and Feng, K. (2022). Monitoring the photosynthetic performance of grape leaves using a hyperspectral-based machine learning model. *Eur. J. Agron.*140, 126589. doi: 10.1016/j.eja.2022.126589

Yuan, D., Zhang, S., Li, H., Zhang, J., Yang, S., and Bai, Y. (2022). Improving the gross primary productivity estimate by simulating the maximum carboxylation rate of the crop using machine learning algorithms. *IEEE Trans. Geosci. Remote Sens.*60. doi: 10.1109/TGRS.2022.3200988

Zeng, Y., Badgley, G., Dechant, B., Ryu, Y., Chen, M., and Berry, J. A. (2019). A practical approach for estimating the escape ratio of near-infrared solar-induced chlorophyll fluorescence. *Remote Sens. Environ* 232, 111209. doi: 10.1016/j.rse.2019.05.028

Zeng, Y., Hao, D., Huete, A., Dechant, B., Berry, J., Chen, J. M., et al. (2022). Optical vegetation indices for monitoring terrestrial ecosystems globally. *Nat. Rev. Earth Environ.*3, 477–493. doi: 10.1038/s43017-022-00298-5

Zhang, X. Y., Huang, Z., Su, X., Siu, A., Song, Y., Zhang, D., et al. (2020b). Machine learning models for net photosynthetic rate prediction using poplar leaf phenotype data. *PloS One* 15 (2). doi: 10.1371/journal.pone.0228645

Zhang, J., Qiu, X., Wu, Y., Zhu, Y., Cao, Q., Liu, X., et al. (2021). Combining texture, color, and vegetation indices from fixed-wing UAS imagery to estimate wheat growth parameters using multivariate regression methods. *Comput. Electron. Agric* 185 (10), 3463–3469. doi: 10.1016/j.compag.2021.106138

Zhang, N., Su, X., Zhang, X., Yao, X., Cheng, T., Zhu, Y., et al. (2020a). Monitoring daily variation of leaf layer photosynthesis in rice using UAV-based multi-spectral imagery and a light response curve model. *Agric. For. Meteorol.*291. doi: 10.1016/j.agrformet.2020.108098

Zhang, X., Zhang, F., Qi, Y., Deng, L., Wang, X., and Yang, S. (2019). New research methods for vegetation information extraction based on visible light remote sensing images from an unmanned aerial vehicle (UAV). *Int. J. Appl. Earth Obs. Geoinf.* 78, 215–226. doi: 10.1016/j.jag.2019.01.001

Zheng, H., Cheng, T., Zhou, M., Li, D., Yao, X., Tian, Y., et al. (2019). Improved estimation of rice aboveground biomass combining textural and spectral analysis of UAV imagery. *Precis. Agric* 20, 611–629. doi: 10.1007/s11119-018-9600-7

Zheng, H., Ma, J., Zhou, M., Li, D., Yao, X., Cao, W., et al. (2020). Enhancing the nitrogen signals of rice canopies across critical growth stages through the integration of textural and spectral information from unmanned aerial vehicle (UAV) multispectral imagery. *Remote Sens.* 12 (6), 957. doi: 10.3390/rs12060957

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 26 January 2023
DOI 10.3389/fpls.2023.1091600

OPEN ACCESS

EDITED BY
Baohua Zhang,
Nanjing Agricultural University, China

REVIEWED BY
Sashuang Sun,
Zhejiang University, China
Chu Zhang,
Huzhou University, China
Changmiao Wang,
The Chinese University of Hong Kong, Shenzhen, China

*CORRESPONDENCE
Rujing Wang
rjwang@iim.ac.cn
Taosheng Xu
taosheng.x@gmail.com

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

RECEIVED 07 November 2022
ACCEPTED 03 January 2023
PUBLISHED 26 January 2023

CITATION
Hu X, Wang R, Du J, Hu Y, Jiao L and Xu T (2023) Class-attention-based lesion proposal convolutional neural network for strawberry diseases identification. *Front. Plant Sci.* 14:1091600. doi: 10.3389/fpls.2023.1091600

Class-attention-based lesion proposal convolutional neural network for strawberry diseases identification

Xiaobo Hu[1,2], Rujing Wang[1,2,3]*, Jianming Du[2], Yimin Hu[2], Lin Jiao[2,4] and Taosheng Xu[2]*

[1]Science Island Branch, University of Science and Technology of China, Hefei, Anhui, China, [2]Institute of Intelligent Machines, Hefei Institutes of Physical Science, Chinese Academy of Sciences (CAS), Hefei, Anhui, China, [3]Institute of Physical Science and Information Technology, Anhui University, Hefei, Anhui, China, [4]School of Internet, Anhui University, Hefei, Anhui, China

Diseases have a great impact on the quality and yield of strawberries, an accurate and timely field disease identification method is urgently needed. However, identifying diseases of strawberries in field is challenging due to the complex background interference and subtle inter-class differences. A feasible method to address the challenges is to segment strawberry lesions from the background and learn fine-grained features of the lesions. Following this idea, we present a novel Class-Attention-based Lesion Proposal Convolutional Neural Network (CALP-CNN), which utilizes a class response map to locate the main lesion object and propose discriminative lesion details. Specifically, the CALP-CNN firstly locates the main lesion object from the complex background through a class object location module (COLM) and then applies a lesion part proposal module (LPPM) to propose the discriminative lesion details. With a cascade architecture, the CALP-CNN can simultaneously address the interference from the complex background and the misclassification of similar diseases. A series of experiments on a self-built dataset of field strawberry diseases is conducted to testify the effectiveness of the proposed CALP-CNN. The classification results of the CALP-CNN are 92.56%, 92.55%, 91.80% and 91.96% on the metrics of accuracy, precision, recall and F1-score, respectively. Compared with six state-of-the-art attention-based fine-grained image recognition methods, the CALP-CNN achieves 6.52% higher (on F1-score) than the sub-optimal baseline MMAL-Net, suggesting that the proposed methods are effective in identifying strawberry diseases in the field.

KEYWORDS

convolutional neural network, strawberry disease identification, complex background, similar diseases, class response map, main lesion object, lesion details

1 Introduction

Strawberry, often praised as the "Queen of Fruits", is rich in vitamin C and antioxidants that support heart health and blood sugar control (Hannum, 2004). It is becoming a new income-producing agricultural product compared with traditional crops. However, strawberries are very delicate and highly susceptible to infection in natural environment. They are prone to various infectious diseases caused by fungal, bacterial and viral pathogens (Iqbal et al., 2021). Up to now, many strawberry diseases have been identified during the whole cultivation period of strawberries. These diseases can occur in strawberries' fruit, leaf, and stem, such as gray mold, powdery mildew and anthracnose. Therefore, disease management is a routine and labor-intensive requirement in strawberry cultivation. Currently, the identification of strawberry diseases is empirically conducted by growers, especially in China. The various types of diseases pose a great challenge to the accurate identification of the growers. Meanwhile, the manual manners are expensive, laborious and subjective, making them hard to wildly apply in modern agriculture. Hence, the current strawberry disease management cannot meet the need for automatic monitoring in agricultural practice (Hu et al., 2021). Furthermore, most strawberry growers lack professional knowledge to distinguish the diseases, resulting in the use of incorrect and overdose fungicides in disease management. The abuse of fungicides greatly harms the health of consumers and has caused substantial economic loss (Wang et al., 2021b). There is an urgent need for a fast and effective method to identify diseases in strawberry farming.

In general, the visual symptoms, including color, texture, shape and location of the lesions are important evidence for disease identification (Sankaran et al., 2010; Cruz et al., 2019; Liang et al., 2019). Given these visual features, various methods based on computer vision (CV) technology have been developed to identify different crop diseases. The CV-based methods for crop disease identification can be summarized into two streams. In the first stream, the traditional CV-based methods (such as color space transform, histogram of oriented gradient and gray level co-occurrence matrix [GLCM]) are applied to extract lesion features from diseased spots (Kim et al., 2009; Revathi and Hemalatha, 2014; Kaur et al., 2016; Johannes et al., 2017). Then, a classifier (e.g., linear/logistic regression, random forest) is constructed to yield classification results based on the extracted features (Huang, 2007; Kaur et al., 2016; Iqbal et al., 2018; Dwivedi et al., 2021). For instance, three phalaenopsis seedlings diseases had been successfully identified by an artificial neural network with the GLCM extracted texture features (Huang, 2007). Besides, (Johannes et al., 2017) designed two descriptors of their segmented hot-spot blobs to validate the effectiveness of the related traditional CV-based methods in identifying diseases at the early stage under a complex field background. The two descriptors were used to describe the color and texture features of the blob lab channels, respectively. These studies have proved that traditional CV-based methods are effective in recognizing the diseases of crops in both laboratory and field environments. However, these methods rely on the manual selection of discriminative features among diseases. The discriminative feature selection in field disease identification is very difficult and time-consuming (Zhao et al., 2022). Furthermore, the identification accuracy could dramatically decrease with a slight change in the input image dataset (Arsenovic et al., 2019). These shortcomings result in the traditional CV-based methods rarely adopted in the practice of crop disease identification. The convolutional neural network (CNN) and its variants lead the second stream for crop disease identification. The CNN-based models can automatically extract basic features like color, texture, edge, and location information. Meanwhile, they are competent to obtain more abstract semantic information from the image of crop diseases (Zeiler and Fergus, 2014). Besides, these CNN-based models have more flexible architectures that can be applied as feature extractors or classifiers. In recent studies, the CNN-based models have become the preferred method to identify crop diseases (Liang et al., 2019; Hu et al., 2021; Yang et al., 2022; Zhao et al., 2022). Earlier studies apply the classical CNN models, such as AlexNet (Krizhevsky et al., 2012), GoogLeNet (Szegedy et al., 2015), and ResNet (He et al., 2016) on some specific crop disease datasets and found the most suitable model for the disease identification tasks (Mohanty et al., 2016; Srdjan. et al., 2016; Ferentinos, 2018; Too et al., 2019; Picon et al., 2019). The related models achieve preferable recognition accuracy on their disease datasets. However, these studies fail to consider the complexity of the practical application of field disease identification. The main challenges of field disease identification are the complex background and a variety of diseases with similar symptoms (Barbedo, 2018). These models cannot be applied to crop cultivation practice. Consequently, some researches aim at reducing the misclassification caused by complex backgrounds and diseases with similar symptoms.

A simple yet effective method to eliminate the influence of complex background on disease identification is to segment the lesion region from their background. Several CNN-based semantic segmentation methods have been proposed to mitigate the adverse impact of the background. (Ngugi et al., 2020) proposed a segmentation network, KijianiNet, to segment tomato leaves from the natural field conditions. (Hu et al., 2021) and (Wang et al., 2021a) adopted U-Net (Ronneberger et al., 2015) and DeepLabV3+ (Chen et al., 2018) in the first stage of their models to segment the diseased leaves from the field scenes, respectively. The related experimental results showed that extracting diseased regions from the background can greatly improve the identification performance of the models. However, CNN-based semantic segmentation methods require pixel-level supervision. Such pixel-level annotation by experts is time-consuming, laborious and costly since plenty of lesions have varied shapes. On the topic of similar disease identification, few studies have proposed effective approaches to tackle this issue. (Cruz et al., 2019) applied transfer learning and data augmentation technologies to enhance the ability of the classical CNN models (e.g., AlexNet, GoogLeNet and ResNet) to distinguish the grapevine yellow from its similar diseases (such as grapevine leafroll and stictocephala bisonia). The experimental results confirmed that the data augmentation technologies were beneficial for classical CNN models to identify grape diseases. Because a suitable data augmentation strategy could increase the differences among similar diseases. However, the strategy was not easy to obtain, it required trial and error. The research of (Yang et al., 2022) was a development in identifying similar diseases of field crops. Similar diseases were classified by increasing the weight of discriminative lesion features.

To locate lesion details and learn discriminative lesion features among similar diseases, they proposed a self-supervised multi-network fusion classification model. However, the locations of the lesion details were randomly generated. Furthermore, all the obtained lesion details need to be fed to a classifier to assess the confidence of these regions as lesions, which greatly increased the time consumption of the model.

Image-based automatic disease identification is a basic need of modern large-scale cultivation agriculture. Field disease identification is challenging due to the complex background and similar symptoms among diseases. To address these problems, this paper focuses on strawberry field disease identification and proposes a novel Class-Attention-based Lesion Proposal Convolutional Neural Network (CALP-CNN) to precisely identify strawberry diseases in the field. The CALP-CNN method first utilizes a class-attention mechanism to enhance the localization of discriminative lesion feature. Two specific modules (i.e., the class object location module, COLM, and the lesion part proposal module, LPPM) are designed to recursively segment the main lesion object and lesion detail from an input image. Finally, the features of the original, main lesion object and lesion details are concatenated for final identification. To our knowledge, the CALP-CNN method is the first attempt to simultaneously address the challenges posed by the complex background and similar symptoms to crop disease identification in the field. The main contributions are summarized as follows:

- We introduce a new class attention mechanism (i.e., the class response map) to improve the ability of the CNN to localize the discriminative lesion features.
- We address the challenges of field disease identification by developing a novel CALP-CNN that simultaneously removes the noisy background and effectively learns discriminative lesion representations among similar diseases in an unsupervised way.
- A series of experiments are conducted on the field strawberry disease dataset to evaluate the effectiveness of the CALP-CNN. The experimental results show that the proposed method has better performance than other state-of-the-art fine-grained methods on field strawberry disease identification.

2 Material and methods

2.1 Material

In this paper, the strawberry diseases with high incidence in planting practice were taken as our research objects. To this end, a strawberry common disease dataset (SCDD) was constructed. The SCDD was collected in two ways: field-collection and internet crawling. We firstly shot 1,326 disease images of three strawberry varieties (Fengxiang, Nvfeng and Hongyan) in ChangFeng County, Anhui Province, China. To increase the diversity of the dataset, the images were deliberately captured in the field at different angles and focal lengths. The second part was from the internet. A crawler was applied to download more than 5,000 images of field strawberry diseases. The collected images were manually screened one by one to discard the poor-quality samples (obscure and the resolution less than 224×224). The disease images in the dataset were annotated by three experts. One was responsible for labeling the dataset, and the other two were responsible for reviewing the results. Finally, a high-quality dataset of strawberry diseases with 3,411 images was constructed for downstream analysis. The SCDD contained 11 common diseases and healthy type. Table 1 shows detailed information of the SCDD. In addition, the typical symptoms of 11 common strawberry diseases are shown in Figure 1.

In our experiments, the dataset was randomly divided into a training set, a validation set and a testing set in the ratio of 6:2:2 (2,047 images for training, 682 images for validation, and the remaining 682 images for testing). In the training process, we adopted the online data augmentation strategies to increase the diversity of the dataset and the robustness of the models. Specifically, the processes of Normalize, RandomHorizontalFlip, RandomVerticalFlip, and RandomResizedCrop (crop to 224×224) were applied during training.

2.2 Methods

In this paper, a class-attention-based lesion proposal CNN is presented to settle the main challenges of CNN-based methods in field disease identification, i.e., the complex background and similar diseases. The framework of CALP-CNN is shown as Figure 2. A cascade architecture is designed for extracting the region-based features from the input images at three scales including the raw image at coarse-grained level, the main lesion object at medium-grained level and the lesion detail images at fine-grained level. Furthermore, a series of modules are developed to extract class-related features in each layer of the cascade architecture. The detailed information of the CALP-CNN is described as follows:

First, a CNN backbone is repeatedly applied to extract region-based features from the input images in three scales. The CNN modules in three scales are given the same parameters. Second, the features are fed forward to three classifiers to predict three probability scores. The computed probability scores represent the prediction

TABLE 1 List of strawberry common disease dataset.

Category label	Strawberry disease	Number
0	healthy	509
1	leaf scorch	287
2	gray mold	332
3	powdery mildew	344
4	brown spot	215
5	fertilizer disorder	308
6	fusarium wilt	145
7	white leaf spot	259
8	calcium deficiency	431
9	magnesium deficiency	197
10	anthracnose	198
11	bacterial leaf spot	186
Total		3411

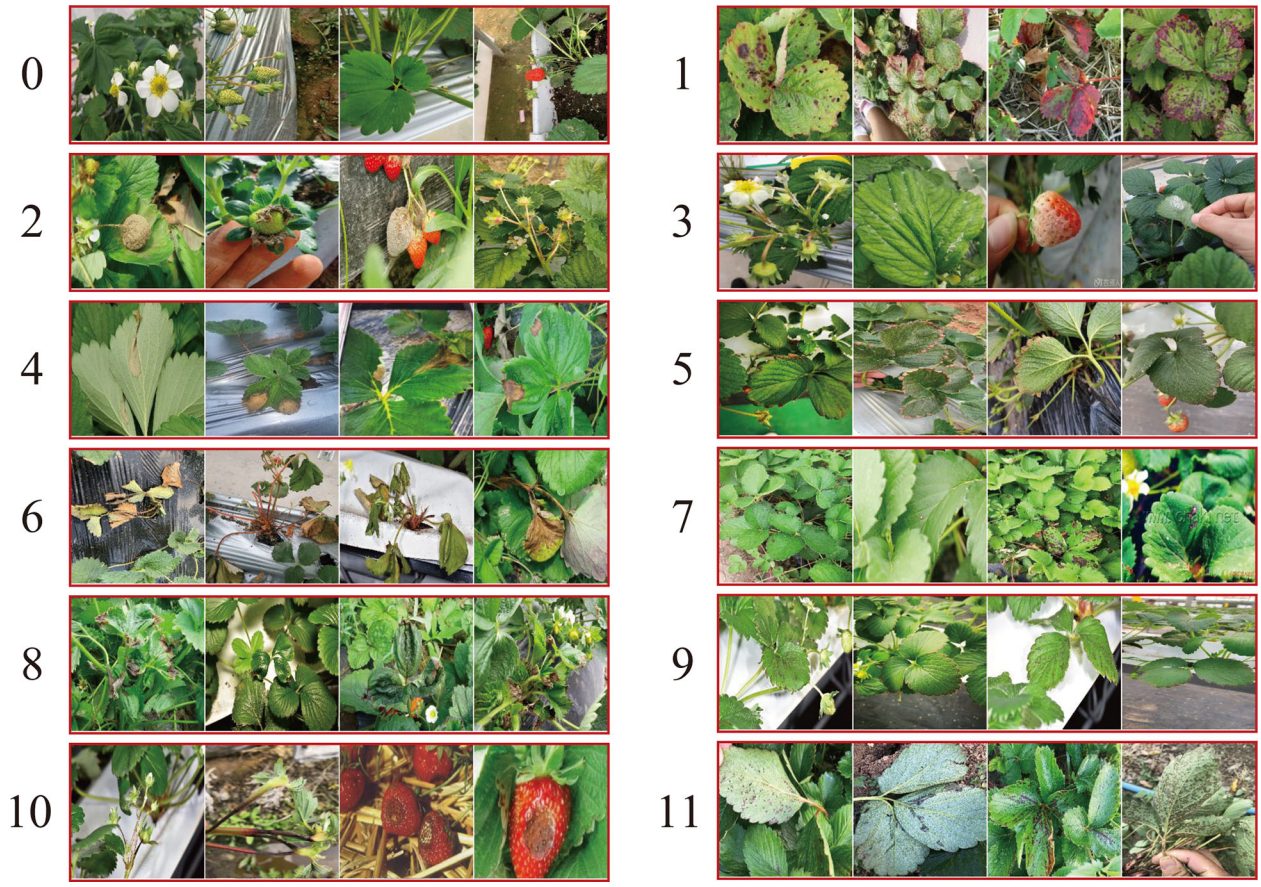


FIGURE 1
The typical symptoms of 11 common strawberry diseases and one healthy type. The annotated labels of the diseases are one-to-one correspondence with Table 1.

confidence of each disease category. Meanwhile, a class response map (CRM) module is constructed to generate a class attention matrix based on the region-based features. Here, the class attention matrix is defined as a class response map in this paper. Third, two different modules (COLM and LPPM) are developed to detect lesion regions based on the corresponding attention matrix from different scales of the input image, respectively. The COLM is used for locating the main lesion object in the image at coarse-grained level, while the LPPM proposes lesion details in the image at medium-grained level. Once an attended region is located, we segment the region and zoom in it to the raw image size. The located regions can be employed to generate a series of highly reliable lesion features. As a whole, the CALP-CNN takes advantage of ensemble learning to integrate the features from three scales for final identification. Moreover, the CALP-CNN combines an intra-scale cross-entropy loss and an inter-scale pairwise ranking loss to ensure rapid convergence.

2.2.1 Class response map

A series of class activation maps can be generated by the product of CNN feature maps with their corresponding class scores. The studies of (Zhou et al., 2016; Ding et al., 2019) have proved that the class-related information in the class activation maps is effective for locating discriminative regions in an image. In this paper, we obtain discriminative information of lesions based on the class activation maps and construct a class response map (also denoted as class attention matrix) to localize the objects of interest. Figure 3 shows the generation process of a class response map.

First, a pre-trained CNN backbone is applied to extract the feature maps of a 3-channel image $I \in R^{3\times H\times W}$, where the $H \times W$ is the spatial size of the image. The extracted feature maps are represented as $S \in R^{C\times H_f\times W_f}$, where C is the channel number and $^{Hf \times Wf}$ is the spatial size of the feature maps. Second, the feature maps S are fed forward to a classifier consisting of a fully connected (FC) layer and a softmax layer. A vector p∈R^{N_c}.(N_C is the pre-set category number of the strawberry diseases) can be computed by the classifier as the predicted probability score of each disease. In addition, the weights of the FC layer are denoted as $w_{fc} \in R^{C\times N_c}$. Third, a CRM module is designed to generate the class-related features maps. It establishes a new convolutional layer with the weight of the w_{fc} (i.e., the formed convolutional layer achieves the same weights as the FC layer). Therefore, it possesses a strong ability to extract class-related features. Based on the constructed convolutional layer, a set of class-related feature maps $Q=\{Q_i\}(Q_i \in R^{Hf\times Wf}, i=1,\ldots,N_c)$ can be generated from the extracted S. The Q_i represents the i-th channel. The features of the Q_i are most relevant to category i. In the training process, the CALP-CNN applies the ground truth label to select the most class-related feature map of the convolutional layer as the class response map. That is to say, if the image is annotated as category c, the class response map is Q_c. In the testing process, there is no ground truth label of the input image. Follow as (Ding et al., 2019), the CALP-

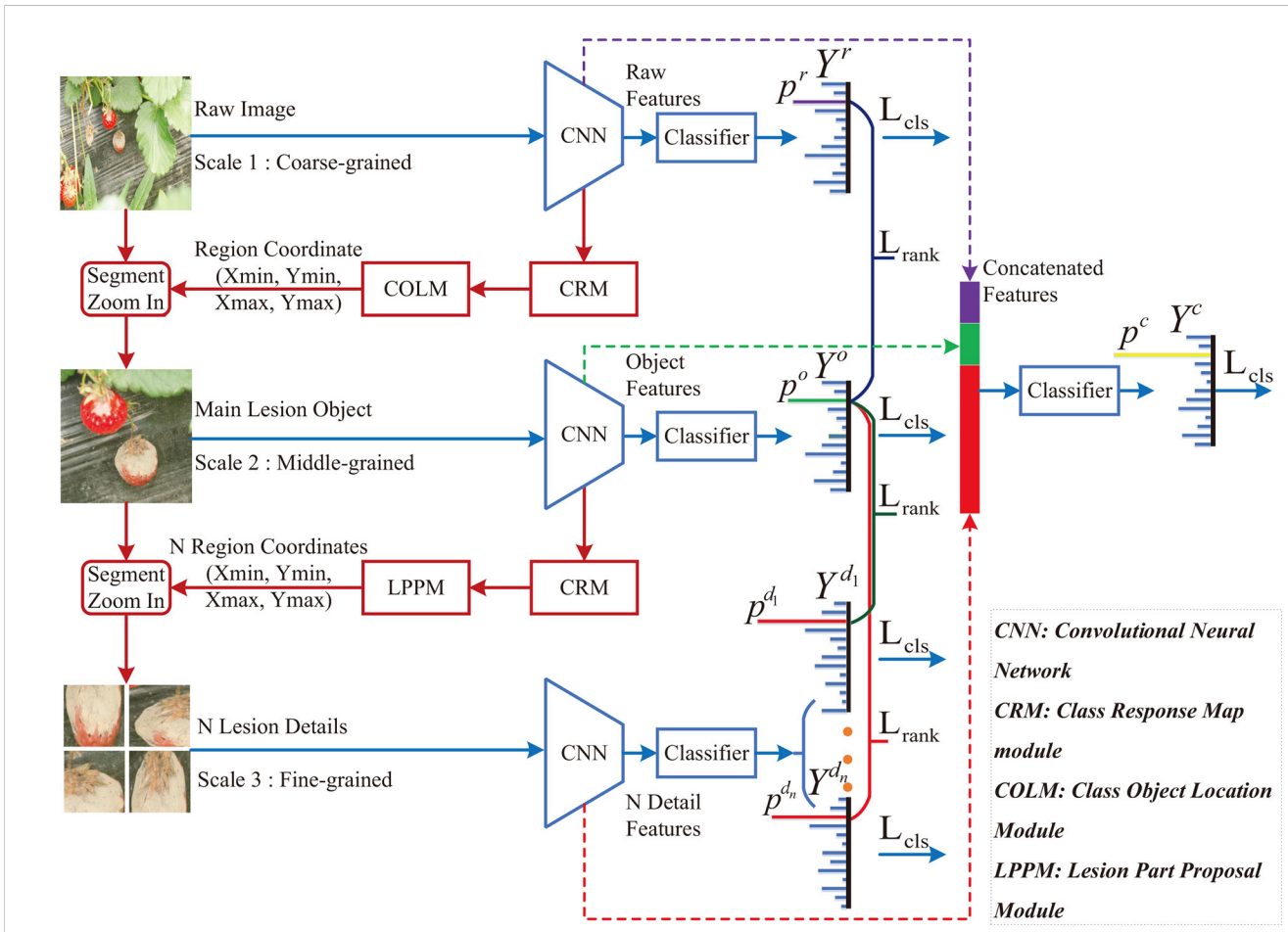


FIGURE 2
The framework of the proposed CALP-CNN. A cascade architecture is designed to construct the lesion details at different scales. A CNN-based backbone is repeatedly used to extract features from the coarse raw image to lesion detail images. The CRM module generates the class response map from the features. The COLM and the LPPM can obtain the coordinates of the lesion object and the lesion details, respectively. All features (the stripes marked with purple, green, and red) are concatenated for final identification. The classification loss L_{cls} (cross-entropy loss between ground truth label Y^* and predict label Y^r, Y^o, Y_i^d,Y^c) and the pairwise ranking loss L_{rank} (the loss between raw probability p^r, object probability p^o, and lesion probabilities p^{d_i}) are combined to optimize the network and make it converge.

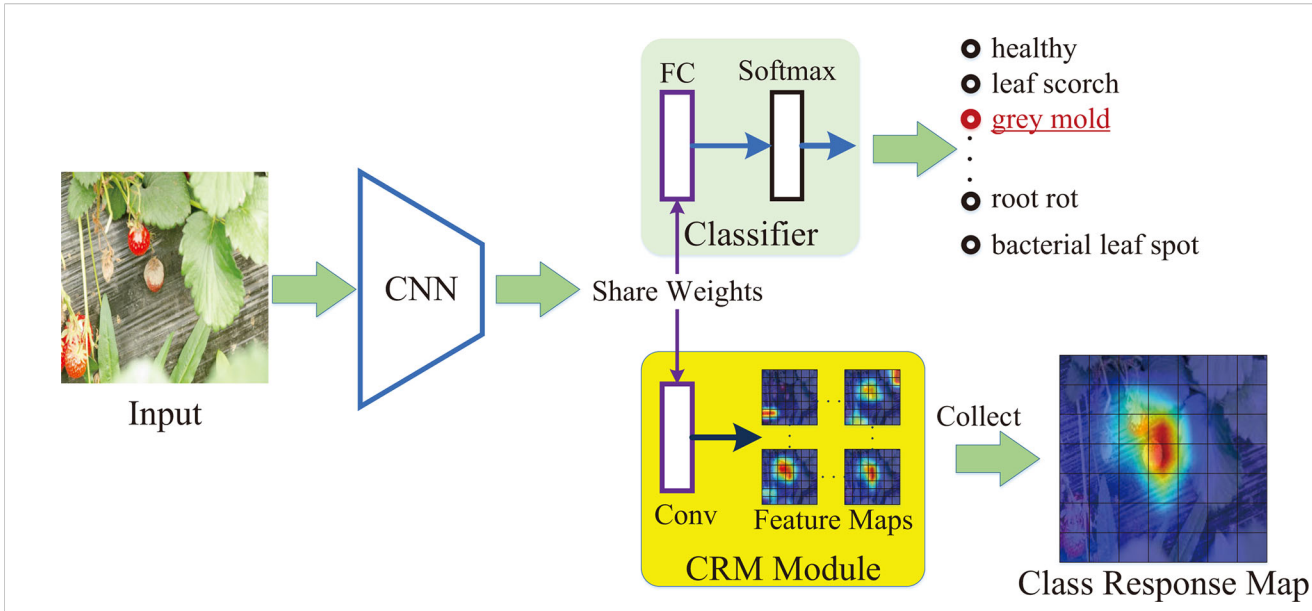


FIGURE 3
The generation process of class response map.

CNN adopts the entropy of the top 5 predicted probabilities to evaluate the lesion information in their corresponding class-related maps. Let $\hat{p} \in R^5$be the subset of p for top 5 predicted class probabilities. We compute the entropy as

$$H = -\sum_{i=1}^{5} p_i \cdot log p_i, \quad p_i \in \hat{p} \tag{1}$$

and construct the class response map Q_c based on the following strategy,

$$Q_c = \begin{cases} \widehat{M}_1, & if \ H \le \epsilon \\ \sum_{i=1}^{5} \widehat{M}_i, & otherwise \end{cases} \tag{2}$$

where $\hat{M} \in R^{5 \times H_f \times W_f}$ is the class-related feature maps correspond to $\hat{p}$ and ε is a threshold (empirically set to 0.2).

2.2.2 Class object location module (COLM)

In most cases, the CNN backbone could extract many irrelevant and noisy features that are adverse to disease identification, especially for a complex background (Barbedo, 2018). To cope with this issue, we design the COLM to locate the main lesion object and discard the irrelevant background region. This module is inspired by the discriminative region location methods of the fine-grained image classification and retrieval domain (Wei et al., 2017; Ding et al., 2019; Zhang et al., 2021). The pipeline of COLM is shown as Figure 4

The class response map Q_c is resized to the same size as the input image I by a bilinear interpolation algorithm. The interpolation result is denoted as $Q_c^{'} \in R^{H \times W}$. Ding et al. have concluded that the larger value in the class response map, the more related of the corresponding pixel to the class (Ding et al., 2019). In most cases, we have no prior knowledge about the location of the lesion objects since most crop disease datasets only have image-level supervision.

$$\bar{q} = \frac{\sum_{i=1}^{H} \sum_{j=1}^{W} Q_c^{'}(i,j)}{H \times W} \tag{3}$$

Then, a mask map M can be generated according to Eq.4.

$$M(i,j) = \begin{cases} 1, & if \ Q_c^{'}(i,j) > \bar{q} \\ 0, & otherwise \end{cases} \tag{4}$$

As shown in Figure 4, the object regions are marked red in the mask map. We can observe some noisy regions (the top-left and bottom-right) in the mask. In fact, the noisy regions could be non-lesion parts, whereas they are activated by the complex background. Fortunately, the sizes of the noisy regions are typically smaller than the main lesion object. Flood-fill algorithm is a common method to connect neighboring and related elements of a matrix. In this paper, we apply it to test the connectivity of all the points in M and find out the largest connected area. The largest connected area is the location of the main lesion object. The minimum enclosing rectangle of the largest connected area is denoted as M. We adopt the top-left point (x_{tl}, y_{tl}) and bottom-right point (x_{br}, y_{br}) to represent the location of $M = M\,[x_{tl}{:}x_{br}, y_{tl}{:}y_{br}]$. With the interpolation algorithm, the pixels in the mask map M are one-to-one corresponding to the pixels in the input image I. Therefore, the location of M can be used to extract the main lesion object and discard the noisy background in I. As a result, the main lesion object I_{obj} is computed as:

$$I_{obj} = I[x_{tl}, x_{br}, y_{tl} : y_{br}] \tag{5}$$

Based on the ablation experiments in section 4.2, the COLM module can effectively improve the classification accuracy.

2.2.3 Lesion part proposal module (LPPM)

Identifying similar diseases in the field is another critical problem for strawberry cultivation, especially for those diseases which have homologous backgrounds and subtle inter-class differences (e.g., the diseases at the early stage and the diseases occurring in the same part). Strengthening the differences between diseases is the key approach to address this issue (Cruz et al., 2019). The similar disease identification is in accord with the characteristics of the fine-grained image recognition (FGIR) (Zheng et al., 2017). The studies of FGIR have concluded that the discriminative features always lie in the details (Fu et al., 2017; Recasens et al., 2018; Ding et al., 2019; Zhang et al., 2021). Hence, we present the LPPM to localize the distinguishing lesion features in the details. The design idea of this module is derived from the region proposal algorithm (RPA) (Ren et al., 2015). The RPA is an effective method to propose candidate regions for object detection. The candidate region is called anchor in object detection. Nevertheless, the RPA requires an additional bounding box to annotate the location of the object. The bounding box annotation process is labor-intensive and subjective. Here, we take the average value of all pixels in the anchor as a confidence of whether the region in the anchor is a lesion detail. In this way, the RPA can be generalized to identify detailed lesions in the images without bounding box annotations.

The pipeline of LPPM is shown as Figure 5. The LPPM takes the output (i.e., class response map) of a CRM module as input. We denote it as $M_c \in R^{Hf \times Wf}$. First, the LPPM propose the coordinates of

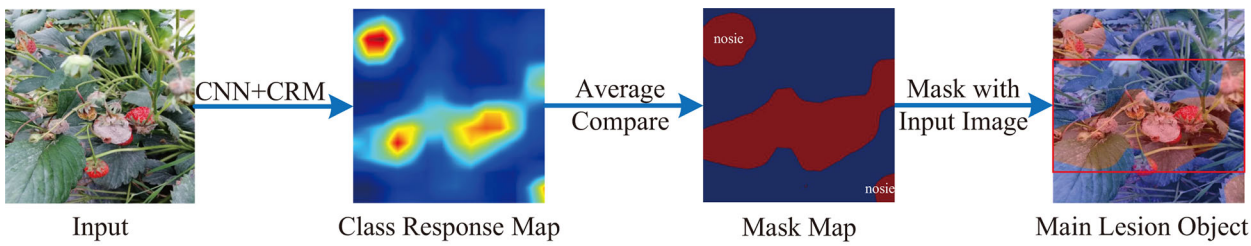


FIGURE 4
The pipeline of COLM. A class response map is generated from a CRM module. The pixels in the class response map are compared to their average value to generate a mask map. Some non-lesion areas are activated by the complex background in the mask map.

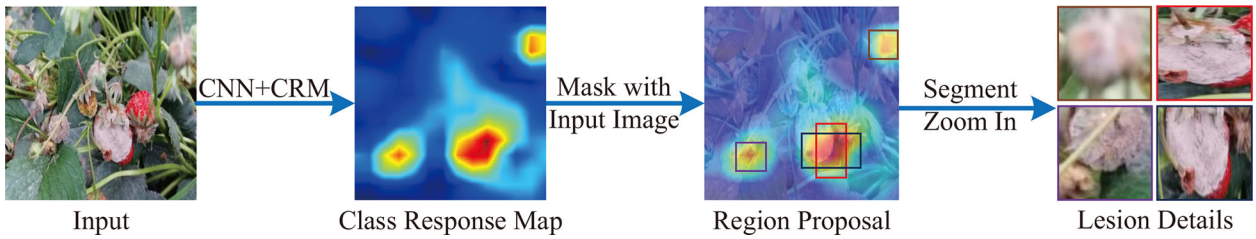


FIGURE 5
The pipeline of LPPM. First, a class response map is generated from a CRM module. Second, the RPA is applied to proposal candidate lesion regions from the class response map. Third, a non-maximum suppression is utilized to pick out the top-*N* lesions.

the anchors on M_c. By default, we use 3 aspect ratios (1:1, 2:1, 1:2) and 1 scale (H_f/2), yielding k=3 anchors at each pixel of M_c. The total number of generated anchors is $k \times H_f \times W_f$. Each anchor is an eligible candidate for the lesion detail. The coordinates of the anchors are denoted by their top-left point (x'_{tl}, y'_{tl}) and bottom-right point (x'_{br}, y'_{br}) Second, we calculate the average value of an anchor at M_c as follows:

$$\bar{a} = \frac{\sum_{i=x'_{\mathrm{tl}}}^{x'_{\mathrm{br}}} \sum_{j=y'_{\mathrm{tl}}}^{y'_{\mathrm{br}}} M_c(i,j)}{\left(x'_{\mathrm{br}} - x'_{\mathrm{tl}}\right) \times \left(y'_{\mathrm{br}} - y'_{\mathrm{tl}}\right)} \tag{6}$$

$\bar{a}$ is the confidence of the anchor to be a lesion detail region. A higher value of ā represents the higher probability of the anchor being a lesion detail. Third, we pick out the top-*N* anchors according to their confidence. In practice, the top-*N* anchors are adjacent and contain almost the same parts (Ren et al., 2015). For this reason, the directly selection of top-*N* anchors will cause information redundancy.

Input: The coordinate list of the anchors; The corresponding confidence list of the anchors; The IoU *threshold*, **Output:** The top-*N* anchor listCombined the confidence list and the coordinate list with an element as $[\bar{a}, x_{tl}, y_{tl}, x_{br}, y_{br}]$. The result is a confidence_coordinate_list; confidence_coordinate_list ← Sort the combined list in descending order with the confidence $\bar{a}$; anchor_list ← Initialize an empty list of selected anchors;**while** *Length (anchor_list) < N and Length (confidence_coordinate_list) > 0* **do**A←Pop out the first anchor element from the confidence_coordinate_list;**If** *anchor_list is empty* **then**Add *A* to the anchor_list;**else** Calculate the *IoU* between A and the other anchors in the anchor_list;**if** *IoU< threshold* **then** *IoU< threshold* Add *A* to the anchor_list;**return** *the anchor_list (is the top-N list);*

ALGORITHM 1.

In this paper, we use the intersection over union (IoU) to indicate the redundant ratio of two anchors. The IoU between anchor A_2 and anchor A_2 is computed as:

$$IoU = \frac{A_1 \cap A_2}{A_1 \cup A_2} \tag{7}$$

The IoU ratios between the anchor with the highest confidence and the other anchors are calculated. The scores of neighboring anchors will be suppressed when their IoU ratios are higher than the pre-set threshold. The threshold is set to 0.7 in this paper. The selection process of the top-*N* anchors is described in Algorithm 1 Note that N is a hyper-parameter which represents the defined number of lesion details. From the ablation experiments (see Section 4.2), the CALP-CNN achieves the best classification results when N is set to 5. Finally, we map the coordinate of the anchors in the top-*N* list to the input image I with the stride ($s = H/H_f$) of the backbone network. The location of the lesion I_{detail} is generated as:

$$I_{\mathrm{detail}} = I\left[s{\star}x_{tl} : s{\star}x_{br}, s{\star}x_{br} : s{\star}y_{br}\right] \tag{8}$$

2.2.4 Optimization strategy

The loss function of the proposed CALP-CNN is composed of two parts, including an intra-scale cross-entropy loss L_{cls} and an inter-scale pairwise ranking loss L_{rank}. The total loss function for an image I is defined as follows:

$$L(I) = L_{cls}(I) + L_{rank}(I) \tag{9}$$

The L_{cls} and L_{rank} are expressed in Eq. 10 and Eq. 11, respectively.

$$L_{cls}(I) = L_{cls}(Y^r, Y^\star) + L_{cls}(Y^o, Y^\star) + L_{cls}(Y^c, Y^\star) + \sum_{i=1}^{N} L_{cls}\left(Y^{d_i}, Y^\star\right) \tag{10}$$

where Y^r, Y^o, and Y^d are the predicted label vectors from the raw, object and detail images. Y^c is the predicted label vector using the concatenated features and $Y^\star$ is the ground truth label vector. N is the number of lesion details. L_{cls} is the chief loss function, which is dominant in the parameter optimization of the CALP-CNN.

$$L_{rank}(I) = L_{rank}(p^r, p^o) + \sum_{i=1}^{N} L_{rank}\left(p^o, p^{d_i}\right) \tag{11}$$

where p^r, p^o and p^d denote the prediction probabilities from the raw, object and detail images, respectively. To be specific, the ranking loss of the probabilities p^i and p^j is defined as:

$$L_{rank}(p^i, p^j) = max\{0, p^i - p^j + \delta\} \tag{12}$$

where δ is a constant (by default, δ=0.05). The ranking loss can force the object image to acquire higher predicted probabilities than

the original image. Meanwhile, the detail images are forced to acquire higher predicted probabilities than the object image. In other words, the L_{rank} takes a coarse prediction as reference and gradually compels the network toward more discriminative region by forcing the finer-scale images to achieve more confident predictions.

2.3 Evaluation metrics

In this paper, the *Accuracy*, *Precision*, *Recall*, and *F1-score* are adopted to evaluate the performance of the proposed CALP-CNN. The *Accuracy*, *Precision*, *Recall*, and *F1-score* of category *i* are defined as follows:

$$Accuracy_i = \frac{TP_i + TN_i}{TP_i + FP_i + TN_i + FN_i} \tag{13}$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \tag{14}$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \tag{15}$$

$$F1 - score_i = \frac{2Precision_i \cdot Recall_i}{Precision_i + Recall_i} \tag{16}$$

where TP_i and TN_i denote the number of samples labeled as category *i* and non-category *i* that are correctly classified, respectively. FP_i denotes the number of samples labeled as non-category *i* but classified as category *i*. FN_i denotes the number of samples labeled as category *i* but classified as non-category *i*.

For a multi-class classification task, the overall *Accuracy*, *Precision*, *Recall*, and *F1-score* can be defined with the average of all the categories in their binary classification case. The formulas of the overall *Accuracy*, *Precision*, *Recall*, and *F1-score* are defined as follows:

$$Accuracy = \frac{\sum_{i=0}^{N_c-1} Accuracy_i}{N_c} \tag{17}$$

$$Precision = \frac{\sum_{i=0}^{N_c-1} Precision_i}{N_c} \tag{18}$$

$$Recall = \frac{\sum_{i=0}^{N_c-1} Recall_i}{N_c} \tag{19}$$

$$F1 - score = \frac{\sum_{i=0}^{N_c-1} F1 - score_i}{N_c} \tag{20}$$

where the N_c represents the number of categories of strawberry diseases in the SCDD.

3 Experimental results and analysis

We conduct a series of experiments on the testing set of the SCDD to verify the effectiveness of the proposed CALP-CNN to identify strawberry diseases by filtering the complex background features and learning the discriminative features among similar diseases. The top-N of the anchors (lesion details) is set to 5 for the LPPM in our experiments.

Baselines: Because the CALP-CNN is an attention-based model and our SCDD only has image-level supervision, here we select six weakly-supervised fine-grained image recognition methods as baselines and compare their disease identification performance with the CALP-CNN method. The six baselines are described in detail as follows:

- MA-CNN (Zheng et al., 2017): Multi-attention convolutional neural network, which uses channel grouping to learn different part features.
- RA-CNN (Fu et al., 2017): Recurrent attention convolutional neural network, which recurrent learns the finer-scale features by an attention proposal network.
- MMAL-Net (Zhang et al., 2021): Multi-branch and multi-scale attention network, which utilizes a saliency map to locate the main object and propose discriminative parts.
- SSN (Recasens et al., 2018): A saliency-based sampling layer for a neural network that samples the raw image based on a saliency map with a non-uniform method.
- TASN (Zheng et al., 2019): Trilinear attention sampling network first uses a trilinear function to enhance saliency values, then samples the raw images with these enhanced values.
- S3N (Ding et al., 2019): Selective sparse sampling network, which captures diverse and fine-grained detail from the raw image based on a class response map with a selective sparse method.

All the baselines achieve state-of-the-art on their fine-grained datasets [e.g., CUB-200-2011 (Welinder et al., 2010), and FGVC Aircraft (Maji et al., 2013)].

Implementation details: The proposed CALP-CNN is implemented on the open-source package Pytorch (Paszke et al., 2019), which can flexibly implement various CNN-based models. A pre-trained ResNet-50 on the ImageNet dataset is used as the backbone for extracting the feature maps. For a fair comparison, all baselines are re-implemented with this backbone. We use the stochastic gradient descent (SGD) to optimize network parameters. All the models are trained for 60 epochs with a batch size of 16. The initial learning rate is set to 1e-3 and will be dropped by 10 at the 20-th and 40-th epoch. The momentum is set to 0.9 and the weight decay is set to 1e-4. The input images are preprocessed to size 224×224. All the experiments are performed on a dell T5820 computer workstation with NVIDIA GeForce RTX 3090 GPU and Intel Xeon W-2200 processor.

3.1 Classification results

We compare the performance of the proposed CALP-CNN with the baselines on the testing set of the SCDD. The classification results are shown in Table 2. The CALP-CNN achieves more accurate classification results on all metrics. The CALP-CNN significantly outperforms the backbone (ResNet-50) by 9.49% on the *F1-score*. The overall *F1-score* of the CALP-CNN is higher than the saliency-based

TABLE 2 The classification performance of different methods on the SCDD.

	Attention Mechanism	*F1-score*	*Accuracy*	*Precision*	*Recall*
ResNet-50 (He et al., 2016)	–	82.47	84.35	83.49	82.13
RA-CNN (Fu et al., 2017)	part attention	83.37	85.71	84.56	83.38
MA-CNN (Zheng et al., 2017)	channel attention	83.56	85.82	84.49	83.92
MMAL-Net (Zhang et al., 2021)	saliency attention	**85.44**	**87.11**	85.79	**85.47**
SSN (Recasens et al., 2018)	saliency attention	82.93	84.40	84.01	82.91
TASN (Zheng et al., 2019)	saliency attention	84.91	87.10	85.72	84.88
S3N (Ding et al., 2019)	class attention	85.41	86.70	**86.56**	84.72
CALP-CNN	class attention	**91.96**	**92.56**	**92.55**	**91.80**

The bold and underlined values indicate the highest and sub-optimal scores in the metric, respectively.

models, for example, 9.03% improvement for SSN, 7.05% improvement for TASN, and 6.52% improvement for MMAL-Net. Additionally, the proposed CALP-CNN is also superior to the recurrent attention method (RA-CNN), the channel grouping attention method (MA-CNN), and the class attention method (S3N). Specifically, it improves 8.59%, 8.4% and 6.55% compared with RA-CNN, MA-CNN, and S3N on *F1-score*, respectively. Note that the improvement of our proposed model is contributed by the introduction of the COLM and LPPM. The COLM can filter the noisy background features, while the LPPM provides discriminative lesion details.

3.2 Ablation experiments

In this paper, four ablation experiments are conducted to investigate the role of 1) different network branches, 2) lesion location methods (saliency map vs. class response map), 3) the number of lesion details, and 4) the ranking loss on field disease identification accuracy. The experiments show that the CNN with three branches and five lesion details (top-5) achieves the best performance. The best model is equipped with the class response map for lesion location and the ranking loss for model optimization.

3.2.1 Contribution of different branches

As shown in Figure 2, the CALP-CNN consists of three main branches, i.e., the raw branch (R-branch), the object branch (O-branch), and the (lesion) details branch (D-branch). In our experiments, we temporarily remove different branches to survey the contribution of each branch. The *F1-score* of the ablation experiments is recorded in Table 3. The following conclusions can be drawn: 1) The *F1-score* of the CALP-CNN with all branches (R+O+D) is 91.96%. It drops to 87.94% when omitting the O-branch. While it drops to 88.42% when the D-branch is removed. These results demonstrate that both the O-branch and the D-branch are capable of locating informative lesion regions. 2) The O-branch has the highest score (88.97%) among the three branches. It shows that the locating and segmenting operation of the class-related lesion object from the complex background can effectively eliminate the influence of the background on disease identification in the field. 3) The D-branch represents detailed information on lesions but does not yield the highest score among the three branches. It demonstrates that the discriminative lesion detail features are not all-inclusive for disease identification. Contextual information is also a key feature for disease identification. On the other hand, the D-branch could provide essential information to the other branches. The overall accuracy of the network features is improved from 83.92% to 87.94% in R+D branches setting and 87.08% to 91.21% in O+D branches setting, respectively. Furthermore, the D-branch can collect important lesion details for similar disease identification cases. 4) Note that the absence of the O-branch results in a bigger loss (4.02%, from 91.96% to 87.94%) than the D-branch (3.54%, from 91.96% to 88.42%), suggesting that removing the background features is critical for disease identification in the field. 5) The concatenated features of the three branches achieved the best performance. It indicates that the share of the object and the lesion detail features can enhance the lesion features and suppress the influence of background features. The disease surrounding context information of disease is preserved in the concatenated features.

3.2.2 Role of different location methods

We re-implement the COLM and LPPM with saliency-based attention (Zhang et al., 2021) to locate the main object and the

TABLE 3 The contribution of each branch.

Experimental Setting	R-branch(%)	O-branch(%)	D-branch(%)	Concatenation(%)
R branch	82.47	–	–	82.47
R+O branches	82.66	88.97	–	88.42
R+D branches	83.92	–	83.01	87.94
O+D branches	–	87.08	84.37	91.21
R+O+D branches	82.44	88.12	86.05	91.96

lesion details. The saliency map adopts a class-agnostic attention mechanism. Different from the saliency map, the class response map is a class-aware attention method. From Table 4, we can observe that the class-aware method has 5.57% higher scores than the class-agnostic method. It further demonstrates that the class-aware method can effectively localize class-related regions.

Number of lesion details: Ten experiments are performed to investigate the relationship between the classification result (*F1-score*) and the number of lesion details. As shown in Figure 6, the *F1-score* improves as the number of lesion details increases. However, the *F1-score* declines when the number of lesion details exceeds 5. It demonstrates that the disease classification performance is not positive to the number of lesion details. The underlying reason is that the contextual information is diluted in numerous detailed lesions.

3.2.3 Effect of ranking loss

To explore the impact of the ranking loss on classification results, we remove the ranking loss and only retain the cross-entropy loss to optimize the parameters of the CALP-CNN model. The best *F1-score* in 60 epochs training is 91.30%, which is 0.66% lower than the original model. The introduction of ranking loss could assist the two modules (COLM and LPPM) in localizing more discriminative regions.

TABLE 4 Comparison between different location methods.

	F1-score(%)	Comments
saliency map	86.39	class-agnostic attention
class response map	91.96	class-aware attention

3.3 Results of similar diseases identification

In practice, some of the diseases of strawberries perform similar visual appearance and contextual information, which could result in false identification among similar diseases. In order to evaluate the effectiveness of the proposed CALP-CNN for distinguishing these similar diseases, two kinds of similar strawberry diseases are chosen in the SCDD for experiments, including (1) the diseases at early stage, (2) the diseases occurring on fruits (e.g., gray mold, powdery mildew, anthracnose). We generate two sub-datasets corresponding to the two kinds of similar strawberry diseases. The disease samples from the two sub-datasets are shown in Figure 7.

The validation results of the trained CALP-CNN and the ResNet-50 on the two sub-datasets are recorded in Table 5. Both of the methods do not achieve the ideal identification performance. However, our CALP-CNN outperforms the ResNet-50 by 5.85% on disease at early stage dataset and 6.73% on disease on fruit dataset, respectively. Overall, the results suggest that the identification of similar strawberry diseases is challenging. While the discriminative

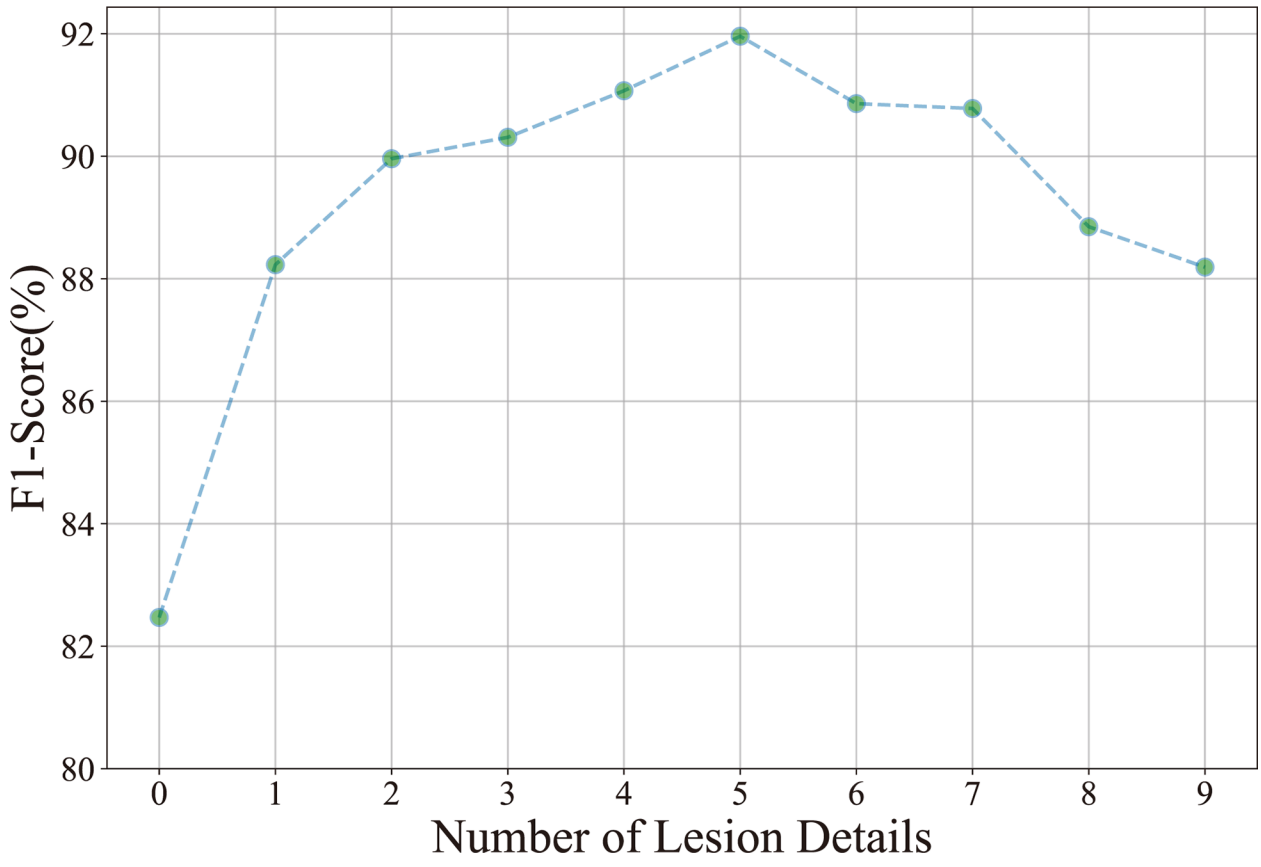


FIGURE 6
Relationship between the classification accuracy (*F1-score*) and the number of lesion details.

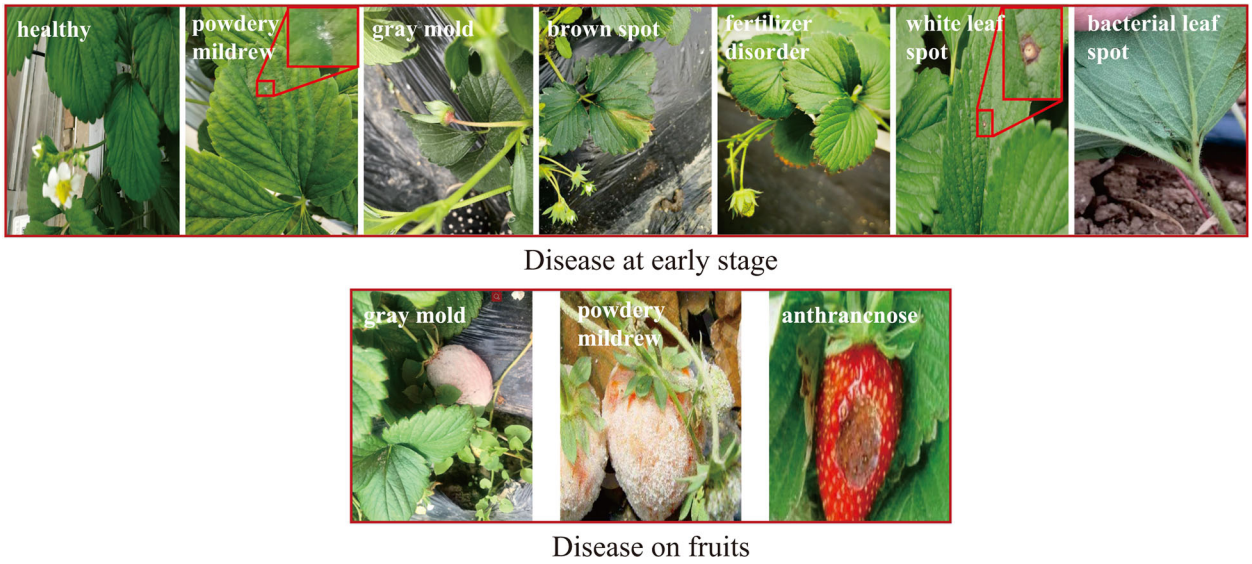


FIGURE 7
The examples of the similar diseases in the SCDD.

TABLE 5 The performance of the CALP-CNN and the ResNet-50 on the similar disease datasets.

Dataset	Amount/Categories	ResNet-50	CALP-CNN
early stage	324/10	59.87	65.72
on fruit	79/3	69.30	76.03

lesion detail features provide helpful information to improve the identification performance.

3.4 Qualitative evaluation of lesion localization performance

Because most of the strawberry datasets (including the SCDD) are image-level annotations. It is difficult to quantitatively evaluate the location accuracy of the main lesion object and the lesion details at the pixel-level. Here, we follow the study of (Wei et al., 2017) to conduct a qualitative evaluation to evaluate the accuracy of the main lesion object and lesion detail detection. We randomly pick out 3 groups of diseased images from the testing set for each strawberry disease and visualize the identification results of the lesions. The experimental results are shown in Figure 8. In Figure 8, the first column of each group is the input image, and the subsequent two columns are the location results of the main lesion object and lesion details of the image, respectively. Note that the images of lesion detail have been amplified to the same size as their input images. Based on the results of the main lesion objects, we can observe that the main lesion objects are all identified in the predicted bounding boxes of the COLM (group 1: 11/11, group 2: 11/11, group 3: 11/11). Furthermore, the predicted boxes contain contextual information by persevering the local background of the main lesion objects. In addition, most lesion details of the diseases can also be predicted by the LPPM (group 1: 54/55, group 2: 52/55, group 3: 55/55). In our experiments, the false predicted lesion areas occur in the images which have only one lesion area and the size of the lesion is relatively small (e.g., line 7, column 3 of group 2).

4 Discussions and conclusions

Existing methods for crop disease identification in the field are not sufficiently accurate because of their poor ability to eliminate the interference from the background and extract discriminative features among similar diseases. Detecting and segmenting the lesion region from the disease image is a simple yet effective way to reduce the influence of the complex background. Meanwhile, learning discriminative features from the lesion details is beneficial for the identification of similar diseases. The CNN-based semantic segmentation methods can effectively segment the lesion regions from the complex background. Hence, recent studies use semantic segmentation networks to segment lesion regions from the background as the first step of their models (Hu et al., 2021; Wang et al., 2021a). The segmentation performance of the networks highly relies on the amount of pixel-level annotated data. The pixel-level annotation is time-consuming, laborious and expensive, which restricts the applications of CNN-based segmentation methods. Besides, many studies have shown that the CNNs can localize discriminative regions from the input image (Selvaraju et al., 2017; Dabkowski and Gal, 2017; Wei et al., 2017; Ding et al., 2019). However, not all the located regions are useful for disease identification. The regions, which are activated by the complicated background, are adverse for disease identification (Barbedo, 2018). Therefore, it is necessary to filter out the most useful region from the located regions. The identification of similar diseases is also a challenging task. Because the discriminative details between the similar diseases are too subtle to be well-represented by the CNNs.

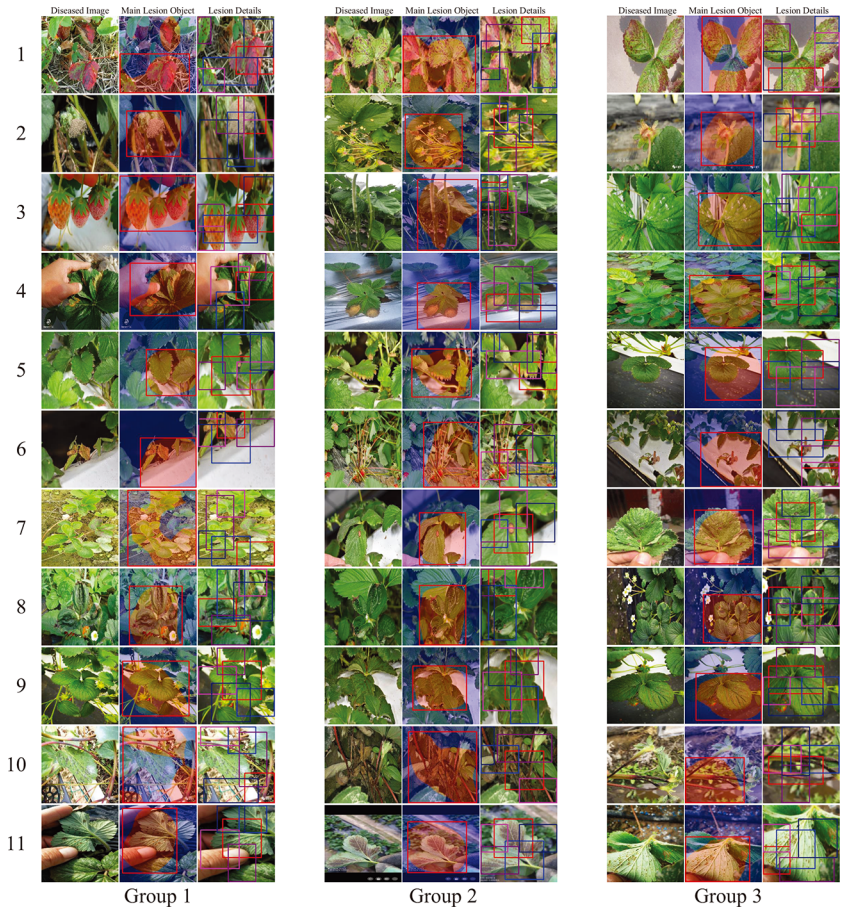


FIGURE 8
Identified main lesion object and lesion details. For each disease, we randomly select three samples from the testing set. The first column of each sample is the diseased image, and the subsequent two columns are the location results of the main lesion object and lesion details. The labels of the diseases are consistent with the Table 1.

Data augmentation technologies can increase the differences among similar diseases. Nevertheless, the increment is not obvious (Cruz et al., 2019). In addition, a suitable augmentation strategy is not straightforward and requires trial and error. Hence, data augmentation technologies are not an appropriate solution for similar disease identification. Fortunately, there are many similarities between crop similar disease identification and FGIR. The FGIR focuses on how to effectively represent the discriminative features between the subordinate classes (Ding et al., 2019). Therefore, the discriminative region localization and feature representation methods in FGIR can be extended to crop similar disease identification.

In this paper, we cite the field strawberry disease identification as our study object and explore innovative methods to address the challenges caused by the complex background and similar diseases. First, we enhance the ability of the CNN backbone to localize discriminative regions through a new class-attention-based mechanism (i.e., class response map). Second, we construct the COLM based on the flood-fill algorithm to filter out the most useful lesion region from the complex background. Third, we raise a new lesion part proposal method (i.e., the LPPM) to propose the discriminative lesion details based on the RPA. The COLM and LPPM are connected in series to form a Class-Attention-based Lesion Proposal Convolutional Neural Network (CALP-CNN), which can simultaneously address the challenges caused by complex background and similar diseases in field disease identification.

A series of experiments are conducted on the constructed field strawberry common disease dataset to testify the effectiveness of the CALP-CNN in eliminating the interference from the complicated background and distinguishing similar strawberry diseases. The classification result on *F1-score* reaches 91.96%, which is greatly higher than other methods, showing that the proposed model outperforms other state-of-the-art methods in the view of field strawberry disease identification. In addition, the ablation results on *F1-score* drop to 87.94% and 88.42%, respectively, when the COLM and LPPM branches in the CALP-CNN are removed. It indicates that both background feature elimination and discriminative lesion detail feature representation are indispensable for field disease identification.

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

XH, RW, JD, and TX conceived the idea and designed the network. XH, LJ, and YH contributed to collecting the dataset. XH wrote the code, validated the method, and wrote the paper. TX, JD and LJ revised the paper. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Key Research and Development Program of China-Intergovernmental International Scientific and Technological Innovation Cooperation (2019YFE0125700).

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

Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., and Stefanovic, D. (2019). Solving current limitations of deep learning based approaches for plant disease detection. *Symmetry* 11, 939-960. doi: 10.3390/sym11070939

Barbedo, J. G. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosyst. Eng.* 172, 84–91. doi: 10.1016/j.biosystemseng.2018.05.013

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 801-818. doi: 10.48550/arXiv.1802.02611

Cruz, A., Ampatzidis, Y., Pierro, R., Materazzi, A., Panattoni, A., De Bellis, L., et al. (2019). Detection of grapevine yellows symptoms in vitis vinifera l. with artificial intelligence. *Comput. Electron. Agric.* 157, 63–76. doi: 10.1016/j.compag.2018.12.028

Dabkowski, P., and Gal, Y. (2017). "Real time image saliency for black box classifiers," in Proceedings of the 31st International Conference on Neural Information Processing Systems, Vol. NIPS'17. 6970–6979 (Curran Associates Inc: Red Hook, NY, USA). doi: 10.48550/arXiv.1705.07857

Ding, Y., Zhou, Y., Zhu, Y., Ye, Q., and Jiao, J. (2019). "Selective sparse sampling for fine-grained image recognition," in Proceedings of the IEEE/CVF International Conference on Computer Vision (IEEE). 6599–6608. doi: 10.1109/ICCV.2019.00670

Dwivedi, P., Kumar, S., Vijh, S., and Chaturvedi, Y. (2021). "Study of machine learning techniques for plant disease recognition in agriculture," in 2021 11th International Conference on Cloud Computing, Data Science Engineering (Confluence) (IEEE). doi: 10.1109/Confluence51648.2021.9377186

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318. doi: 10.1016/j.compag.2018.01.009

Fu, J., Zheng, H., and Mei, T. (2017). "Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Los Alamitos, CA, USA: IEEE Computer Society), 4476-4484. doi: 10.1109/CVPR.2017.476

Hannum, S. M. (2004). Potential impact of strawberries on human health: A review of the science. *Crit. Rev. Food Sci. Nutr.* 44, 1–17. doi: 10.1080/10408690490263756

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE). 770–778. doi: 10.1109/CVPR.2016.90

Huang, K. Y. (2007). Application of artificial neural network for detecting phalaenopsis seedling diseases using color and texture features. *Comput. Electron. Agric.* 57, 3–11. doi: 10.1016/j.compag.2007.01.015

Hu, G., Wei, K., Zhang, Y., Bao, W., and Liang, D. (2021). Estimation of tea leaf blight severity in natural scene images. *Precis. Agric.* 22, 1239–1262. doi: 10.1007/s11119-020-09782-8

Iqbal, M., Jamshaid, M., Zahid, M. A., Andreasson, E., Vetukuri, R., and Stenberg, J. (2021). Biological control of strawberry crown rot, root rot and grey mould by the beneficial fungus aureobasidium pullulans. *BioControl* 66, 535-545. doi: 10.1007/s10526-021-10083-w

Iqbal, Z., Khan, M. A., Sharif, M., Shah, J. H., ur Rehman, ,. M. H., and Javed, K. (2018). An automated detection and classification of citrus plant diseases using image processing techniques: A review. *Comput. Electron. Agric.* 153, 12–32. doi: 10.1016/j.compag.2018.07.032

Johannes, A., Picon, A., Alvarez-Gila, A., Echazarra, J., Rodriguez-Vaamonde, S., Navajas, A. D., et al. (2017). Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Comput. Electron. Agric.* 138, 200–209. doi: 10.1016/j.compag.2017.04.013

Kaur, I., Aggarwal, G., and Verma, A. (2016). Detection and classification of disease affected region of plant leaves using image processing technique. *Indian J. Sci. Technol.* 9, 1-13. doi: 10.17485/ijst/2016/v9i48/104765

Kim, D. G., Burks, T. F., Qin, J., and Bulanon, D. M. (2009). Classification of grapefruit peel diseases using color texture feature analysis. *Int. J. Agric. Biol. Eng.* 2, 41–50. doi: 10.3965/j.issn.1934-6344.2009.03.041-050

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, vol. 25 . Eds. F. Pereira, C. Burges, L. Bottou and K. Weinberger (Curran Associates, Inc). doi: 10.1145/3065386

Liang, Q., Xiang, S., Hu, Y., Coppola, G., Zhang, D., and Sun, W. (2019). Pd2se-net: Computer-assisted plant disease diagnosis and severity estimation network. *Comput. Electron. Agric.* 157, 518–529. doi: 10.1016/j.compag.2019.01.034

Maji, S., Rahtu, E., Kannala, J., Blaschko, M. B., and Vedaldi, A. (2013). Fine-grained visual classification of aircraft. *ArXiv*. doi: 10.48550/arXiv.1306.5151

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01419

Ngugi, L. C., Abdelwahab, M., and Abo-Zahhad, M. (2020). Tomato leaf segmentation algorithms for mobile phone applications using deep learning. *Comput. Electron. Agric.* 178, 105788. doi: 10.1016/j.compag.2020.105788

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). "Pytorch: An imperative style, high-performance deep learning library," in *Advances in neural information processing systems*, vol. 32 . Eds. H. Wallach, H. Larochelle, A. Beygelzimer, F. D. Alché-Buc, E. Fox and R. Garnett (Curran Associates, Inc). doi: 10.48550/arXiv.1912.01703

Picon, A., Alvarez-Gila, A., Seitz, M., Ortiz-Barredo, A., Echazarra, J., and Johannes, A. (2019). Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Comput. Electron. Agric.* 161, 280–290. doi: 10.1016/j.compag.2018.04.002. BigData and DSS in Agriculture.

Recasens, A., Kellnhofer, P., Stent, S., Matusik, W., and Torralba, A. (2018). "Learning to zoom: a saliency-based sampling layer for neural networks," in Proceedings of the European Conference on Computer Vision (ECCV) (Springer International Publishing), 51–66. doi: 10.1007/978-3-030-01240-3\s\do5(4

Ren, S., He, K., Girshick, R., and Sun, J. (2015). "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, vol. 28 . Eds. C. Cortes, N. Lawrence, D. Lee, M. Sugiyama and R. Garnett (Curran Associates, Inc). doi: 10.1109/TPAMI.2016.2577031

Revathi, P. B., and Hemalatha, M. (2014). Cotton leaf spot diseases detection utilizing feature selection with skew divergence method. *Int. J. Sci. Eng. Technol.* 3, 22–30. Available at: https://www.ijset.com/publication/v3/005.pdf.

Ronneberger, O., Fischer, P., and Brox, T. (2015). *U-Net: Convolutional networks for biomedical image segmentation* (Springer International Publishing), 234–241. doi: 10.1007/978-3-319-24574-4\s\do5(2)8

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

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). "Grad-cam: Visual explanations from deep networks *via* gradient-based localization," in 2017 IEEE International Conference on Computer Vision (ICCV) (IEEE). 618–626. doi: 10.1109/ICCV.2017.74

Srdjan., S., Marko, A., Anderla, A., Dubravko, ,.u., and Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. *Comput. Intell. Neurosci* 3289801. doi: 10.1155/2016/3289801

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (IEEE), 1–9. doi: 10.1109/CVPR.2015.7298594

Too, E. C., Yujian, L., Njuki, S., and Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* 161, 272–279. doi: 10.1016/j.compag.2018.03.032. BigData and DSS in Agriculture.

Wang, Z., Di, S., Qi, P., Xu, H., Zhao, H., and Wang, X. (2021b). Dissipation, accumulation and risk assessment of fungicides after repeated spraying on greenhouse strawberry. *Sci. Total Environ.* 758, 144-153. doi: 10.1016/j.scitotenv.2020.144067

Wang, C., Du, P., Wu, H., Li, J., Zhao, C., and Zhu, H. (2021a). A cucumber leaf disease severity classification method based on the fusion of deeplabv3+ and u-net. *Comput. Electron. Agric.* 189, 106373. doi: 10.1016/j.compag.2021.106373

Wei, X.-S., Luo, J.-H., Wu, J., and Zhou, Z.-H. (2017). Selective convolutional descriptor aggregation for fine-grained image retrieval. *IEEE Trans. Image Process.* 26, 2868–2881. doi: 10.1109/TIP.2017.2688133

Welinder, P., Branson, S., Mita, T., Wah, C., and Perona, P. (2010). *Caltech-Ucsd birds 200* (california institute of technology). Available at: https://authors.library.caltech.edu/27468/1/WelinderEtal10_CUB-200.pdf.

Yang, G. F., Yang, Y., Zi-Kang, H. E., Zhang, X. Y., and Yong, H. E. (2022). A rapid, low-cost deep learning system to classify strawberry disease based on cloud service. *J. Of Integr. Agric.* 21, 460–473. doi: 10.1016/S2095-3119(21)63604-3

Zeiler, M. D., and Fergus, R. (2014). "Visualizing and understanding convolutional networks," in *Computer Vision – ECCV 2014* (Cham: Springer International Publishing), 818–833. doi: 10.1007/978-3-319-10590-1\s\do5(5)3

Zhang, F., Li, M., Zhai, G., and Liu, Y. (2021). "Multi-branch and multi-scale attention learning for fine-grained visual categorization," in International Conference on Multimedia Modeling. 136–147 (Springer). doi: 10.48550/arXiv.2003.09150

Zhao, S., Liu, J., and Wu, S. (2022). Multiple disease detection method for greenhouse-cultivated strawberry based on multiscale feature fusion faster r-cnn. *Comput. Electron. Agric* 199, 107176. doi: 10.1016/j.compag.2022.107176

Zheng, H., Fu, J., Mei, T., and Luo, J. (2017). "Learning multi-attention convolutional neural network for fine-grained image recognition," in *2017 IEEE International Conference on Computer Vision (ICCV)* (IEEE), 5219–5227. doi: 10.1109/ICCV.2017.557

Zheng, H., Fu, J., Zha, Z.-J., and Luo, J. (2019). "Looking for the devil in the details: Learning trilinear attention sampling network for fine-grained image recognition," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE), doi: 10.1109/CVPR.2019.00515

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). "Learning deep features for discriminative localization," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE), 2921–2929. doi: 10.1109/CVPR.2016.319

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 07 February 2023
DOI 10.3389/fpls.2022.1009727

OPEN ACCESS

EDITED BY
Shizhuang Weng,
Anhui University, China

REVIEWED BY
Alireza Sanaeifar,
Zhejiang University, China
Yujie Wang,
Anhui Agricultural University, China

*CORRESPONDENCE
Furong Xu
xfrong99@163.com
Yuanzhong Wang
boletus@126.com

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

RECEIVED 02 August 2022
ACCEPTED 28 November 2022
PUBLISHED 07 February 2023

CITATION
Liu C, Zuo Z, Xu F and Wang Y (2023) Study of the suitable climate factors and geographical origins traceability of *Panax notoginseng* based on correlation analysis and spectral images combined with machine learning.
Front. Plant Sci. 13:1009727.
doi: 10.3389/fpls.2022.1009727

Study of the suitable climate factors and geographical origins traceability of *Panax notoginseng* based on correlation analysis and spectral images combined with machine learning

Chunlu Liu[1,2], Zhitian Zuo[1], Furong Xu[2*] and Yuanzhong Wang[1*]

[1]Medicinal Plants Research Institute, Yunnan Academy of Agricultural Sciences, Kunming, Yunnan, China, [2]Collge of Traditional Chinese Medicine, Yunnan University of Chinese Medicine, Kunming, Yunnan, China

Introduction: The cultivation and sale of medicinal plants are some of the main ways to meet the increased market demand for plant-based drugs. *Panax notoginseng* is a widely used Chinese medicinal material. The growth and accumulation of bioactive constituents mainly depend on a satisfactory growing environment. Additionally, the occurrence of market fraud means that care should be taken when purchasing.

Methods: In this study, we report the correlation between saponins and climate factors based on high performance liquid chromatography (HPLC), and evaluate the influence of climate factors on the quality of *P. notoginseng*. In addition, the synchronous two-dimensional correlation spectroscopy (2D-COS) images of near infrared (NIR) data combined with the deep learning model were applied to traceability of geographic origins of *P. notoginseng* at two different levels (district and town levels).

Results: The results indicated that the contents of saponins in *P. notoginseng* are negatively related to the annual mean temperature and the temperature annual range. A lower annual mean temperature and temperature annual range are favorable for the content accumulation of saponins. Additionally, high annual precipitation and high humidity are conducive to the content accumulation of Notoginsenoside R1 (NG-R1), Ginsenosides Rg1 (G-Rg1), and Ginsenosides Rb1 (G-Rb1), while Ginsenosides Rd (G-Rd), this is not the case. Regarding geographic origins, classifications at two different levels could be successfully distinguished through synchronous 2D-COS images combined with the residual convolutional neural network (ResNet) model. The model accuracy of the training set, test set, and external validation is achieved at 100%,

and the cross-entropy loss function curves are lower. This demonstrated the potential feasibility of the proposed method for *P. notoginseng* geographic origin traceability, even if the distance between sampling points is small.

Discussion: The findings of this study could improve the quality of *P. notoginseng*, provide a reference for cultivating *P. notoginseng* in the future and alleviate the occurrence of market fraud.

KEYWORDS

***Panax notoginseng*, active components, climate factors, synchronous 2D-COS images, deep learning model, geographical traceability**

Introduction

Because of the high price of precious Chinese medicinal materials, some criminals often blend pure substances with less expensive materials in order to earn illegal profits (Liu et al., 2019; Ichim and de Boer, 2021). Recently, as a popular medicinal material for the treatment and prevention of diseases and for keeping healthy, *Panax notoginseng* has also been affected by the same situation. A large variety of *P. notoginseng* is sold on the market, which results in some illegal traders mixing the cheap and sub-quality materials with the genuine product, the inferior with the superior (Yao et al., 2021; Yu et al., 2022; Cui et al., 2022). As a common Chinese medicinal material for alleviating blood stasis, hemostasis, swelling, and pain relief, *P. notoginseng* is found in the dried roots of *Panax notoginseng* (Burk) *F. H. Chen* of the Araliaceae family. It is especially suitable for patients with hypertension, hyperlipidemia, hyperglycemia, heart and cerebrovascular diseases, and patients who have low immunity, anemia, and are prone to falling and sprains (Hawthorne et al., 2022; Jiang et al., 2022; Zheng et al., 2022). Additionally, it improves blood circulation, moisturizes the skin, and slows down aging (Peng et al., 2017; Teseo et al., 2021). Phytochemical and pharmacological studies of *P. notoginseng* have demonstrated that its main biologically active components are dammarane-type saponins consisting of protopanaxadiol and protopanaxatriol glycosides (Qiao et al., 2018; Marianela et al., 2021).

The composition of *P. notoginseng* in nature is complex and is highly related to the cultivation years, processing methods, geographical origin, etc. (Wang et al., 2012; Bai et al., 2021; Wan et al., 2021; Zhang et al., 2021). The collection location points of samples are an important factor of geographical origin, which may be related to the content of active components and the market price of *P. notoginseng*. Therefore, it is important to be able to trace the origins of *P. notoginseng*. Climate factors (temperature, light, rainfall) in different geographical collection location points are some of the main factors that cause quality changes in medicinal plants (Liu et al., 2022). Every type of Chinese medicinal material has its own growth preferences, which results in different suitable growth areas. Therefore, it is key to analyze the correlation between the content accumulative of active components and climate factors, and comprehensively assess the influence of climate factors on the quality of *P. notoginseng*. In addition, Yunnan province (especially Wenshan Prefecture) is one of the important growth and export geographic origins of *P. notoginseng*. To prevent confusion about the contents of the *P. notoginseng* that is on the market, it is essential to develop a simple and quick method of geographical origin traceability.

In recent studies, methods have been reported for the traceability of the geographic origins of *P. notoginseng*, such as sensory analysis (macroscopic and microscopic), inductively coupled plasma mass spectrometry (ICP-MS), electronic tongue or electronic nose, and isotope (Tian et al., 2021; Liang et al., 2021; Ji et al., 2022). However, these methods have some disadvantages, such as large variation and subjectivity (sensory analysis), and being expensive, complex, time-consuming, and labor-intensive. Infrared (IR) spectroscopy has the advantages of being rapid, simple, and pollution-free. It has occupied a unique position in the analytical field since its creation, which illustrates its capabilities. With the continuous development of modern technology and the increasing demand for quality detection, IR technology has been widely applied in the research of Chinese medicinal materials (Li et al., 2018; Zhou et al., 2020), food (Wildea et al., 2019; Candoğan et al., 2021), biology (Huber et al., 2021; Kirschbaum et al., 2021), chemistry (Cura et al., 2021; Mishra et al., 2021), and other fields. Among its applications, an IR-based approach to understand the complex composition of *P. notoginseng*, where chemometrics and machine learning models have been developed, has gained great popularity in terms of the possibility of authenticating and tracing the origins of *P. notoginseng*. However, traditional one-dimensional (1D) linear spectra may not be specific enough and can create overlaps of data, which can limit the amount of useful information

extracted from data. Being more versatile, two-dimensional correlation spectroscopy (2D-COS) could be used to overcome this drawback and extract useful information from a series of spectra under chemical or physical stimuli (Noda, 1989; Noda, 1990). On the other hand, with the improvement in data processing and analysis, deep learning has become a promising research algorithm for the qualitative detection of Chinese medicinal materials, and it could be used as an auxiliary method for the study of 2D-COS images (Lecun et al., 2015; Jogin et al., 2018). Compared with other methods to trace geographic origins, 2D-COS images combined with the deep learning model do not require complex procedures, such as data processing and feature extraction. 2D-COS is more focused on processing problems of simple digital images, which are easier, faster, and more representative than analyzing complex spectral data itself.

In the past, several reports have studied the traceability of the geographic origins of *P. notoginseng* geographic origins. For example, Bai et al. (2021) generated high performance liquid chromatography (HPLC) characteristic fingerprints of *P. notoginseng* extract samples by a multi-wavelength fusion profiling (MWFP) method. They used the averagely linear quantified fingerprint method (ALQFM) and an unsupervised statistical method based on fusion fingerprint matching to identify the geographical origins of *P. notoginseng*. Chen et al. (2018) preprocessed through standard normal variables (SNV) and first derivative (FD) for near infrared (NIR) spectra and established a partial least-squares discriminant analysis (PLS-DA) model to quickly identify the geographic origins of *P. notoginseng*. Similarly, Zhou et al. (2020) carried out a single-spectrum analysis and multi-sensor information fusion strategy for Fourier transform mid-infrared (FT-MIR) and NIR data combined with the multivariate classification algorithm to successfully identify the geographic origins of *P. notoginseng*. In contrast, another study used ultraviolet-visible (UV-Vis) spectrophotometry, Fourier transform infrared (FT-IR) spectrum and HPLC combined with chemometrics to determine the total flavonoid content of *P. notoginseng* from different geographic origins. The total flavonoid content was analyzed and predicted by the standard linear equation of rutin and the orthogonal signal corrected partial least squares regression (OSC-PLSR) model, respectively (Li et al., 2017). Meanwhile, some articles have studied the influence of ecological factors on the growth of *P. notoginseng*. For example, He et al. (2016) applied fingerprints of stable oxygen isotope to study the "Dao-di" authenticity of *P. notoginseng* and trace its geographical origins. The dominant ecological factors and their weights affecting the taproot $\delta^{18}O$ of *P. notoginseng* were studied through correlation analysis, stepwise regression analysis, partial correlation analysis, and path analysis. A total of 16 main ecological factors affecting the taproot $\delta^{18}O$ of *P. notoginseng* were screened from 49 ecological factors, and the size, direction, decisive factors, restrictive factors, and the dominant factor were analyzed. Additionally, Yue et al. (2022) proposed the theory of *P. notoginseng* regionalization modeling. The ecological suitability of *P. notoginseng* under current and future climates was analyzed by the maximum entropy model (MaxEnt). The study found that the current most suitable habitat for *P. notoginseng* was mainly located in southwest China. Global climate change is not conducive to the development of *P. notoginseng* planting, and climate warming may lead to serious shrinkage of the growth areas of *P. notoginseng*. Considering future climate change, Yunnan Province was still the most suitable habitat area for *P. notoginseng*, and Sichuan Province was an important potential suitable habitat area. The research provided a new perspective on the ecological suitability of other medicinal plants in the southwest mountainous area. Nevertheless, none of these reports were based on HPLC to analyze the correlation between active component content accumulation and climate factors to alleviate the influence of climate factors on the quality of *P. notoginseng*. In addition, there are no reports of using 2D-COS images of NIR data combined with deep learning models to trace the geographic origin of *P. notoginseng* at the levels of district and town.

In this study, to ensure authenticity and traceability, all *P. notoginseng* samples were collected from cultivation bases. HPLC combined with the principal component analysis (PCA) model was used to analyze the differences of *P. notoginseng* between different districts and towns. Correlation analysis and a partial least squares regression (PLSR) model were constructed to research the correlation between the content accumulation of the main components and climate factors of *P. notoginseng* and to analyze the effect of climatic factors on the variation of saponin content under different growth environments. On this basis, in order to prevent the alteration of the product on the market and associated consumer confusion, the geographic origin traceability of *P. notoginseng* from different district levels was further explored by converting raw spectral data into 2D-COS images combined with the ResNet model. In addition, the reliability of the model was verified by identifying the geographic origin of *P. notoginseng* samples from different town levels. The findings of this study could improve the quality of *P. notoginseng*, provide a reference for cultivating *P. notoginseng* in the future, and alleviate the phenomenon of market fraud.

Materials and methods

Sample information

As the main objective of the present study was to evaluate the quality of P. notoginseng under the influence of different environmental factors and discrimination the geographical origins. The sampling points were selected from more dispersed

locations to be more representative and to allow an analysis of environmental factors. Therefore, the geographical origins were divided into four parts: DDB (Northeastern Yunnan), DDN (Southeast Yunnan), DX (Western Yunnan), and DZ (Central Yunnan). In addition, considering the Yunnan Province, especially Wenshan Prefecture is the main geographical origin of P. notoginseng, it is more meaningful and representative for analysis. Therefore, four town-level samples from Wenshan Prefecture, Yunnan Province were selected for analysis and validation, respectively YS (Yanshan, Wenshan Prefecture), XC (Xichou, Wenshan Prefecture), MG (Maguan, Wenshan Prefecture) and QB (Qiubei, Wenshan Prefecture).

A total of 229 *P. notoginseng* samples were collected from the cultivation base of Yunnan province, which meant the authenticity and traceability of the sample could be guaranteed. The altitude ranged from 1150 to 2382 m a.s.l. Detailed sample information of the geographical origins, collection locations, and the corresponding amount have been demonstrated in Figure 1 and Table S1. The collected samples were cleaned with tap water. The different parts were divided and dried at 50°C, then weighed and recorded. Among them, part of the main roots used as the main research object of this research was pulverized and passed through 90 mesh screen. All samples were packaged and labeled in zip-lock bags and stored at room temperature for further use.

Chemicals reagents

All methanol and acetonitrile used for HPLC analysis were of HPLC grade, and the other chemicals were of analytical grade. Notoginsenoside R1 (NG-R1), Ginsenosides Rg1 (G-Rg1), Ginsenosides Rb1 (G-Rb1), and Ginsenosides Rd (G-Rd) were supplied by the China Institute of Food and Drug Verification (Beijing, China). The UPTL-II-40L system (Chengdu, China) was applied for water purification.

Climate factors sources

The climate factors (bioclimatic variables and elevation information) were bioclimatic variable layers, which included Bio 1-Bio 19 and a spatial resolution of 30 s. These variables were downloaded from WorldClim (https://www.worldclim.org/), and the detailed information has been shown in Table S2. The data (".tif" format) were opened in ArcGIS 10.6 software, and the climate factor indicators corresponding to the GPS coordinates of *P. notoginseng* at different sampling points were extracted by the "Sampling" tool.

Reference climate factors and chemical analysis

Screening for climate factors

The 19 climate factors may correlate with each other. To avoid co-linearity among these climate factors, a Pearson autocorrelation analysis of the 19 climate factors was conducted by the SPSS 20.0 statistical program. In this study, climate factors that were higher than correlation coefficient (|R|

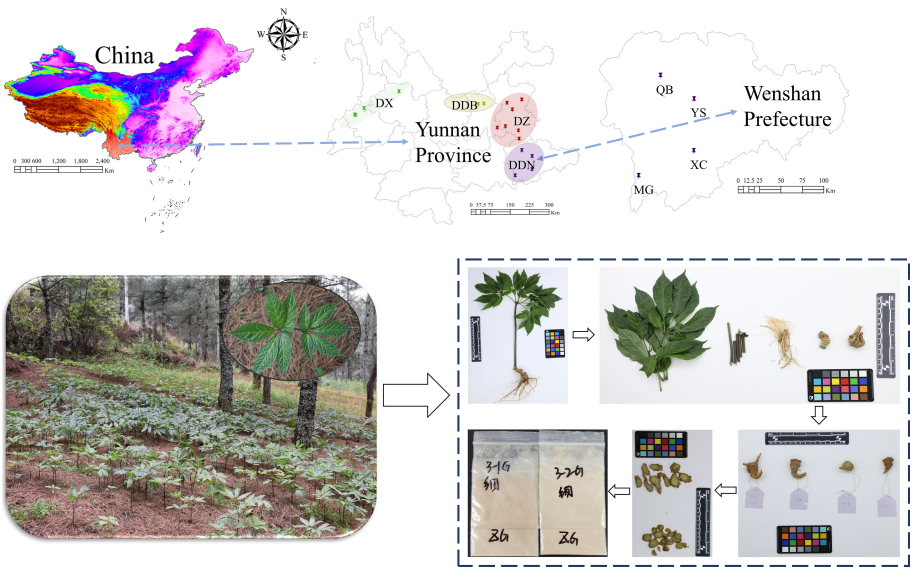


FIGURE 1
Detailed sample information of the geographical origins, collection locations, and picture of *P. notoginseng*.

>0.8) and less significant for the distribution of *P. notoginseng* were excluded.

Determination of climate factors weights

Principal component analysis (PCA) is an exploratory data analysis technique that uses a smaller number of principal components to represent changes in data sets. The original variables of centered and scaled may be on different measurement scales. After the orthogonal transformation of the normalization, principal components (PCs) were calculated as linear combinations of the original variables. The first PC accounts for more variance and the remaining PC for most likely to occupy the variance not covered by the first PCs. In general, the value of the cumulative variance should be greater than 80% (variance criterion) to be meaningful (Margaritis et al., 2020). PLSR analysis is a multivariate linear regression method that could provide information on the correlation structure of variables and structural similarity or dissimilarity. It can be used to discover correlation models between predictor variables and evaluate the response variables on an equal number of samples. In this study, the variables with more influence in the corresponding models were selected by PLSR (Farrés et al., 2015). The variable importance in projection (VIP) selection method can summarize the effect of each *X* variable on the PLS model and select the variable that contributes most to the explanation of *y* variance. In general, the VIP scores were greater than 1 (the average of the squared), which indicates that the variable makes a significant contribution to the model (Tran et al., 2014). In this study, PCA was used to classify the four saponins in *P. notoginseng*. The linear regression equations of NG-R1, G-Rg1, G-Rb1, G-Rd, and the selected climate factors were established by the PLSR method. Then, according to the linear regression equation, the normal distribution plots of VIP value were obtained. The VIP index value was normalized as the weight coefficient of each climate factor.

HPLC analysis

The Shimadzu Nexera LC-40 (Kyoto, Japan) device was equipped with an LC-40 binary pump, the SIL-40 automatic sampling device was connected to an SPD-M40 detector, and a Shim-pack VP-ODS column (250 × 4.6 mm, 5 μm) was applied. The mobile phase contained A (water) and B (acetonitrile). The gradient program was as follows: 0-5 min, 20% B; 5-10 min, 20%-25% B; 10-20 min, 25%-28% B; 20-30 min, 28%-30% B; 30-40 min, 30%-36% B; 40-45 min, 36%-40% B; 45-55 min, 40%-45% B; 55-60 min, 45%-90% B; 60-65 min, 90%-20% B; 65-70 min, 20% B. The injection volume for each sample was 10 μL, and the flow rate was 1 mL/min. After each run was balanced (maintain) every 10 min. The column temperature was set at 33-35°C, and the results were monitored at 203 nm. The methodology (linearity ranges, stability, repeatability, precision, and spiked recovery) was investigated by referring to the 2020 edition of the guiding principles of *Chinese Pharmacopoeia* (National Pharmacopoeia Committee, 2020).

Spectra acquisition

The NIR spectrometer (Thermo Fisher Scientific INC., USA) equipped with a diffused reflection mode was used to measure the spectra of *P. notoginseng*. The sample was placed into a sample cup (confirmed to be radiopaque), and the scanning range was 10000-4000 cm^{-1}. The acquisition parameters of each spectrum were scanned 64 times with a resolution of 4 cm^{-1}. Each collection was collected twice, and the average spectra were taken for analysis. In addition, it is worth noting that the spectra were corrected by collecting the background to remove atmospheric interference information.

2D-COS images acquisition and ResNet model establishment

The 2D-COS is a perturbation-based method first proposed by Noda. In this study, we extend it to generalized 2D-COS image analysis based on 2D spectral theory and literature references (Yang et al., 2013; Yang et al., 2014; Yang et al., 2015; Yang et al., 2020). The variable-variable correlation spectroscopy can set any kind of perturbation variables, such as temperature, concentration, pressure, and time. Synchronized 2D-COS images have sharper characteristic peaks for better characterization of different types of images (Dong et al., 2021). When measuring spectra with equal perturbation intervals t in steps m, dynamic spectral intensity was represented as a column vector S at variable v, it was defined as the following:

$$S(v) = \begin{bmatrix} y(v,t_1) \\ y(v,t_2) \\ y(v,t_3) \\ \vdots \\ y(v,t_m) \end{bmatrix} \quad (1)$$

The synchronous two-dimensional correlation intensities between variables v_1 and v_2 are calculated as $\Phi(v_1, v_2)$.

$$\Phi(v_1,v_2) = \frac{1}{m-1} S(v_1)^T \cdot S(v_2) \quad (2)$$

According to the full-band 2D-COS images (Figure 2), the bands of 7000-4000 cm^{-1} fingerprint area were selected for

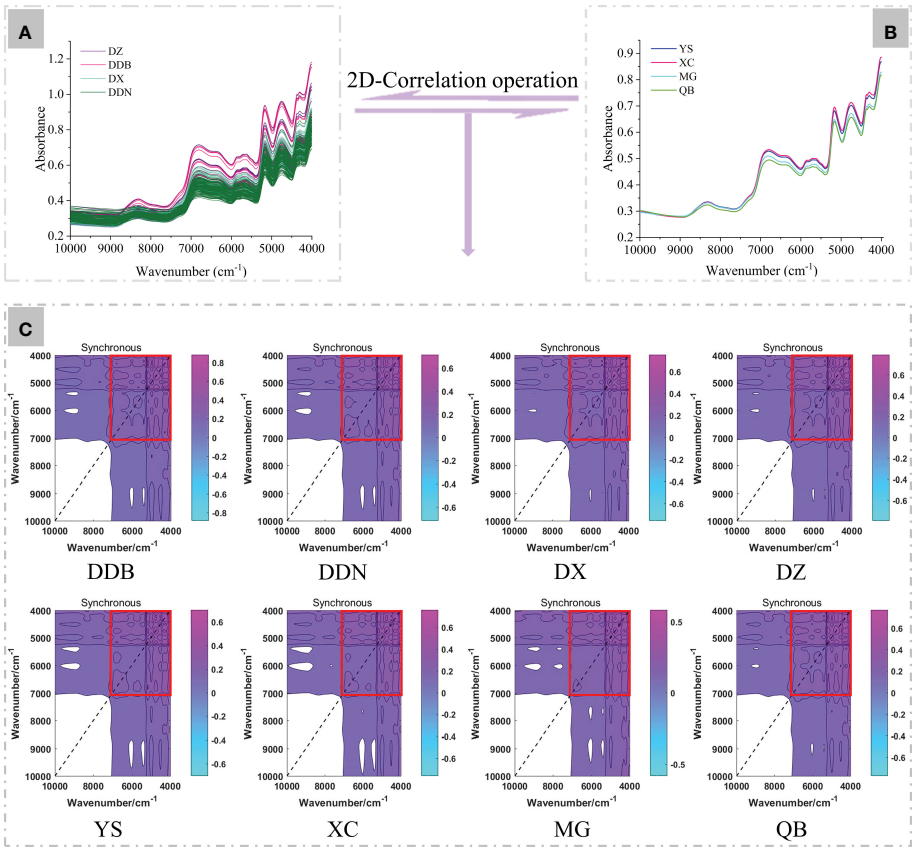


FIGURE 2
The generation of synchronous 2D-COS **(A, B)**. The original average NIR spectra of *P. notoginseng* from four different districts and towns. **(C)** Full-band synchronous 2D-COS images from different districts and towns. The red box shows the selected synchronous 2D-COS images in the range of 7000-4000 cm^{-1}.

subsequent analysis. In this study, 90% of the samples (60% as the training set, 30% as the test set) were chosen to establish the ResNet model and the remaining 10% for external validation. Then, a self-built script in MATLAB 2017a was run to generate synchronous 2D-COS images (in the form of JPEG images). This provided a foundation for deep learning modeling. Moreover, we set normalization and resizing in the script to keep the size of images consistent (128×128pixel). We used the MxNet deep learning framework and anaconda3-4.2.0 that comes with Python 3.5.2 to further our learning. Additionally, the TensorBoard and MxBoard were installed for training process visualization and networking.

In this study, the traceability model of *P. notoginseng* from different districts and towns was established by ResNet technology in Convolutional Neural Network (CNN) network. The ResNet technology of deep learning realized residuals with a "shortcut connections" structure, which could simplify learning objectives, reduce training difficulty, speed up the training, and improve the accuracy of the model. The residual module was applied to simplify learning objectives; the detailed process is presented in Figure S1. In addition, dimensional consistency of input and output data to determine the structure as identity block or convolution (conv) block was applied. The schematic diagrams of conv and the identity block are shown in Figure S2A, B.

The synchronous 2D-COS images acted as the input data. First, a layer of convolution operation is performed on the input data. Then, the BatchNorm normalization and Relu nonlinear activation processing were performed, and the data were input into a 32-layer convolutional neural network (11*2 identity blocks and 4*2 conv blocks) to extract features. The parameters of the fully connected layer were simplified. Additionally, the important features were extracted by global average pooling. Finally, the learned "distributed feature representation" was mapped to the sample label space using the full connection layer output data. The traceability flow chart of geographical origins is shown in Figure 3.

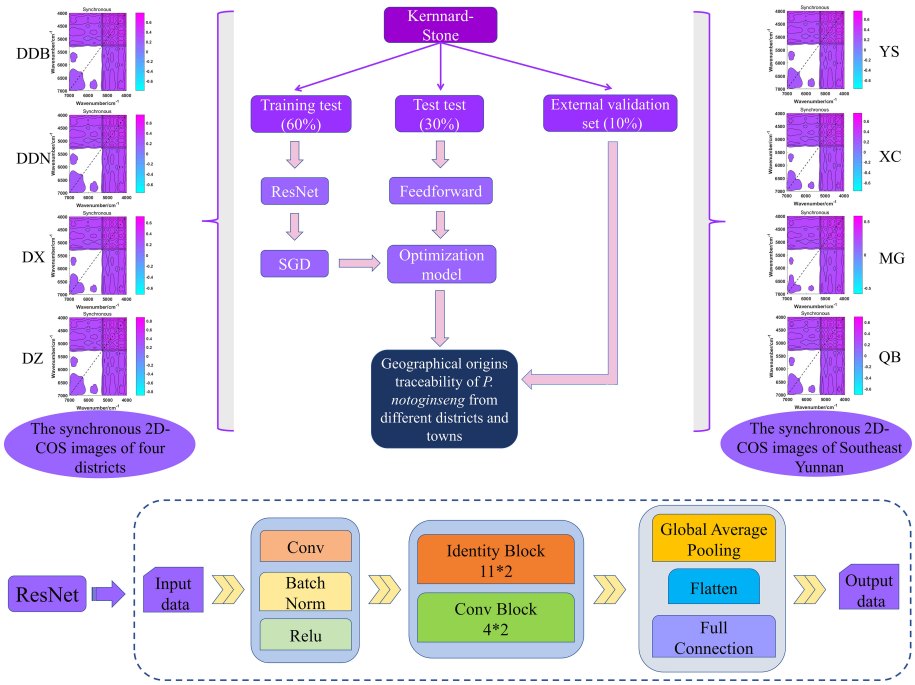


FIGURE 3
The geographical origins traceability flow chart of *P. notoginseng* from different districts and towns based on ResNet model, and the structure of ResNet model.

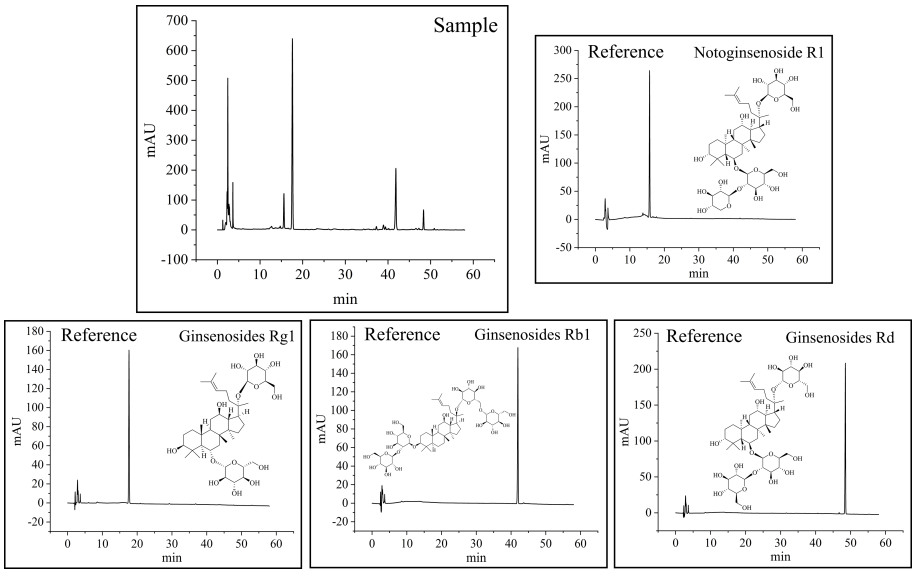


FIGURE 4
HPLC chromatograms at 203 nm, and the structure of four active compounds in *P. notoginseng*.

Results and discussion

Analysis of saponin contents in *P. notoginseng*

P. notoginseng is rich in saponins. Among all saponins, NG-R1, G-Rg1, G-Rb1, and G-Rd have the highest content in *P. notoginseng*. They are the most popular compounds applied for quality control of *P. notoginseng* in most studies owing to their excellent biological activity. The HPLC chromatograms and linear regression data of the four saponins are shown in Figure 4 and Table S3. The assay method of HPLC was validated. It is evident from Table S3 that concentrations and peak areas of the four components show an obvious linear relationship (R^2>0.9995). The relative standard deviation (RSD) value of stability repeatability, precision, and spiked recovery of each reference compound were all less than 3%. In view of this result, the established method fulfills the requirements for qualitative and quantitative analyses.

According to previous studies, saponins are one of the most important components to exert the drug efficacy of *P. notoginseng*. Saponins are typically used as a quality indicator for evaluating *P. notoginseng*. In this study, combined with *Chinese Pharmacopoeia* (*Chinese Pharmacopoeia* Committee 2020) and literature reports, four saponins in *P. notoginseng* were selected as indicators (Wei et al., 2018; Bai et al., 2021). The level of the four saponins in *P. notoginseng* from different districts and towns is shown in Table 1. The results showed that in *P. notoginseng* from different districts and towns, the highest content was of G-Rg1, followed by G-Rb1 and finally G-Rd and NG-R1. This is in line with previous studies by Wei et al. (Wei et al., 2018). In addition, the four saponins in DDB were lower than other districts, and the total average content was 16.55 mg/g. Relatively speaking, the total average content from DDN was the highest at 26.47 mg/g. Among the samples collected from different towns, the total average content from XC was the highest (32.78 mg/g), and the QB provided the lowest (22.04 mg/g). The total average content from YS and MG were similar (27.04 and 26.25 mg/g, respectively). These results indicated that there were certain differences between the content of saponin from different districts and towns. Therefore, comprehensively and properly evaluating the quality of *P. notoginseng* using the four saponins as quality control indicators is feasible. Simultaneously, the above results show the necessity and importance of the identification of *P. notoginseng*. However, there is uncertainty in analyzing the differences of *P. notoginseng* from different districts and towns only based on the content of four saponins. Therefore, the unsupervised multivariate method (PCA) was employed for further analysis.

PCA analysis

In order to further reflect the differences of four saponins (NG-R1, G-Rg1, G-Rb1 and G-Rd), PCA was applied to analyze *P.notoginseng* from different districts and towns, respectively. Be seen from Figures 5A, 5B, the first two components accounted for 97.94% of the total variance, which could explain most of the information in the sample. From the PCA score plots (Figure 5A), *P. notoginseng* from different districts is distributed in different quadrants. The *P. notoginseng* from DDB is located in the fourth quadrant and had a larger dispersion, indicating that saponins contribute substantially to the principal components. Furthermore, they are shown to be negatively correlated with both the first and second principal components. The dispersion of saponins in DZ, DX, and DDN samples is relatively low, which indicated that the component structures of the saponins are relatively similar and could be clustered into one category. However, the contribution rate for the principal components is not high. Figures 5C, D present the PCA score plots of *P. notoginseng* from four different towns. The samples of QB and XC have large dispersion, located in the second and third quadrants, respectively. In other words, saponins substantially contribute to the principal components. In addition, there is an overlapping trend between YS and MG, which could be clustered

TABLE 1 Each content and total contents of four main components in *P. notoginseng* from different districts and towns. ($\bar{x} \pm SD$) %.

	Notoginsenoside R1	Ginsenosides Rg1	Ginsenosides Rb1	Ginsenosides Rd	Total (NG-R1+G-Rg1+G-Rb1+G-Rd)
DDB	8.26 ± 1.52	31.28 ± 3.98	19.95 ± 2.47	6.71 ± 1.48	16.55 ± 3.63
DDN	11.39 ± 2.69	44.67 ± 4.04	34.41 ± 3.33	15.44 ± 2.85	26.47 ± 4.21
DX	10.88 ± 2.17	38.68 ± 3.36	32.80 ± 3.57	14.70 ± 2.67	24.26 ± 3.89
DZ	11.01 ± 2.79	36.67 ± 3.60	26.43 ± 3.17	12.01 ± 2.46	21.53 ± 3.75
YS	11.38 ± 2.68	48.07 ± 4.34	35.11 ± 3.22	13.59 ± 3.02	27.04 ± 4.42
XC	15.41 ± 2.98	53.25 ± 4.14	41.83 ± 3.44	20.64 ± 2.94	32.78 ± 4.43
MG	12.90 ± 2.42	46.53 ± 3.54	30.69 ± 3.38	14.87 ± 2.82	26.25 ± 4.09
QB	7.73 ± 2.23	34.52 ± 3.00	31.62 ± 2.86	14.28 ± 2.21	22.04 ± 3.65

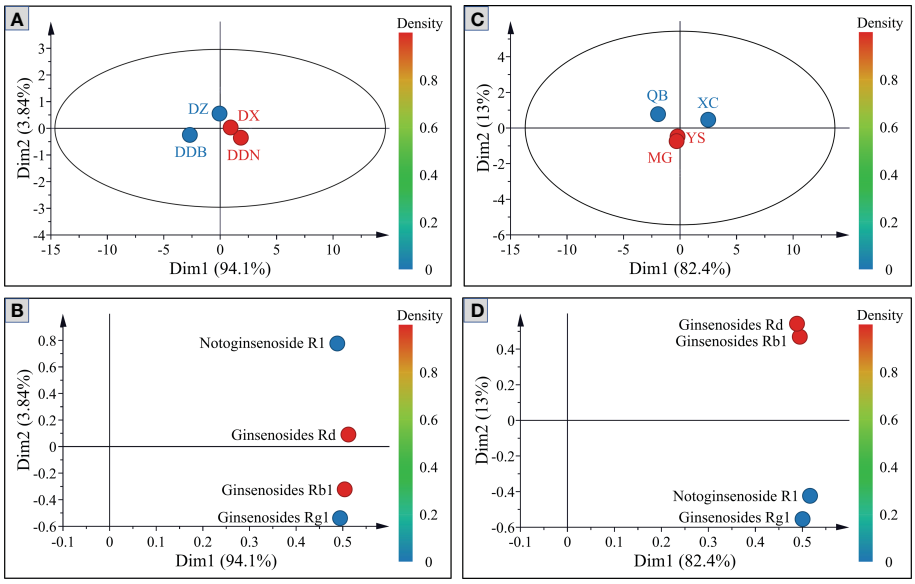


FIGURE 5
Principal component score plots and loading plots of saponins in *P. notoginseng* from different districts **(A, B)** and towns **(C, D)**.

into one category and have a low contribution rate for principal components. The PCA scores scatter plot was established after further analysis. The color of the point represents the contribution of different variables to the principal components. As shown in Figures 5B, D, the four saponins varied considerably between different geographic origins. Therefore, the correlation analysis was further carried out, with a view to observing the influence of climate factors on the content accumulation of saponins from different geographic origins.

Correlation analysis between saponin contents and climate factors

Climate factors have a critical effect on the distribution and secondary metabolites of plants. Generally speaking, linear correlation between the independent variables should be examined before constructing a regression model to prevent affecting the fitting effect of the regression model. Therefore, Pearson correlations were used to eliminate climate factors with high correlation coefficients (|R|>8) and less significance for the distribution of *P. notoginseng* samples. The results are shown in Figure 6. In the end, a total of seven climate factors (Bio 1, Bio 4, Bio 7, Bio 12, Bio 14, Bio 15, and Bio 17) were obtained for analysis. There was a significant correlation between the level of the four saponins and seven climate factors. Therefore, these seven climate factors were selected as independent variables.

The regression equations between the level of the four saponins and seven climate factors established by PLSR are shown below: NG-R1: Y=18.806-0.545 Bio1+0.04 Bio4-0.505 Bio7-0.001 Bio12-0.903 Bio14+0.133 Bio15+0.226 Bio17; G-Rg1: Y=197.720-2.552 Bio1-0.043 Bio4-4.425 Bio7+0.029 Bio12-0.927 Bio14-0.196 Bio15-0.104 Bio17; G-Rb1: Y=110.283-1.221 Bio1-0.046 Bio4-1.185 Bio7+0.037 Bio12-1.144 Bio14-0.406 Bio15-0.141 Bio17, and G-Rd: Y=38.450-0.477 Bio1+0.012 Bio4-0.728 Bio7+0.012 Bio12-1.055 Bio14-0.156 Bio15-0.102 Bio17. The results show that temperature

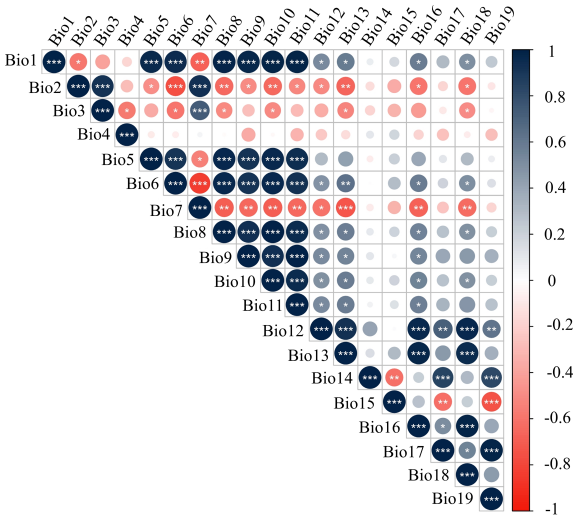


FIGURE 6
The autocorrelation test of climate factors. The definitions of climate factors are shown in Table S2.

and precipitation are crucial climate factors impacting the content of the four saponins in *P. notoginseng*. The content of NG-R1 correlated negatively with Bio1, Bio7, Bio12, and Bio14, and correlated with Bio4, Bio15, and Bio 17. For G-Rg1 and G-Rb1, their content displayed a negative correlation with Bio 1, Bio 4, Bio 7, Bio 14, Bio 15, and Bio 17, and demonstrated a positive correlation with Bio12. The correlation between the contents of G-Rd and climate factors is similar to G-Rg1 and G-Rb1, the only difference is the positively correlation with Bio4. In the analysis of PLSR, the explanatory power of the independent variable to the dependent variable is measured by the VIP. Therefore, the VIP values of the contents of the four saponins and seven climate factors were analyzed (Figure 7). The variables with larger contributions (VIP>1) were screened as important variables. From Figure 7, it is clear that Bio7 and Bio12 have a greater impact on the content accumulation of the four saponins. In addition, the content accumulation of the four saponins was also affected by Bio1, Bio1, Bio17, and Bio15, respectively.

To sum up, all the regression coefficients of Bio1 and Bio7 selected based on the VIP value were negative for the four saponins. It shows that these are negatively correlated with the annual mean temperature and the temperature annual range. That is to say, the lower annual mean temperature and the temperature annual range are favorable for the content accumulation of four saponins. DDN (Wenshan Prefecture) is located near the Tropic of Cancer and has a subtropical climate, where the temperature does not experience extremely high or low temperatures. Its annual mean temperature is 15.8°C-19.3° C, and the temperature annual range is small. This may be one of the reasons why Wenshan Prefecture could be regarded as the "Sanqi Hometown". In addition, the regression coefficients of G-Rg1, G-Rb1, G-Rd, and Bio12 were all positive values, which showed a positive correlation. That is to say, high annual precipitation and high humidity are suitable for the content accumulation of G-Rg1, G-Rb1, and G-Rd. Interestingly, the content accumulation of NG-R1 negatively correlated with Bio12. That is, low annual precipitation could be more suitable for the content accumulation of NG-R1. According to the actual climate analysis of DDN, we speculate that this may be the reason for the low contents of NG-R1 among four saponins. These results are consistent with the traditional production areas of *P. notoginseng*.

Analysis of original NIR spectra and 2D-COS images

The original average NIR spectra of *P. notoginseng* from four different districts and towns are shown in Figure 2A, B. It can be clearly discovered that the spectra at 10000-7600 cm^{-1} have got low signal-to-noise ratios and intensities, which this region probably unsuited for spectra differentiation (Gierlinger et al., 2004). At 7600-5200 cm^{-1} are the first overtone C-H that stretches vibrations in different groups. The peak of 5200-4000 cm^{-1} is the maximal value, which reflects the combined C-H absorption of amino acids, sugars, and proteins (Li et al., 2018; Liu et al., 2019). The broad bands at 8320 cm^{-1} correspond to the second overtone of the C-H stretching in different groups. The bands around 6356

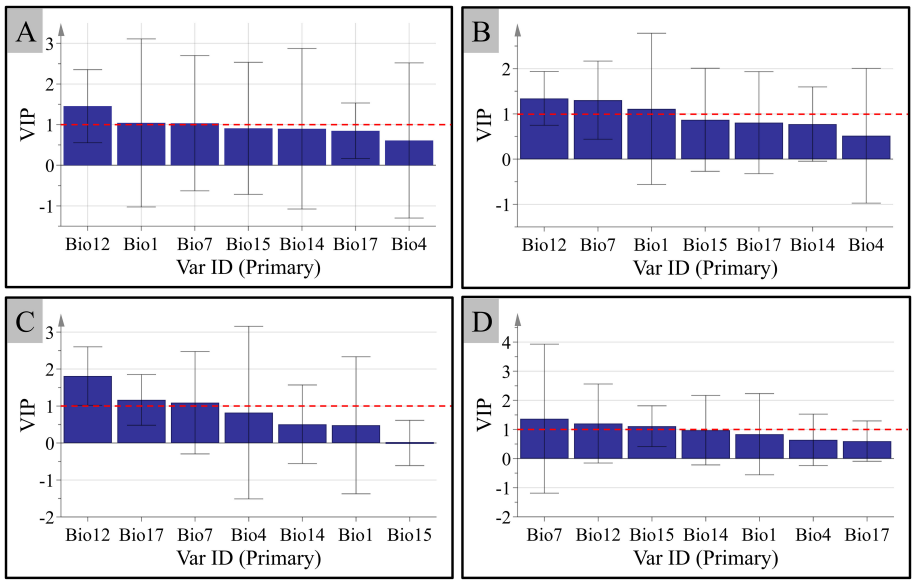


FIGURE 7
The variable importance in the projection (VIP) between the accumulation of the contents of the four main components in *P. notoginseng* and the climate factors. **(A)** Notoginsenoside R1. **(B)** Ginsenosides Rg1. **(C)** Ginsenosides Rb1. **(D)** Ginsenosides Rd.

and 6800 cm^{-1} are assigned to the first overtones of the O-H and the N-H stretching. The bands located around 5168 cm^{-1} correspond to the combination of O-H stretching and the first overtone of C-O deformation, and the 4756 cm^{-1} are from the combination of O-H deformation and the C-O stretching. In addition, the absorption band of 4300 cm^{-1} is assigned the combination overtone of C-H and C-C stretching (Nie et al., 2013; Fu et al., 2017; Li et al., 2018; Yang et al., 2019). However, from the original and average NIR spectra, there were less significant differences between the four districts and towns. This may be because the complex composition information of Chinese medicinal materials leads to their similarly existing chemical bonds. Another possible reason would be that the NIR spectra are C-H, O-H, and N-H stretched overtones and combined bands. They are characterized by absorption bandwidths, overlap, and weak absorption, which leads to the characteristics being similar (Nie et al., 2013). As a result, the geographic origins of *P. notoginseng* may be difficult to discriminate directly by the NIR spectra with the naked eye. Therefore, we converted the spectral data into corresponding 2D-COS images combined with the deep learning model for further analysis.

The synchronous 2D-COS images of *P. notoginseng* from different districts and towns are displayed in Figure 2C. It can be seen from the synchronous 2D-COS images that the feature peaks are mainly distributed in the 7000-4000 cm^{-1} bands. Therefore, the bands of 7000-4000 cm^{-1} were used for further deep learning modeling.

Geographical origins traceability analysis of *P. notoginseng* based on ResNet model

The samples of *P. notoginseng* from the four different districts and towns were collected in a relatively large number. The content of *P. notoginseng* from different geographic origins shows great differences due to the influence of climate and human factors. In Wenshan Prefecture, the "Sanqi Hometown, there were also slight differences in the content of samples from different towns. Therefore, tracing geographical origins was carried out of the district level and the town level.

In this study, the weight attenuation coefficient λ of the ResNet model was set to 0.0001, and the learning rate was set to 0.01. In addition, accuracy curves and cross-entropy cost function curves (smoothing parameter is 0.6) of the training set and test set were generated by Mxboard to evaluate the identification ability of the model. The value of accuracy curves is closer to 1, and the cross-entropy cost function is closer to 0, which indicated that the identification ability and convergence effect of the model is better.

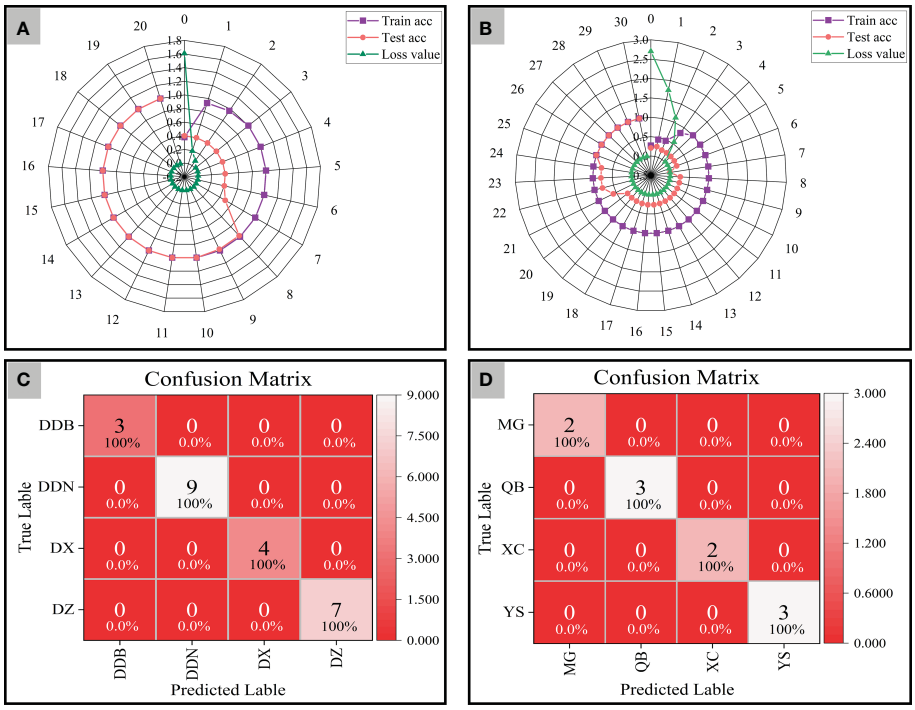


FIGURE 8
The radar plots of accuracy and cross-entropy cost function of models based on synchronous 2D-COS images (**A**: District; **B**: Town) and the confusion matrix of ResNet models (**C**: District; **D**: Town).

We performed ResNet model analysis on synchronized 2D-COS images of 152 training sets and 54 testing sets from four different districts. The radar plots show the classification accuracy and cross-entropy cost function of the model generated based on synchronized 2D-COS images. As can be seen from Figure 8A, the accuracy of both the training set and the test set shows a rising trend. When the number of epochs reaches 10, the accuracy of both the training set and the test set is 1, the loss value is reduced to 0.001, and the model training time is only 675 s. Furthermore, this study applied the established ResNet model for validation on 23 external validation sets, and all external validation samples from four different districts were correctly identified (Figure 8C). It shows that the model has strong robustness and could accurately distinguish *P. notoginseng* from different districts in a short time.

As verification, a total of 93 P. *notoginseng* samples were collected from four towns in Wenshan Prefecture and were also analyzed by the same method. As shown in Figure 8B, when the number of epochs is 25, the accuracy of both the training set and the test set reaches 1, and the cross-entropy loss value reaches the minimum (0.001). In addition, the results of the confusion matrix demonstrate that the external validation samples are classified correctly (Figure 8D). That is to say, the classification models of *P. notoginseng* at different town levels are as good as those at the district level. However, it was not difficult to see that the training time of the model reduced to 597 s.

From the above results, the established model could successfully trace the geographical origins of *P. notoginseng* from different district and town levels. However, the training time of the model may be affected due to the distance between the collected samples. That is to say, the model did not have the phenomenon of overfitting and had strong robustness. The geographical origin of *P. notoginseng* can also be accurately traced when the distance between sampling points is small. The only difference was in the training time of the model, which may be related to the sample number and the differences within the group.

Conclusion

Some studies from previous literature show that 2D-COS images combined with deep learning can authenticate different herbal and boletus samples, including origin, growth year, and species. In addition, there have been studies that analyzed different bands and different types of 2D-COS images (synchronous, asynchronous, and integrated 2D-COS images). The results of all these studies indicate that synchronous 2D-COS images combined with deep learning is the most suitable method for discrimination analysis. Comparatively, few studies have analyzed climatic factors and quality differences of *P. notoginseng* from different geographical origins .In this study, the method was used to identify the geographical origin. In this study, an identification model of geographical origins of *P. notoginseng* in different districts was proposed and verified by town level samples, the results have indirectly proven the reliability of the model.

Climate is one of the major factors that affects the growth suitability of Chinese medicinal materials, including *P. notoginseng*. Therefore, an investigation into the effects that climate has on the accumulation of active components is essential to improve the quality of *P. notoginseng*. In this study, four saponins of *P. notoginseng* from different districts and towns were determined using HPLC. The correlation between the level of saponins and climate factors was evaluated using PLSR and VIP, and the influence of climate factors on the quality of *P. notoginseng* was analyzed. The results showed that the presence of each saponin was negatively correlated with annual mean temperature and temperature annual range. A lower annual mean temperature and temperature annual range were favorable for the accumulation of the four saponins. In addition, high annual precipitation and high humidity are suitable for the content accumulation of NG-R1, G-Rg1, and G-Rb1, while this is not the case for G-Rd.

In addition, as a traditional Chinese medicinal material with high medicinal value and a high price, *P. notoginseng* is often fraudulently traded on the market. Therefore, a simple and reliable method was proposed to conduct a comprehensive geographic origin traceability study on *P. notoginseng* (from different districts), where the reliability of the model (from different towns) was verified. The results of the accuracy curve, cross-entropy loss function curve, and confusion matrix show that the synchronous 2D-COS model has a strong tendency for generalization. The method proposed in this study could achieve geographical origin traceability of *P. notoginseng*, even though the distance between sampling points is small. The findings of this study could lead to improvements in the quality of *P. notoginseng*, provide a reference for cultivating *P. notoginseng* in the future, and alleviate the phenomenon of market fraud.

Data availability statement

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

Author contributions

CL: Data curation and analysis, software, validation, and writing – review and editing. ZZ: validation, project administration, funding acquisition. FX: review – editing, supervision, project administration. YW: Supervision,

investigation, resources, project administration. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Natural Science Foundation of Yunnan Province of China (202101AT070260).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1009727/full#supplementary-material

References

Bai, J., Yue, P., Dong, Q., Wang, F., He, C. Y., Li, Y., et al. (2021). Identification of geographical origins of *Panax notoginseng* based on HPLC multi-wavelength fusion profiling combined with average linear quantitative fingerprint method. *Sci. Rep.-UK* 11, 5126. doi: 10.1038/s41598-021-84589-9

Candoğan, K., Altuntas, E. G., and İğci, N. (2021). Authentication and quality assessment of meat products by Fourier-transform infrared (FTIR) spectroscopy. *Food Eng. Rev.* 13, 66–91. doi: 10.1007/s12393-020-09251-y

Chen, H., Lin, Z., and Tan, C. (2018). Fast discrimination of the geographical origins of notoginseng by near-infrared spectroscopy and chemometrics. *J. Pharmaceut. Biomed.* 161, 239–245. doi: 10.1016/j.jpba.2018.08.052

Cui, Z. Y., Liu, C. L., Li, D. D., Wang, Y. Z., and Xu, F. R. (2022). Anticoagulant activity analysis and origin identification of *Panax notoginseng* using HPLC and ATR-FTIR spectroscopy. *Phytochem. Anal.* 33, 971–981. doi: 10.1002/pca.3152

Cura, K., Rintala, N., Kamppuri, T., Saarimäki, E., and Heikkilä, P. (2021). Textile recognition and sorting for recycling at an automated line using near infrared spectroscopy. *Recycling* 6, 11. doi: 10.3390/recycling6010011

Dong, J. E., Zuo, Z. T., Zhang, J., and Wang, Y. Z. (2021). Geographical discrimination of *Boletus edulis* using two dimensional correlation spectral or integrative two dimensional correlation spectral image with ResNet. *Food Control.* 129, 108132. doi: 10.1016/j.foodcont.2021.108132

Farrés, M., Platikanov, S., Tsakovski, S., and Tauler, R. (2015). Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *J. Chemometr.* 29 (10), 528–536. doi: 10.1002/cem.2736

Fu, H. Y., Yin, Q. B., Xu, L., Wang, W. Z., Chen, F., and Yang, T. M. (2017). A comprehensive quality evaluation method by FT-NIR spectroscopy and chemometric: Fine classification and untargeted authentication against multiple frauds for Chinese *Ganoderma lucidum*. *Spectrochim. Acta A.* 182, 17–25. doi: 10.1016/j.saa.2017.03.074

Gierlinger, N., Schwanninger, M., and Wimmer, R. (2004). Characteristics and classification of Fourier- transform near infrared spectra of the heartwood of different larch species (Larix sp.). *J. Near Infrared Spectrosc.* 12, 113–119. doi: 10.1255/jnirs.415

Hawthorne, B., Lund, K., Freggiaro, S., Kaga, R., and Meng, J. (2022). The mechanism of the cytotoxic effect of *Panax notoginseng* extracts on prostate cancer cells. *Biomed. Pharmacother.* 149, 112887. doi: 10.1016/j.biopha.2022.112887

He, Z. J., Liang, S. W., Ding, Y., Liu, Y., and Chen, Z. J. (2016). Relationships between $\delta^{18}O$ in taproot of *panax notoginseng* and ecological factors. *J. Nucl. Agric. Sci.* 03, 556–564.

Huber, M., Kepesidis, K. V., Voronina, L., Bozic, M., Trubetskov, M., Harbeck, N., et al. (2021). Stability of person-specific blood-based infrared molecular fingerprints opens up prospects for health monitoring. *Nat. Commun.* 12, 1511. doi: 10.1038/s41467-021-21668-5

Ichim, M. C., and de Boer, H. J. (2021). A review of authenticity and authentication of commercial ginseng herbal medicines and food supplements. *Front. Pharmacol.* 11. doi: 10.3389/fphar.2020.612071

Jiang, Y. X., Li, H. M., Huang, P. L., Li, S. L., Li, B. C., Huo, L. N., et al. (2022). *Panax notoginseng* saponins protect PC12 cells against aβ induced injury *via* promoting parkin-mediated mitophagy. *J. Ethnopharmacol.* 285, 114859. doi: 10.1016/j.jep.2021.114859

Ji, C., Liu, J. Y., Zhang, Q., Li, J., Wu, Z. Q., Wang, X. Y., et al. (2022). Multi-element analysis and origin discrimination of *Panax notoginseng* based on inductively coupled plasma tandem mass spectrometry (ICP-MS/MS). *Molecules* 27, 2982. doi: 10.3390/molecules27092982

Jogin, M., Mohana,, Madhulika, M. S., Divya, G. D., Meghana, R. K., and Apoorva, S. (2018). "Feature extraction using convolution neural networks (CNN) and deep learning," in *2018 3rd IEEE international conference on recent trends in electronics*. 2319–2323 (Bangalore, India: Information & communication technology (RTEICT)).

Kirschbaum, C., Greis, K., Mucha, E., Kain, L., Deng, S., Zappe, A., et al. (2021). Unravelling the structural complexity of glycolipids with cryogenic infrared spectroscopy. *Nat. Commun.* 12, 1201. doi: 10.1038/s41467-021-21480-1

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

Liang, S. W., He, Z. J., Xiong, J. F., and Hong, C. Q. (2021). Origin traceability of main root of spring *Panax notoginseng* based on stable isotope fingerprint. *China J. Chin. Mater. Med.* 46, 560–566. doi: 10.19540/j.cnki.cjcmm.20201121.103

Liu, P., Wang, J., Li, Q., Gao, J., Tan, X. Y., and Bian, X. H. (2019). Rapid identification and quantification of *Panax notoginseng* with its adulterants by near infrared spectroscopy combined with chemometrics. *Spectrochim. Acta A.* 206, 23–30. doi: 10.1016/j.saa.2018.07.094

Liu, C. L., Zuo, Z. T., Xu, F. R., and Wang, Y. Z. (2022). Authentication of herbal medicines based on modern analytical technology combined with chemometrics approach: A review. *Crit. Rev. Anal. Chem.* in press doi: 10.1080/10408347.2021.2023460

Li, Y., Zhang, J. Y., and Wang, Y. Z. (2018). FT-MIR and NIR spectral data fusion: A synergetic strategy for the geographical traceability of *Panax notoginseng*. *Anal. Bioanal. Chem.* 410, 91–103. doi: 10.1007/s00216-017-0692-0

Li, Y., Zhang, J., Xu, F. R., Wang, Y. Z., and Zhang, J. Y. (2017). Rapid prediction study of total flavonoids content in *panax notoginseng* using infrared spectroscopy combined with chemometrics. *Spectrosc. Spect. Anal.* 37, 70–74.

Margaritis, A., Soenen, H., Fransen, E., Pipintakos, G., Jacobs, G., Blom, G., et al. (2020). Identification of ageing state clusters of reclaimed asphalt binders using principal component analysis (PCA) and hierarchical cluster analysis (HCA) based on chemo-rheological parameters. *Constr. Build. Mater.* 244, 118276. doi: 10.1016/j.conbuildmat.2020.118276

Marianela, G. S., Chavarría-Rojas, M., Zúñiga, R. V., Berrocal, G. C., and Redondo, G. M. (2021). Ginsenosides and other phytochemicals of *Panax* spp. properties and uses in the pharmaceutical field. *Pharmacogn. Commn.* 11, 177–181. doi: 10.5530/pc.2021.4.36

Mishra, P., Woltering, E., Brouwer, B., and Hogeveen-Van, E. (2021). Improving moisture and soluble solids content prediction in pear fruit using near-infrared

spectroscopy with variable selection and model updating approach. *Postharvest Biol. Tec.* 171, 111348. doi: 10.1016/j.postharvbio.2020.111348

National Pharmacopoeia Committee. (2020). *Chinese pharmacopoeia*. (Beijing: China Medical Science and Technology Press).

Nie, P. C., Wu, D., Sun, D. W., Cao, F., Bao, Y. D., and He, Y. (2013). Potential of visible and near infrared spectroscopy and pattern recognition for rapid quantification of notoginseng powder with adulterants. *Sens. (Basel Switzerland).* 13, 13820–13834. doi: 10.3390/s131013820

Noda, I. (1989). Two-dimensional infrared spectroscopy. *J. Am. Chem. Soc* 111, 8116–8118. doi: 10.1021/ja00203a008

Noda, I. (1990). Two-dimensional infrared (2D IR) spectroscopy: Theory and applications. *Appl. Spectrosc.* 44, 550–561. doi: 10.1366/0003702904087398

Peng, Y., Wu, Z. J., Huo, Y. P., Chen, Y. W., Lu, F. H., Peng, Q., et al. (2017). Simultaneous determination of ginsenosides Rg1, re, and Rb1 and notoginsenoside R1 by solid phase extraction followed by UHPLC-MS/MS and investigation of their concentrations in various kinds of cosmetics. anal. *Methods-U. K.* 9, 5441–5448. doi: 10.1039/C7AY01651D

Qiao, Y. J., Shang, J. H., Wang, D., Zhu, H. T., Yang, C. R., and Zhang, Y. J. (2018). Research of *Panax* spp. in kunming institute of botany. *CAS. Nat. Prod. Bioprospect.* 8, 245–263. doi: 10.1007/s13659-018-0176-8

Teseo, S., Houot, B., Yang, K., Monnier, V., Liu, G., and Tricoire, H. (2021). *G. sinense* and *P. notoginseng* extracts improve healthspan of aging flies and provide protection in a huntington disease model. *Aging Dis.* 12, 425. doi: 10.14336/AD.2020.0714-1

Tian, L. X., Li, J. H., Zhang, L., Ahmad, B., and Huang, L. F. (2021). Discrimination of five species of panax genus and their geographical origin using electronic tongue combined with chemometrics. *World J. Tradit. Chin. Med.* 7, 104. doi: 10.4103/wjtcm.wjtcm_80_20

Tran, T. N., Afanador, N. L., Buydens, L. M., and Blanchet, L. (2014). Interpretation of variable importance in partial least squares with significance multivariate correlation (sMC). *Chemometr. Intell. Lab.* 138, 153–160. doi: 10.1016/j.chemolab.2014.08.005

Wang, D., Liao, P. Y., Zhu, H. T., Chen, K. K., Xu, M., Zhang, Y. J., et al. (2012). The processing of *Panax notoginseng* and the transformation of its saponin components. *Food Chem.* 132, 1808–1813. doi: 10.1016/j.foodchem.2011.12.010

Wan, G. Z., Wang, L., Jin, L., and Chen, J. (2021). Evaluation of environmental factors affecting the quality of *Codonopsis pilosula* based on chromatographic fingerprint and MaxEnt model. *Ind. Crop Prod.* 170, 113783. doi: 10.1016/j.indcrop.2021.113783

Wei, G. F., Dong, L. L., Yang, J., Zhang, L. J., Xu, J., Yang, F., et al. (2018). Integrated metabolomic and transcriptomic analyses revealed the distribution of saponins in *Panax notoginseng*. *Acta Pharm. Sin. B* 8, 458–465. doi: 10.1016/j.apsb.2017.12.010

Wildea, A. S., Haughey, S. A., Galvin-King, P., and Elliott, C. T. (2019). The feasibility of applying NIR and FT-IR fingerprinting to detect adulteration in black pepper. *Food Control* 100, 1–7. doi: 10.1016/j.foodcont.2018.12.039

Yang, R. J., Dong, G. M., Sun, X. S., Yu, Y. P., Liu, H. X., Yang, Y. R., et al. (2015). Synchronous-asynchronous two-dimensional correlation spectroscopy for the discrimination of adulterated milk. *Anal. Methods-UK* 7, 432–437. doi: 10.1039/c5ay00134j

Yang, R. J., Liu, R., and Xu, K. X. (2013). Detection of adulterated milk using two-dimensional correlation spectroscopy combined with multi-way partial least squares. *Food Biosci.* 2, 61–67. doi: 10.1016/j.fbio.2013.04.005

Yang, R., Liu, C. Y., Yang, Y. R., Wu, H. Y., Jin, H., Shan, H. Y., et al. (2020). Two-trace two-dimensional(2T2D) correlation spectroscopy application in food safety: A review. *J. Mol. Struct.* 1214, 128219. doi: 10.1016/j.molstruc.2020.128219

Yang, X. D., Song, J., Peng, L., Gao, L. T., Liu, X. W., Xie, L., et al. (2019). Improving identification ability of adulterants in powdered *Panax notoginseng* using particle swarm optimization and data fusion. *Infrared Phys. Techn.* 103, 103101. doi: 10.1016/j.infrared.2019.103101

Yang, R. J., Yang, Y. R., Dong, G. M., Zhang, W. Y., and Yu, Y. P. (2014). Multivariate methods for the identification of adulterated milk based on two-dimensional infrared correlation spectroscopy. *Anal. Methods-UK* 6, 3436–3441. doi: 10.1039/c4ay00442f

Yao, C. L., Wang, J., Li, Z. W., Qu, H., Pan, H. Q., Li, J. Y., et al. (2021). Characteristic malonyl ginsenosides from the leaves of *Panax notoginseng* as potential quality markers for adulteration detection. *J. Agr. Food Chem.* 69, 4849–4857. doi: 10.1021/acs.jafc.1c00382

Yue, J. Q., Li, Z. M., Zuo, Z. T., and Wang, Y. Z. (2022). Evaluation of ecological suitability and quality suitability of *panax notoginseng* under multi-regionalization modeling theory. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.818376

Yu, N., Xing, R. R., Wang, P., Deng, T. T., Zhang, J. K., Zhao, G. M., et al. (2022). A novel duplex droplet digital PCR assay for simultaneous authentication and quantification of *Panax notoginseng* and its adulterants. *Food Control* 132, 108493. doi: 10.1016/j.foodcont.2021.108493

Zhang, J. Y., Xu, X. Z., Kuang, S. B., Cun, Z., Wu, H. M., Shuang, S. P., et al. (2021). Constitutive activation of genes involved in triterpene saponins enhances the accumulation of saponins in three-year-old *Panax notoginseng* growing under moderate light intensity. *Ind. Crop Prod.* 171, 113938. doi: 10.1016/j.indcrop.2021.113938

Zheng, Y. R., Fan, C. L., Chen, Y., Quan, J. Y., Shi, L. Z., Tian, C. Y., et al. (2022). Anti-inflammatory, anti-angiogenetic and antiviral activities of dammarane-type triterpenoid saponins from the roots of *Panax notoginseng*. *Food Funct.* 13, 3590–3602. doi: 10.1039/d1fo04089h

Zhou, Y. H., Zuo, Z. T., Xu, F. R., and Wang, Y. Z. (2020). Origin identification of *Panax notoginseng* by multi-sensor information fusion strategy of infrared spectra combined with random forest. *Spectrochim. Acta A* 226, 117619. doi: 10.1016/j.saa.2019.117619

frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 09 February 2023
DOI 10.3389/fpls.2023.1003150

OPEN ACCESS

EDITED BY
Yijun Yan,
Robert Gordon University, United Kingdom

REVIEWED BY
Zichen Huang,
Kyoto University, Japan
Jinling Zhao,
Anhui University, China
Tanvir Kaur,
Eternal University, India

*CORRESPONDENCE
Sruti Das Choudhury
S.D.Choudhury@unl.edu

†These authors have contributed equally to this work

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

RECEIVED 25 July 2022
ACCEPTED 09 January 2023
PUBLISHED 09 February 2023

CITATION
Das Choudhury S, Saha S, Samal A, Mazis A and Awada T (2023) Drought stress prediction and propagation using time series modeling on multimodal plant image sequences.
Front. Plant Sci. 14:1003150.
doi: 10.3389/fpls.2023.1003150

Drought stress prediction and propagation using time series modeling on multimodal plant image sequences

Sruti Das Choudhury[1,2*†], Sinjoy Saha[3†], Ashok Samal[2], Anastasios Mazis[1,4] and Tala Awada[1,5]

[1]School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, United States, [2]School of Computing, University of Nebraska-Lincoln, Lincoln, NE, United States, [3]Institute of Radio Physics and Electronics, University of Calcutta, Kolkata, West Bengal, India, [4]Department of Civil and Environmental Engineering, University of California, Merced, Merced, CA, United States, [5]Agricultural Research Division, University of Nebraska-Lincoln, Lincoln, NE, United States

The paper introduces two novel algorithms for predicting and propagating drought stress in plants using image sequences captured by cameras in two modalities, i.e., visible light and hyperspectral. The first algorithm, VisStressPredict, computes a time series of holistic phenotypes, e.g., height, biomass, and size, by analyzing image sequences captured by a visible light camera at discrete time intervals and then adapts dynamic time warping (DTW), a technique for measuring similarity between temporal sequences for dynamic phenotypic analysis, to predict the onset of drought stress. The second algorithm, HyperStressPropagateNet, leverages a deep neural network for temporal stress propagation using hyperspectral imagery. It uses a convolutional neural network to classify the reflectance spectra at individual pixels as either stressed or unstressed to determine the temporal propagation of stress in the plant. A very high correlation between the soil water content, and the percentage of the plant under stress as computed by HyperStressPropagateNet on a given day demonstrates its efficacy. Although VisStressPredict and HyperStressPropagateNet fundamentally differ in their goals and hence in the input image sequences and underlying approaches, the onset of stress as predicted by stress factor curves computed by VisStressPredict correlates extremely well with the day of appearance of stress pixels in the plants as computed by HyperStressPropagateNet. The two algorithms are evaluated on a dataset of image sequences of cotton plants captured in a high throughput plant phenotyping platform. The algorithms may be generalized to any plant species to study the effect of abiotic stresses on sustainable agriculture practices.

KEYWORDS

stress prediction, image sequence analysis, time series modeling, dynamic time warping, temporal stress propagation, spectral band difference segmentation, deep neural networks

1 Introduction

Increasing demands for food, fuel, fiber, and feed to meet the needs of the growing population, under climate change, and dwindling natural resources constitute a major challenge confronting sustainable agriculture in the 21st century. In addition, climate change has affected the intensity and frequency of drought and extreme weather events in many regions, increasing food insecurity and affecting the livelihoods of many communities (Sheffield et al., 2012; Ort et al., 2015). In fact, it is estimated that about two thirds of crop losses in the last half century were caused by drought (Comas et al., 2013). Thus, an improved understanding of the plant's response to increased water stress as a function of time is an important step in shepherding breeding efforts, developing smart agricultural practices, and enhancing the decision making process to mitigate and adapt to climate change.

A time series is an ordered sequence of values of a variable measured at successive points in time, often at regular time intervals, e.g., weather forecasts, stock prices, biometrics, and exchange rates in finance. Based on the variable, a time series can be classified as either continuous or discrete. In the case of a continuous time series, observations are measured continuously over time, e.g., temperature readings, and the flow of a river. On the other hand, a discrete time series is characterized by recordings at typically equally spaced time intervals, e.g., daily, weekly, or yearly. High throughput plant phenotyping (HTPP) refers to the imaging of plants captured at regular intervals for a significant time period to extract the salient information about a plant's development and metabolism that are manifested at different wavelengths of the electromagnetic spectrum. Visible light image sequences are used to extract morphological characteristics of the plants or their organs (Dyrmann, 2015; Das Choudhury et al., 2018). In contrast, infrared images can serve as a proxy for a plant's temperature, which in turn can be used to detect differences in stomatal conductance, a measure of a plant's response to water status and transpiration rate for abiotic stress adaptation (Li et al., 2014). Hyperspectral cameras typically capture a scene in hundreds of bands covering a broad range of wavelengths at very narrow intervals. Since hyperspectral imaging has the highest coverage of the electromagnetic spectrum, it has the potential for a wide variety of applications, including the detection of abiotic and biotic stresses in plants and the measurements of chlorophyll content, canopy senescence, and water content (Gampa and Quinones, 2020). In this paper, we used time-series image sequences captured by two types of cameras, i.e., visible light and hyperspectral, for stress prediction and temporal stress propagation.

The images in an HTPP platform are captured at regular intervals with timestamps to compute phenotypes, i.e., the observable traits of plants as a result of the complex interaction between genotype and environment. Imaging at regular intervals facilitates the extraction of smart phenotypic traits, e.g., predicting the onset of stress and its temporal propagation patterns in a plant. Since the process of phenotypic trait extraction based on image analysis is nondestructive in nature, the traits may be extracted at multiple timestamps in a plant's life cycle. The phenotypes computed by analyzing the images captured in an HTPP may be modeled as a discrete time series. These abstractions and subsequent computations are not possible from manual measurements.

The phenotypic time series can be classified into the following four categories (Das Choudhury, 2020b):

a. *Nonlinear*: A phenotypic time series that tends to increase, decrease or stagnate over time is referred to as a nonlinear time series. The total leaf area of a plant increases over time under normal growth conditions, however, it often starts to decrease as some leaves experience curling or shedding due to exposure to stress, e.g., drought, thermal, and salinity. Note that for many cereal crops, e.g., maize and sorghum, the plant height increases monotonically with time and then remains stagnant upon completion of the vegetative stage.
b. *Recovery*: The normal growth of a plant is significantly affected under stress. However, if the stress condition is reverted, i.e., re-watering of a drought-stressed plant or adjusting the temperature of a plant under thermal stress, normal growth may resume under certain circumstances.
c. *Seasonal*: Plants undergo internal physiological seasonal changes leading to changing leaf colors, shedding, blooming, and generating new leaves. A time series representing the total number of leaves over a growing season can demonstrate this effect.
d. *Catastrophic*: A catastrophic phenotypic time series reflects any significant impact on a plant's development due to unprecedented events, e.g., floods, storms, and earthquakes, and hence does not follow any specific pattern.

This paper presents two algorithms to understand the dynamics of stress in plants from image sequences. It first describes a predictive model to determine if a plant is under stress, using the time series of holistic phenotypes or traits computed by analyzing visible light image sequences using dynamic time warping (DTW) - a statistical method extensively used to analyze temporal sequences, including applications in speech recognition and biometric verification (Das Choudhury and Tjahjadi, 2013). The paper introduces a novel dynamically growing subsequence based DTW matching algorithm for stress prediction using the phenotypic time series.

Deep neural networks have been successfully employed in high throughput temporal plant phenotyping for a variety of applications (Bashyam et al., 2021; Zheng et al., 2021; Das Choudhury et al., 2022). The method in (Das Choudhury et al., 2022) performs automated flower detection from multi-view image sequences to determine a set of novel phenotypes, e.g., the emergence time of the first flower, the total number of flowers present in the plant at a given time, flower growth trajectory, and blooming trajectory. A graph theoretic approach has been used by (Bashyam et al., 2021) to detect and track individual leaves of a maize plant for automated growth stage monitoring. The method by (Azimi et al., 2021) uses Convolutional Neural Network - Long Short Term Memory (CNN-LSTM) for water stress classification in chickpea plants, whereas the method by (Taha et al., 2022) uses deep convolutional neural networks (DCNNs) to diagnose the nutrient status of lettuce grown in aquaponics. In this paper, we present a novel algorithm based on convolutional neural networks to determine the qualitative and quantitative propagation of

drought stress in cotton plants by classifying reflectance spectra generated from hyperspectral image sequences.

The efficacy of the two algorithms is demonstrated using a set of cotton (*Gossypium* spp.) plants. Cotton, a C3 crop known for its valued fiber (cotton lint), supplies about 79% of the global natural fiber used in the textile industry (Dabbert and Gore, 2014; Townsend and Sette, 2016), while its seeds provide nutrition to both humans and animals (Bertrand et al., 2005). Drought stress has been identified as a major impediment to cotton production. In cotton, drought stress causes a reduction in both quantity and quality of lint (Pettigrew, 2004; Wang et al., 2016) with a severe negative impact on a farmer's income and supply of raw material for the textile industry. Although the algorithms introduced in this paper are evaluated on a cotton dataset, they are generic and, thus, are applicable to any plant species subjected to any kind of stress, i.e., drought, salinity, and thermal, to quantitatively determine the impact of stress as a function of time.

2 Materials and methods

In this section, we first describe the dataset used to develop and evaluate the two algorithms, i.e., VisStressPredict and HyperStressPropagateNet, followed by the detailed descriptions of these algorithms.

2.1 Dataset

The image sequences used for algorithm development and evaluation were obtained at the Innovation Campus greenhouse of the University of Nebraska-Lincoln (Lincoln, Nebraska, U.S.) using High Throughput Plant Phenotyping Core Facilities (HTPP, Scanalyzer 3D, LemnaTec Gmbh, Aachen, Germany). Chemically-delinted black cotton seeds (variety PHY 499 WRF) were planted in 5.7 L pots (22 cm diameter and 20 cm height) filled with 25% sand and 75% standard greenhouse mix, at approximately 24°C, and RH 58%. The daytime Photosynthetic Active Radiation (PAR) was supplemented with LED red/blue light, with an intensity of 200 $\mu mol\ m^{-2}s^{-1}$. The photoperiod in the greenhouse was set at 17 hours throughout the study to standardize the light conditions. After germination, plants were maintained on the bench where nutrients and water were applied following a standard greenhouse management regime. After two weeks, plants were randomly divided into two groups of 10 corresponding to the two experimental groups (i.e., Experiments 1 and 2). Each experimental group was further split into two groups of 5 plants and assigned to treatment groups (control and drought stress). The onset of the dry-down and the duration of the experiment varied across the two experiments. In Experiment 1, dry-down was initiated 12 days after the onset of plant imaging and lasted for 8 days. A week later, a similar dry-down was initiated for the second experimental group and lasted for 9 days.

Each plant was placed in a metallic carrier (dimension: 236 mm × 236 mm × 142 mm) on an automated conveyor belt that moves the plants from the greenhouse to the four imaging chambers successively for capturing images in different modalities. It has three watering stations with a balance that can add water to a target weight or specific volume and records the specific quantity of water added daily. The images of the greenhouse with plants placed on the automated conveyor belt, the watering station, and plants entering into the imaging cabinets are available in (Das Choudhury et al., 2018; Das Choudhury et al., 2022). The cameras installed in the four imaging chambers are (a) visible light - side view and top view (Prosilica GT6600 29 megapixel camera with a Gigabit Ethernet interface [1]), (b) infrared - side view and top view (Pearleye p-030 LWIR), (c) fluorescent - side view and top view (Basler Scout scA1400-17gm/gc), and (d) hyperspectral - side view (Headwall Hyperspec Inspector x-vnir [2]) and near-infrared - top view (Goldeye p-008 SWIR), respectively. Each imaging chamber has a rotating lifter for up to 360 side view images. In this study, we used visible light images (captured from five side-views, i.e., 0°, 72°, 144°, 216°, 288°) for VisStressPredict algorithm and hyperspectral images for HyperStressPropagateNet algorithm. The average time interval between a plant entering into and exiting from each of the first three imaging chambers for capturing five side view images is approximately 1 minute and 10 seconds. Since a hyperspectral camera typically captures a scene in hundreds of bands at a narrow interval over a broad range of the spectrum, its image capturing time is significantly higher than that of the other imaging modalities. In our HTPP facility, the time to capture a single side view image of a plant using a hyperspectral camera (total number of bands: 243; spectrum range: 546 nm to 1700 nm) is approximately 2 minutes and 15 seconds. All images are exported as PNG file types. Pots were automatically weighed upon exiting the hyperspectral chamber, and water was applied daily to designated levels to reach a predetermined percentage of field capacity (50%). Table 1 provides detailed information on the specifications of the cameras of our HTPP system.

2.2 VisStressPredict: DTW based stress prediction using visible light imagery

2.2.1 Image-based phenotypic time series computation

In this section, we describe the steps to compute phenotypic time series based on analyzing image sequences. Visible light images are used to compute structural phenotypes that characterize a plant's morphology. Image-based structural phenotypes can either be computed by considering the whole plant as a single object (holistic phenotypes) or by considering individual components of the plants, e.g., stem, leaves, fruits, and flowers (component phenotypes). Figure 1 shows the intermediate images in the computation of three holistic structural phenotypes, i.e., the height of the plant, the area of the convex hull enclosing the plant, and the total number of plant pixels, all of which contribute to the measurement of plant growth and development. First, the original plant image sequences are cropped to a fixed size to remove the frames of the imaging cabinet and the pot. Figure 1A shows a sample original image, and Figure 1B shows the corresponding cropped image that retains the plant. The

1 https://www.alliedvision.com/en/support/technical-documentation/prosilica-gt-documentation

2 https://www.manualslib.com/products/Headwall-Hyperspec-Inspector-10421605.html

TABLE 1 Camera specifications of the HTPP system at the UNL, USA.

Image type	Camera sensor	Spectral range (nm)	Spatial resolution	Bit depth	Frame rate
Visible light	AVT Prosilica GT6600	400-700	6576 × 4384	14 (mono) - 12 (color)	4
Fluorescent	Basler Scout scA1400-17GC	620-900	1390 × 1038	12	17
Near infrared	Goldeye P-008 SWIR	900-1700	320 × 256	12	118
Infrared	Pearleye P-030 LWIR	800-1400	640 × 480	14	24
Hyperspectral	Headwall Hyperspec Inspector X-VNIR	546-1700	320 × 561	8	–

cropped image is then binarized in the LAB color space using color thresholding (Figure 1C). Finally, the binary image thus obtained is enclosed by primitive geometric shapes, e.g., bounding rectangle and convex hull (Figure 1D), to compute holistic phenotypes.

Figures 2A, B show the image sequences of cotton plants enclosed by bounding rectangle and convex hull under normal condition and drought stress, respectively. Figures 3A, B show the nonlinear phenotypic time series of the plant height for a set of controlled and drought-stressed plants, respectively. Similarly, Figures 3C, D show the nonlinear phenotypic time series for plant biomass (measured by the total number of plant pixels as the function of time) for a set of controlled and drought-stressed plants, respectively. Figures 3E, F show the nonlinear phenotypic time series for plant size (measured by the area of the convex hull enclosing the plant) for a set of controlled and drought-stressed plants, respectively.

To validate the phenotypic traits measured noninvasively based on analyzing images captured in the HTPP system against the destructive handheld (low-throughput) techniques, we correlated the projected leaf area (pixels) and plant height (pixels), derived from the RGB camera of the HTPP, against values derived from low-throughput destructive methods (Figure 4). Image-derived projected plant biomass and plant height were highly and significantly correlated with the measured leaf area (R^2 = 0.92, p< 0.01) and plant height (R^2 = 0.94, p< 0.01) respectively, confirming the hypothesis for the HTPP methods' ability to accurately estimate morphological traits.

2.2.2 Time series smoothening

The noise introduced during the binarization process and the natural change of orientation of plants' leaves results in unevenness in the phenotypic time series, which poses significant challenges to subspace matching based on dynamic time warping. We use a moving average (MA) filter to smooth the time series to address this. The MA filter is the most common filter in digital signal processing to smooth functions. It is effective for time-domain encoded signals due to its simplicity. The high frequencies to be removed can be controlled by the length of the window of the MA filter.

Given a phenotypic time series x of length m and a window size N, the filtered time series y is given by Equation 1 as follows:

$$y[i] = \frac{1}{N}\sum_{j=0}^{N-1} x[i+j-1], \quad 1 \le i \le m-N+1 \tag{1}$$

In this paper, a window size of N=3 is used for the MA filter. The smoothened time series are used as input to the DTW-based drought stress prediction algorithm. Figure 5 shows the smoothened time series of the phenotypes (plant height, plant biomass and plant size) for control and stressed plants of Experiment 2.

2.2.3 Stress prediction using DTW

The goal of time series modeling is to study past observations to develop an appropriate model that describes its underlying structure for making predictions. Dynamic time warping (DTW) (Sakoe and Chiba, 1978) is widely used to find the optimal alignment between two given time series. It has been successfully used in automatic speech recognition, gait recognition, and data mining to compare time series with different speeds and deformations. DTW uses dynamic programming to compute a warping function that optimally aligns two time series of variable lengths and measures their similarity. Given two plant phenotypic time series, i.e., $P=(P_1,P_2,\ldots,P_M)$ and $Q=(Q_1,Q_2,\ldots,Q_N)$ of respective lengths $M\in N$ and $N\in N$, and P_i and Q_j are the respective phenotypic value on the i^{th} and j^{th} days, DTW constructs an $M\times N$ warping path which is a sequence of length p of L index pairs $((i_1,j_1),(i_2,j_2),\ldots,(i_L,j_L))$ and $A(P,Q)$ is a set of

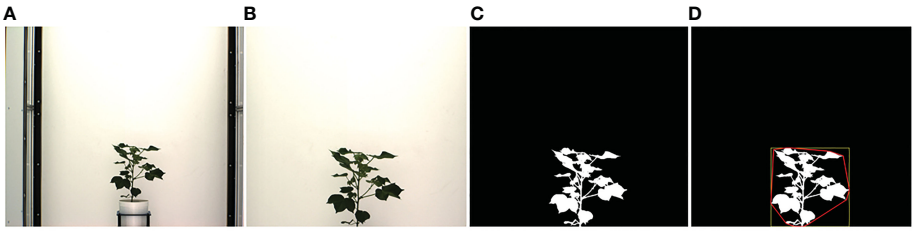


FIGURE 1
Illustration of holistic phenotype computation based on image analysis: **(A)** original image; **(B)** cropped image; **(C)** binary image; **(D)** plant enclosed by bounding rectangle and convex hull.

10.3389/fpls.2023.1003150

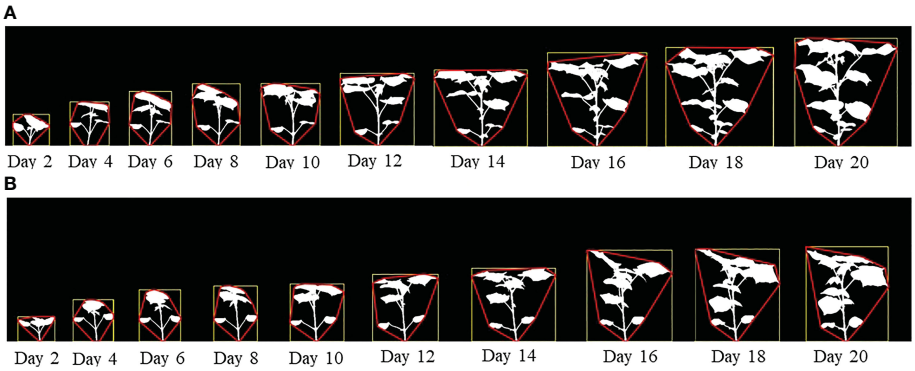


FIGURE 2
(A) An image sequence of a sample plant for side view angle of 0° (Experiment 1) enclosed by their bounding rectangles and convex hulls under controlled condition; and **(B)** An image sequence of a sample plant for side view angle of 0° (Experiment 1) enclosed by their bounding rectangles and convex hulls subjected to drought stress.

all admissible paths. For a path to be admissible, it should satisfy the following conditions: (a) boundary: $p_1=(1,1)$ and $p_L=(M,N)$; (b) monotonicity and all indices should appear at least once: $i_{l-1}\leq i_l\leq i_{l-1}+1$ and $j_{l-1}\leq j_l\leq j_{l-1}+1$. DTW minimizes the cost of warping P and Q together, i.e.,

$$DTW(P,Q)=\min_{p\in A(P,Q)}\left(\left(\sum_{(i,j)\in p}dist\left(P_i,Q_j\right)\right)^{1/2}\right) \quad (2)$$

Dynamic programming provides an exact solution to the optimization problem at hand. DTW constructs the $M\times N$ matrix of

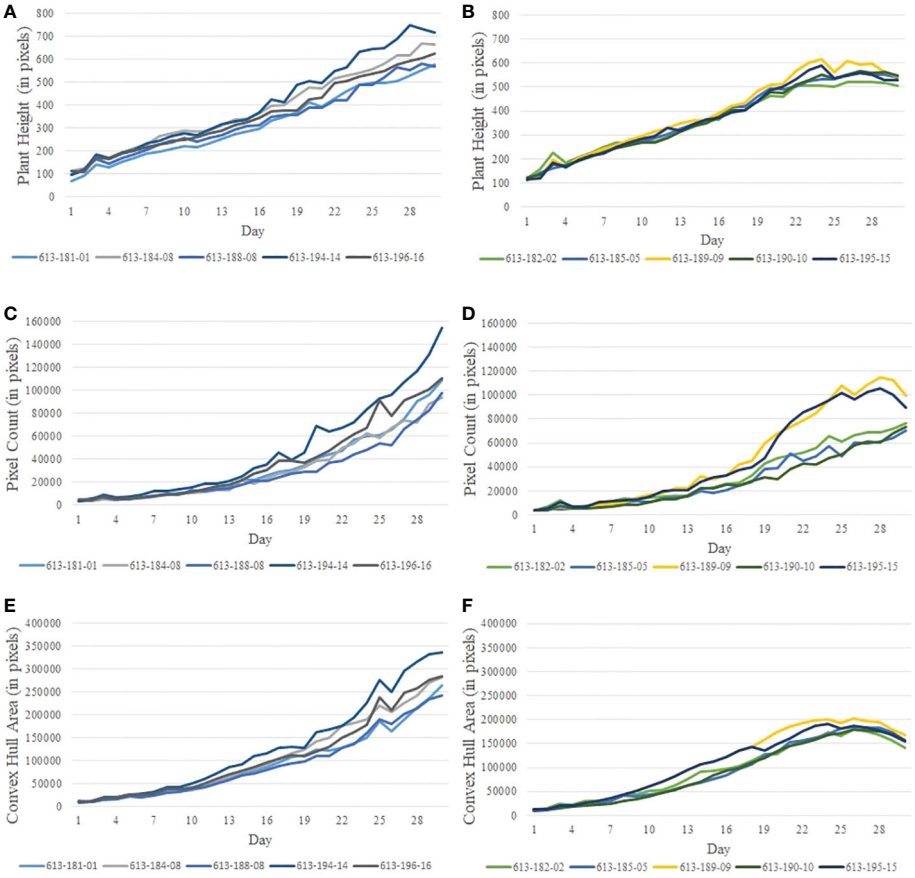


FIGURE 3
Illustration of nonlinear phenotypic time series using plants from Experiment 2: **(A, B)**- phenotypic time series for the height of plants under the controlled environment and subjected to drought stress, respectively; **(C, D)**- phenotypic time series for plant biomass (measured by pixel count) under the controlled environment and subjected to drought stress, respectively; and **(E, F)**- phenotypic time series for plant size (measured by the area of convex hull) under the controlled environment and subjected to drought stress, respectively.

Euclidean distances of corresponding phenotypes where $DTW_{i,j}$ is the distance between $P[1:i]=(P_1, P_2,\ldots,P_i)$ and $Q[1:j]=(Q_1, Q_2,\ldots,Q_i)$ with the best alignment given by the recurrence function given in Equation 3, i.e.,

$$DTW_{i,j} = dist(P_i, Q_j) + minimum(DTW_{i-1,j}, DTW_{i,j-1}, DTW_{i-1,j-1}) \quad (3)$$

where $dist(P_i,Q_j)=(P_i,Q_j)^2$.

Differences in environmental conditions, even in controlled environments, including water content and induced stress, result in variations in the phenotypic sequences for different plants of the same species. However, the plants undergoing stress will have fundamentally different phenotypic trajectories than those growing in normal conditions. Thus, dynamic time warping (DTW) distance is an ideal fit to compare the phenotypic trajectories of plants. The DTW distance between the phenotypic sequences of plants under similar conditions will be significantly different from those of plants under other conditions and can therefore form the basis to differentiate a normal growth sequence from a (drought) stress sequence. Note that all plant image sequences used in this study are of the same length, i.e., $M = N$. However, mechanical breakdown or the time-shared based imaging policy in an HTTP often results in the generation of image sequences of unequal lengths. Since DTW effectively compares time series of varying lengths, our proposed VisStressPredict algorithm will be suitable to deal with unforeseeable situations of generating unequal phenotypic time series in any phenotypic measurement environment. This also proves the generalizability of the algorithm.

In this paper, we propose a DTW-based approach to differentiate between control and stressed plants based on their phenotypic time series. Given a sequence $S = (S_1,S_2,\ldots,S_n)$ of length n and its subsequence $S_{sub}=(S_1,S_2,\ldots,S_i)$ of length i where $1 \le i \le n$, we classify the subsequence S_{sub} as either control or stressed. Two representative sequences R_c and R_s are calculated by element-wise averaging of a set of control and stressed sequences, respectively. Figures 6A, B show the representative phenotypic sequences of the height of a plant for control and stressed plants, respectively. Figures 6C, D show the representative phenotypic sequences of plant biomass (measured by total plant pixels) for control and stressed plants, respectively. Figures 6E, F show the representative phenotypic sequences of plant size (measured by convex-hull area) for control and stressed plants, respectively. The DTW distances D_c and D_s are calculated between S_{sub} and R_c and between S_{sub} and R_s, respectively. D_c and D_s are referred to as control DTW distance and stress DTW distance, respectively. The distance is normalized to obtain D_{norm}, which is then smoothened using a MA filter of window size N=6 (Equation 1) to obtain a stress factor, i.e., SF.

D_{norm} is given by:

$$D_{norm}(i) = \frac{|D_s(i) - D_c(i)|}{D_s(i)} \quad (4)$$

where $D_s(i)$ and $D_c(i)$ are given by
$D_s(i) = DTW(S_{sub},R_s)$ and $D_c(i) = DTW(S_{sub},R_c)$.
Finally, the stress factor, SF, for the subsequence is computed by:

$$SF(i) = Average(D_{norm}, i, N), \quad (5)$$

where $Average$ gives the average of the normalized distances in the window i–N+1 to i or 0. If the stress factor is above a predefined threshold, t^*, we label that subsequence as stressed. The threshold value t^* is defined as:

$$t^* = Median(Max_{SF}) \quad (6)$$

where Max_{SF} is the set of maximas of the SF's of control plants.

The stress factor threshold t^* is taken as the median of the maximas of SF's rather than the mean or maximum is to avoid any outliers in the control set from drastically affecting the threshold t^*.

Equation 7 gives the conditions for the predicted class.

$$Predicted\ \ Class(i) = \begin{cases} Stressed \text{ if } & SF(i) \ge t^* \\ Control \text{ otherwise} & \end{cases} \quad (7)$$

Finally, onset of the stress can be determined by identifying the first time stamp in the sequence to have the predicted class to be labeled "Stressed."

$$Onset(S) = \Theta : \forall \ \ 1 \le i < \Theta \ \ PC(i) = Control \wedge PC(\Theta) = Stressed \quad (8)$$

The proposed method is summarized in Algorithm 1.

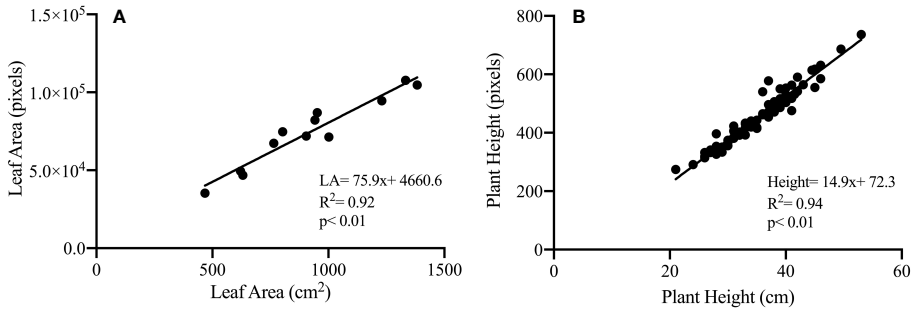


FIGURE 4
Illustration of correlation between phenotypic traits measured destructively and based on visible light image analysis: projected leaf area (cm^2) measured destructively and plant biomass (total plant pixels) derived from image analysis (cm^2) **(A)**; plant height (cm) measured destructively and plant height (pixels) derived from image analysis **(B)** for control and dry-down groups.

```
Require: S:array[1…n], S_sub: S[1:i] 1 ≤ i≤ n,listOfControlSeqs, listOfStressedSeqs
  function CALCREPRESENTATIVESEQ(listOfSeqs)
    SumOfSeqs←array[1…n]
    for each s in listOfSeqs do
      for i:= 1 to n do
          SumOfSeqs [i]← SumOfSeqs [i] + s[i]
      end for
    end for
    R_t←SumOfSeqs/length(listOfSeqs)
    return R_t
end function
function VISSTRESSPREDICT(S, S_sub, listOfControlSeqs, listOfStressedSeqs)
    R_c ← CALCREPRESENTATIVESEQ(listOfControlSeqs) ▷ Representative sequence for control
    R_s ← CALCREPRESENTATIVESEQ(listOfStressedSeqs) ▷ Representative sequence for stress
    D_norm←array[1…n]
    D_normMA←array[1…n]
    flag ← 0
    day ← 0
    for i = 1 to length(S) do
        S_sub← S[1:i]
        D_c ← DTWDistance(R_c,S_sub)
        D_s ← DTWDistance(R_s,S_sub)
        D_norm[i]← |D_s - D_c |/D_s
        D_normMA[i]←Average(D_norm, i, N)
        if DNormMA[i]>t* then
          predictedClass ← "Stressed"
          if flag == 0 then
            firstStressDay ← i
            flag ← 1
          end if
        else
          predictedClass ← "Control"
        end if
      end for
      return predictedClass, firstStressDay
end function
```

ALGORITHM 1
Classify control and stressed sequences and predict onset of stress

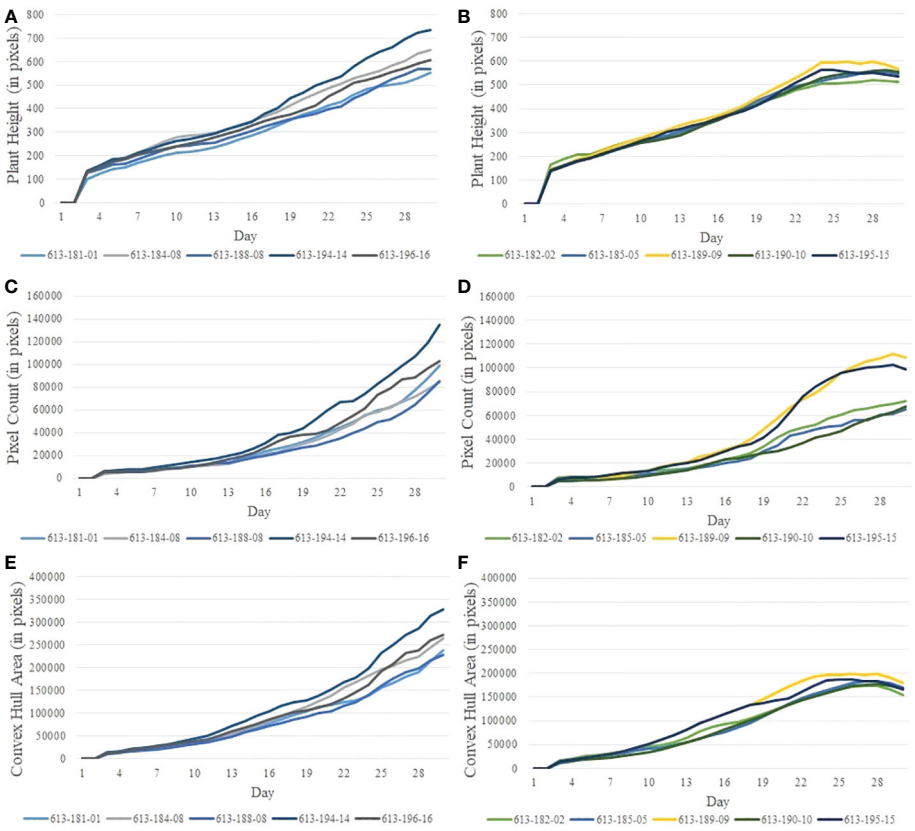


FIGURE 5
Illustration of smoothened time series using plants from Experiment 2. **(A, B)** plant height for control and stressed plants, respectively; **(C, D)** plant biomass (measured by pixel count) for control and stressed plants, respectively; and **(E, F)** plant size (measured by the area of convex hull enclosing the plant) for control and stressed plants, respectively.

2.3 HyperStressPropagateNet: Deep neural network based temporal stress propagation using hyperspectral imagery

A hyperspectral image can be represented by a three-dimensional array of intensities, $H(x,y,\lambda)$, where (x,y) represents the location of a pixel and λ denotes the wavelength. Thus, it is often referred to as a hyperspectral cube. Intensities at a given wavelength can be represented as a two-dimensional image, and intensity information at a specific location for all wavelengths can be represented by a spectral reflectance curve.

2.3.1 Segmentation

We use a spectral band difference-based segmentation approach to create the mask of the plant for subsequent analysis. This segmentation method is useful and efficient for plant phenotyping analysis using hyperspectral or multispectral imagery, since the goal is to analyze only the plant ignoring the background. The segmentation process is illustrated in Figure 7. In this approach, two bands of specific wavelengths that have significant contrast in intensity are first identified (Figures 7A, B), then enhanced by multiplying a constant factor (Figures 7C, D), and finally subtracted from each other to isolate the plant pixels, i.e., the foreground (Figure 7E). Based on empirical analysis, the two wavelengths that are effective are 770 nm and 680 nm, and the constant factor is 2. Thus, the enhanced foreground image, (I_f), is given by:

$$I_f = 2{\star}I_{770} - 2{\star}I_{680}, \tag{9}$$

where I_{770} and I_{680} are the images at 770 nm and 680 nm wavelengths, respectively. The enhanced foreground image is then binarized using Otsu's automatic thresholding technique (Otsu, 1979) to generate a binary mask for the plant (Figure 7F), which is then used to segment the plant in all bands of a hyperspectral cube for subsequent analysis. Otsu's method chooses a global threshold so as to maximize the separability of the resultant classes in gray levels. This threshold is then used to convert a grayscale image to a binary image. In this paper, we used graythresh() function of Matlab to generate the global threshold followed by imbinarize() to create the binary mask.

2.3.2 Hyper-pixel generation

A hyper-pixel (HP) is defined as HP = $\{P_{410,\ldots}\ P_{800}\}$, where P_i denotes a plant pixel at the wavelength i. A reflectance spectrum is generated at each hyper-pixel by plotting the grayscale value of the hyper-pixel over the wavelength range. Figures 8A, B show the reflectance spectra generated at randomly selected pixels from a controlled and a stressed plant, respectively. Stomatal response, reactive oxygen species scavenging, metabolic rate, water absorption, and photosynthetic capability are all affected when plants are subjected to drought stress. These collective responses lead to an adjustment in the growth rate of plants as an adaptive

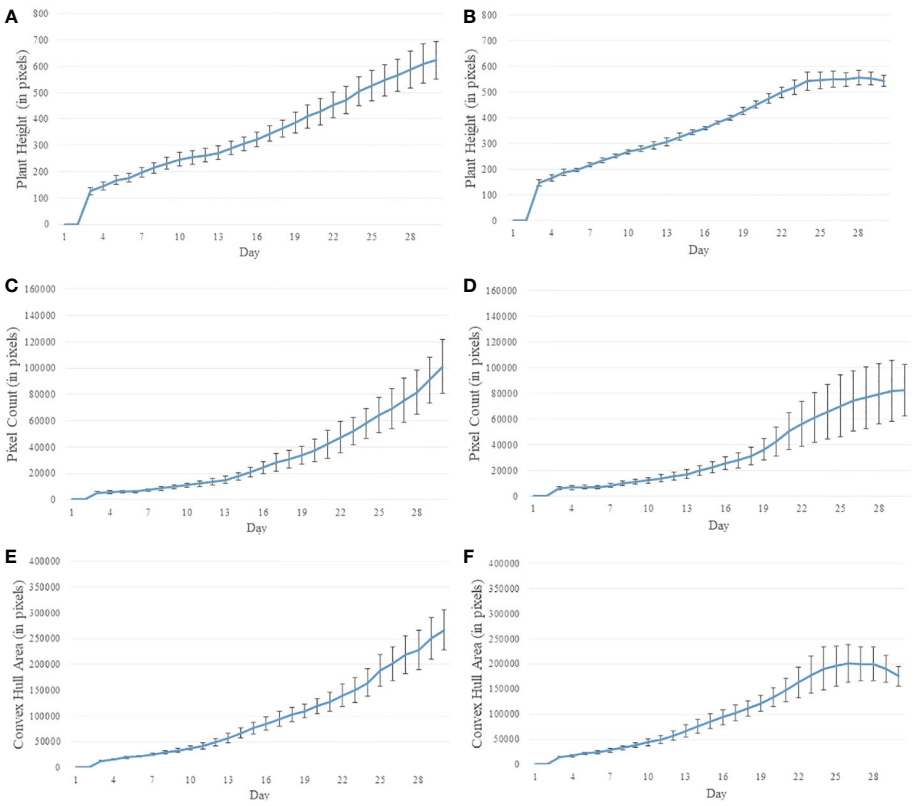


FIGURE 6
Illustration of representative phenotypic time series with mean and standard deviation using plants from Experiment 2. **(A, B)** plant height for control and stressed plants respectively; **(C, D)** plant biomass for control and stressed plants respectively; and **(E, F)** plant size for control and stressed plants respectively.

response for survival (Osakabe et al., 2014). This phenomenon creates differences in the reflectance spectra at different wavelength ranges generated from hyperspectral imagery of the stressed and controlled plants. It is seen from Figure 8A that the reflectance spectra from the controlled plant are very similar. The comparatively dispersed nature of the reflectance spectra of the stressed plant (Figure 8B) can be attributed to the varying stress at different parts of the plant. We observe a relatively sharp dip in the reflectance spectra of the stressed plant compared to the controlled plant approximately in the wavelength range of 1200–1300 nm. The difference in the reflectance spectra between the controlled and the stressed plant forms the basis of this algorithm. Note that a sharp decrease in reflectance between 1400–1600 nm wavelength range is guided by the physiological characteristics of the plants. This wavelength range is known for atmospheric water absorption, and is sensitive to vapor reflectance. In this range, light absorption by the plants is significantly high resulting in low gray-scale values in their hyperspectral imagery.

2.3.3 Training and classification

Convolutional neural network (CNN) models have been effective in various computer vision applications, including segmentation, classification, object recognition, biometrics, and medical imaging (LeCun and Bengio, 1995; Kolhar and Jagtap, 2021). Recently, 1-dimensional (1D) CNNs have been used in natural language processing, speech recognition, and biomedical signal processing where they can perform feature extraction and classification tasks in a single end-to-end model (Jang et al., 2020). In this paper, we use a 1D CNN to classify the reflectance spectra into two classes, i.e., stressed and unstressed. These convolutional layers learn from the representation learning component. Each convolutional layer consists of multiple (N) filters. Each filter of the convolutional layer learns a different feature. The goal of representation learning is to learn the different features in the convolution layers and then use them in the subsequent dense layers for the final classification. The architecture of the proposed network is shown in Figure 9. The proposed network consists of two components: representation learning and classification. The details of the network architecture are given below.

The representation learning component consists of four 1D convolutional layers. The size of the input vector is (m, 243), where m is the number of training examples, each consisting of 243 reflectance values corresponding to a reflectance spectrum. The dimension of this input vector is increased to (m, 243, 1) to feed it into a 1D convolutional layer. The 1D CNN layer is followed by a rectified linear unit (ReLU) activation function. There are four such successive 1D CNN layers with ReLU activation. They each have a kernel size of 5 and a stride of 1. The first two convolutional layers have 64 filters, and the 'same' padding is used while the last two convolutional layers have 128 filters with the 'valid' padding.

The feature vectors obtained after the convolutions are fed to the classification component, which consists of two dense layers. First, the output of the convolutional step, which is a vector of size ($m \times 235 \times 128$), is 'flattened' to a vector of size (m, 30080). The flattened vector is then fed to a dense layer in the classification component, which has 32

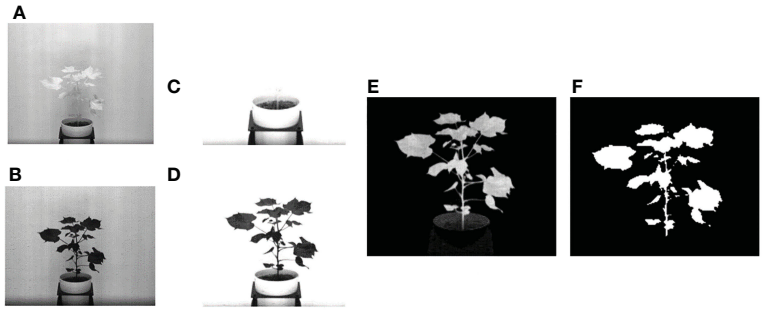


FIGURE 7
Illustration of spectral band difference based segmentation: **(A, B)** - hyperspectral images of a cotton plant at wavelengths 770 nm and 680 nm, respectively; **(C, D)** - corresponding enhanced images; **(E)** image obtained after subtracting **(C)** from **(D)**; and **(F)** binary image.

filters with ReLU activation. The (m, 32) vector thus obtained from the dense layer is passed to another dense layer with a sigmoid activation for binary classification between stressed and unstressed classes. The groundtruth for training is developed based on visual inspection of the RGB images of the plants. Drooping can be seen in plants in dry down stage for the last three days of the experiment. The hyperpixels for these last three days are labeled as stressed. The hyperpixels of the plants for all days under the controlled environment are marked as unstressed for the purpose of groundtruth generation. The labeled dataset is split into training, validation, and test sets in the ratio of 0.64, 0.16, and 0.20.

2.3.4 Evaluation metrics

HyperStressPropagateNet has been evaluated using a confusion matrix, precision-recall curve, and F_1-score. These metrics are defined as follows:

- Confusion matrix is a specific tabular representation that allows the visualization of the performance of an algorithm, and is extensively used in the case of statistical classification problems. For a confusion matrix C, $C_{i,j}$ is equal to the number of observations known to be in class i but predicted to be in class j. Thus, $C_{0,0}$ is the true negatives (T_N), $C_{1,0}$ is the false negatives (F_N), $C_{0,1}$ is the false positives (F_P), and $C_{1,1}$ is the true positives (T_P).
- F1-Score is the harmonic mean of precision and recall. The range for F1-Score is [0, 1], with 0 being the worst and 1 being the best prediction. It is defined as:

$$F1 - Score = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{10}$$

Precision (P) is defined as $T_P/(T_P+F_P)$ and recall (R) is defined as $T_P/(T_P+F_N)$. F_1-score is a better measure than accuracy for unbalanced datasets.

3 Experimental results

3.1 VisStressPredict: DTW based stress prediction using visible light imagery

The stress factor (SF) for each plant is calculated using Equation 5. If the stress factor (SF) for a particular plant on a certain day crosses the threshold t^*, it is predicted to be stressed from that day. The predicted class and onset of stress are given by Equations 7 and 8, respectively. Figures 10A, B show the stress factor as a function of time (called as a stress factor curve) for a set of normal and stressed plants, respectively. The figures show that the plants demonstrate similar group behavior. The stress factor curves for normal plants gradually increase, peak around the threshold t^*, and then gradually decrease. The stress factor curves for the stressed plants, on the other hand, generally keep increasing for the duration of the study.

It is seen from Figure 10B that the stress factor curves for stressed plants with plant IDs 613-182-02, 613-185-05, 613-190-10, and 613-195-15 cross the threshold t^* on Day 24, Day 23, Day 23, and Day 22, respectively, whereas stress factor curves for the control plants remain below the threshold during the course of the study (Figure 10A). The

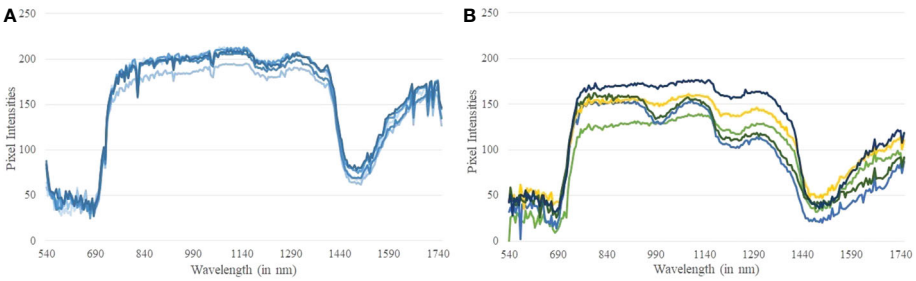


FIGURE 8
(A) Reflectance spectra generated at random pixels of a controlled plant; and **(B)** Reflectance spectra generated at random pixels of a stressed plant.

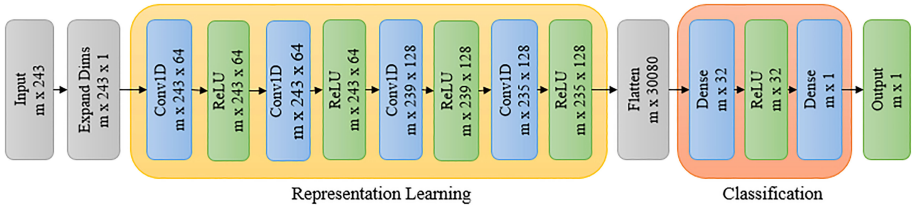


FIGURE 9
Deep learning architecture for classification of stressed and unstressed pixels.

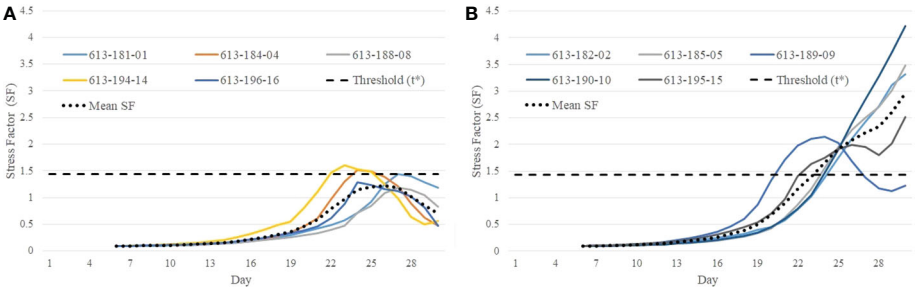


FIGURE 10
Illustration of difference in behavioral characteristics between control and stressed plants in terms of stress factor curves: **(A)** stress factor curves for control plants; and **(B)** stress factor curves for stressed plants.

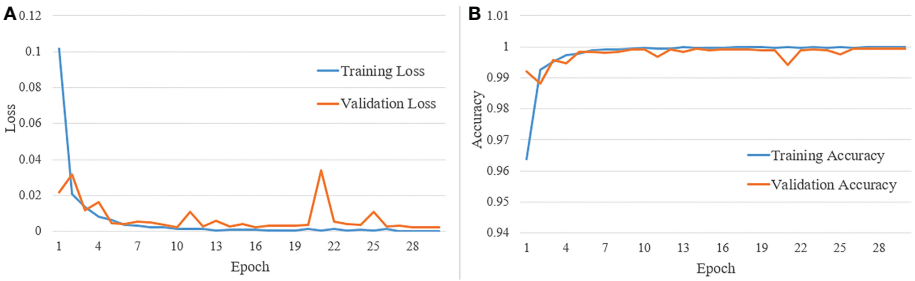


FIGURE 11
(A) Training and validation loss vs number of epochs; and **(B)** training and validation accuracy vs number of epochs.

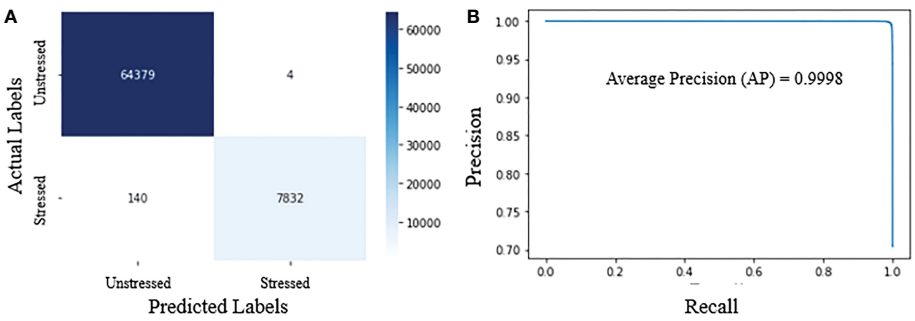


FIGURE 12
Performance metrics for HyperStressPropagateNet: **(A)** confusion matrix; and **(B)** precision-recall curve.

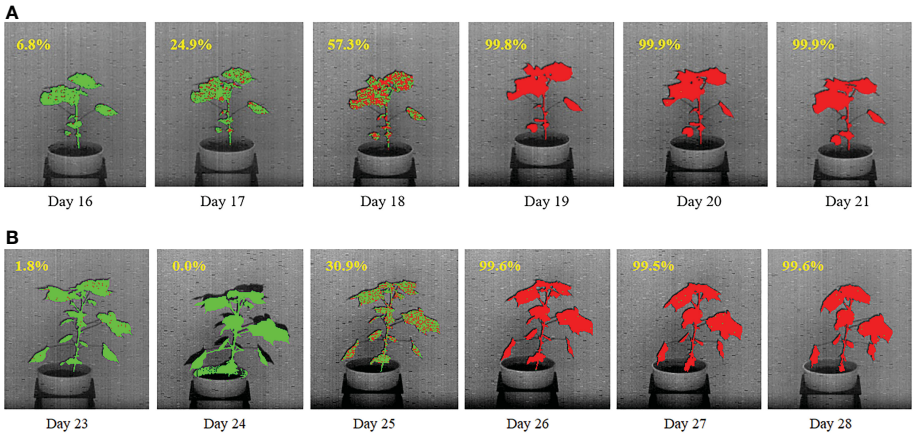


FIGURE 13
Illustration of qualitative and quantitative temporal propagation of stress using **(A)** a plant from DD1 group and **(B)** a plant from DD2 group. The percentage of stress pixels are shown at the top-left corner of each image.

only exception is a stressed plant (ID 613-189-09) in Figure 10B whose stress factor curve keeps decreasing after it reaches a peak. This is likely an outlier and may be caused due to imaging artifacts or an anomaly in the computation of the convex-hull area due to plant rotation. These defects may occur in plant images and can be fixed with a more rigorous image processing and correction pipeline.

3.2 HyperStressPropagateNet: Deep neural network based temporal stress propagation using hyperspectral imagery

Figure 11A shows the training and validation loss versus the number of epochs, and Figure 11B shows the training and validation accuracy versus the number of epochs. The total number of epochs used during training is 30. From the two sets of graphs, it is evident that the validation loss and accuracy closely follow the training loss and accuracy, respectively. Also, the model converges, and validation accuracy reaches above 95% within 10 epochs.

Figure 12A shows the confusion matrix, demonstrating the accuracy of classifying hyperpixels into stressed and unstressed classes. The confusion matrix in Figure 12A shows that at a threshold probability of 0.5, the false positives and false negatives are extremely low. Precision and recall for our proposed classifier are 0.99 and 0.98, respectively. F_1-score is 0.98. The very high values for precision, recall, and F_1-score show that the model can accurately distinguish between stressed and unstressed spectra.

Figure 12B shows the precision-recall curve for different thresholds for the predicted probabilities. The model outputs the probabilities of pixels being stressed from which the predictions are obtained using a threshold. This threshold is generally kept as 0.5. As the threshold is increased from 0 to 1.0, the predictions obtained from the probabilities vary, and so do the precision and recall values. The model with the highest area under the precision-recall curve is generally deemed optional. Figure 12B shows that the precision and recall values are very high for the entire range of threshold for the proposed model, thus giving a very high area under the precision-recall curve close to 1.0. The average precision for the model is also very high, i.e., 0.9998. The various performance metrics demonstrate the efficacy of the proposed algorithm.

Figures 13A, B show the temporal propagation of stress using hyperspectral image sequences of cotton plants from Experiment 1 (Plant ID: 613-200-20) and Experiment 2 (Plant ID: 613-195-15), respectively. In this figure, the hyperpixels classified as stressed and unstressed are shown in red and green, respectively, for qualitative visualization of temporal stress propagation. The percentage of the

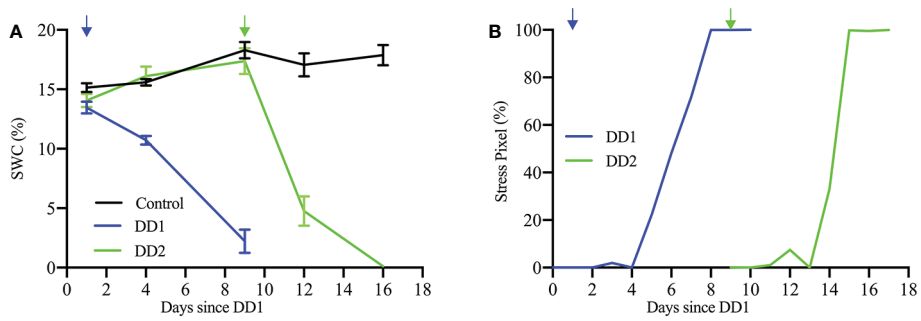


FIGURE 14
(A) SWC (%) for the control and the two dry-down groups (DD1, Plant ID: 613-200-20 and DD2, Plant ID: 613-195-15); and **(B)** stress pixel (%) over days since DD1 for the same plants.

stressed pixels to the total plant pixels is shown at the top left of each image.

For the plant in Figure 13A, drought stress was introduced on Day 13. The figure shows a gradual increase in the stress symptoms that started on Day 16, when 6.8% of the plant is labeled as stressed (marked in red). The percentage of stress pixels increased to 24.9% on Day 17 and 57.3% on Day 18. There is almost no green pixel present in the plant on the last two days, i.e. on Day 20 and Day 21, which implies that the whole plant is stressed. For the plant in Figure 13B, drought stress was introduced on Day 20. The figure shows that stress symptoms did not appear in the plant until Day 25 when, 30.9% of the plant are labeled as stressed (marked in red). The figure shows that the plant is considerably stressed on Days 26 and 27, with very few unstressed pixels (shown in green). There is almost no green pixel present in the plant on Day 28, which implies that the whole plant is stressed.

The limited water availability in the soil is confirmed by changes in the soil water content (SWC), as shown in Figure 14A. Soil water content was measured using a HH2 type Moisture Meter (Eijkelkamp, NL) connected to a ML3 ThetaProbe Soil Moisture Sensor (Delta-T Devices, UK). Dry-down treatment resulted in a decline of SWC from initial conditions (field capacity ~ 16% SWC) to 15.6% of field capacity (~ 2.5% SWC) and 6.3% of field capacity (~ 1.0% SWC) for DD1 and DD2, respectively, at the end of each dry-down period. Figure 14A also indicates the immediate effect in the SWC that follows the cessation of watering. Figure 14B shows the temporal progression of percentage of stress pixels for a plant from Experiment 1 (Plant ID: 613-200-20) and a plant from Experiment 2 (Plant ID: 613-195-15). The quantitative visualization of temporal stress propagation of these two plants are shown in Figures 13A, B, respectively. The excellent correlation between the SWC and the corresponding temporal progression of the percentage of stress pixels computed by HyperStressPropagateNet for both experiments demonstrates the efficacy of the proposed method.

4 Discussion

The paper introduces two novel algorithms to understand the impact of stress on plants. First, an approach to predict the onset of stress in drought-affected plants is presented. The algorithm, named as VisStressPredict, uses an extension of dynamic time warping based on the time-series analysis of plant phenotypes derived from visible light image sequences. The paper also introduces a novel method, i.e., HyperStressPropagateNet, to examine the propagation of stress in plants over time. The deep learning based algorithm uses a convolutional neural network to classify hyperpixels into stressed and unstressed categories. Although both methods have been evaluated using cotton plant image sequences, they can be generalized to any plant species to study the temporal effect of any kind of stress, e.g., thermal and salinity. Thus, the methods have the potential to help differentiate between stress-tolerant and stress-susceptible genotypes for sustainable agriculture. Note that VisStressPredict and HyperStressPropagateNet fundamentally differ in their goals and hence in the input image sequences, underlying approaches, and final outcomes. VisStressPredict identifies the onset of stress on the plant as a whole, but HyperStressPropagateNet maps the stress in a plant at a finer resolution. However, the onset of stress as predicted by VisStressPredict (Figure 10) correlates extremely well with the day of appearance of stress pixels in the plants as computed by HyperStressPropagateNet (Figure 13). This establishes the dove-tailed relationship between the two proposed algorithms.

The efficacy of the proposed algorithms depends on the reliability of the phenotypes. The accuracy of the computed phenotypic time series depends on many factors, including accuracy in image segmentation, the effectiveness of denoising, and the stability of the plant structure derived from images. The phenotypes such as plant height, plant area, and convex-hull area are derived from the RGB images through a series of image processing steps (See Section 2.2.1). Segmentation of the plant is the basis for image-based phenotypic computation, and inaccuracies in segmentation will result in imprecise computation of plant phenotypes, including its height and convex-hull area (Das Choudhury, 2020a). In addition, inherent challenges introduced while imaging the plant, such as those due to plant rotation, also impact the accuracy of some phenotypes, including plant area and convex-hull area (Bashyam et al., 2021). The plant rotation may cause shrinking of the convex-hull area computed from the imagery from the previous day, although the plant has grown bigger (Maddonni et al., 2002; Das Choudhury et al., 2016). All these factors result in unevenness in the phenotypic time series (Figure 3). This unevenness affects the performance of the subsequence-based DTW matching, which explains the outlier stress factor curve (plant ID 613-189-09) in Figure 10B. The impact of the error may be ameliorated to some extent by smoothening, as explained in Section 2.2.2.

Finally, it is worth noting that even with a very limited number of stress days in the dataset, the proposed VisStressPredict algorithm shows excellent performance as expressed by the empirically determined stress factor. The mean stress factor curve in Figure 10A remains below the threshold during the course of the study for the control plants, whereas, it crosses the threshold on Day 23 and keeps on increasing during the rest of the days for the stressed plants (Figure 10B). The method's potential to predict stress, even in its early stages, demonstrates its efficacy. However, in future work, we will explore the generality of the method by examining the performance of the algorithms on a large dataset with different plant species where plants are subjected to stress for a longer duration.

The dataset used in the study consists of images of cotton plants that are visibly drooped (but not visibly dried as seen by a change of color) under stress. Thus, it is not possible to quantify the stress at the fine pixel scale based on analyzing color features using visible light images. The hyperspectral image analysis for temporal stress propagation achieves the novel objective of identifying the stress location in the plant before the visible stress symptoms appear in the plant. Our study shows an excellent correlation between the soil water content and the percentage of stress pixels in the plants (Figure 14). The figure shows that as the soil water content decreases, the stress in plants increases. The Pearson correlation coefficients calculated for SWC and stress pixel percentage for the said plants from the two dry-down groups (DD1, Plant ID: 613-200-20 and DD2, Plant ID: 613-195-15) are -0.972 and -0.735, respectively. The early detection of stress susceptibility acts as an alarm to the deteriorating plant health, and appropriate intervention, e.g., adequate watering of the plant, may help recover the plant's health. Future work will consider the

identification of wavelengths that carry the most salient information on drought stress prior to the classification for improved computational complexity.

5 Conclusion

The paper introduces two novel algorithms, i.e., VisStressPredict and HyperStressPropagateNet, to study stress response in plants in greater spatial and temporal resolution by analyzing visible light and hyperspectral imagery. While RGB cameras capture the visible part of the light spectrum in only three broad bands (red, green, and blue), hyperspectral cameras typically capture a broad range of wavelengths at very narrow intervals of a few nanometers. The VisStressPredict algorithm predicts the onset of stress in plants using an enhanced dynamic time warping approach from the phenotypic time series derived from visible light images. The HyperStressPropagateNet algorithm, in contrast, identifies the location of stress in the plants using a deep learning approach from the hyperspectral imagery. The algorithm has been used to illustrate the temporal propagation of stress both qualitatively and quantitatively. The efficacy of the two algorithms is demonstrated using a set of control and drought-stressed cotton plants imaged in an HTTP system. Both the algorithms have the potential to examine the response to other kinds of biotic and abiotic stresses in plants, and can be applied to any kind of plant species.

Data availability statement

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

Author contributions

SD conceived the idea, developed and implemented the algorithm, conducted experimental analysis, led the manuscript writing and supervised the research. SS developed and implemented the algorithm, conducted experimental analysis and contributed to the manuscript writing. AM led the dataset design, performed validation experiments and contributed to the manuscript writing. AS and TA critically reviewed the manuscript and provided constructive feedback throughout the process. All authors contributed to the article and approved the submitted version.

Funding

This research is partially supported by the Nebraska Agricultural Experiment Station with funding from the Hatch Act capacity program (Accession Number 1011130) from the USDA National Institute of Food and Agriculture.

Acknowledgments

Authors are thankful to Dr. Vincent Stoerger, the Plant Phenomics Operations Manager at the University of Nebraska-Lincoln, USA, for his support in setting up experiments to create the dataset used in this study.

Conflict of interest

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

Publisher's note

References

Azimi, S., Wadhawan, R., and Gandhi, T. K. (2021). Intelligent monitoring of stress induced by water deficiency in plants using deep learning. *IEEE Trans. Instrument. Meas* 70, 1–13. doi: 10.1109/TIM.2021.3111994

Bashyam, S., Choudhury, S. D., Samal, A., and Awada, T. (2021). Visual growth tracking for automated leaf stage monitoring based on image sequence analysis. *Remote Sens.* 13. doi: 10.3390/rs13050961

Bertrand, J., Sudduth, T., Condon, A., Jenkins, T., and Calhoun, M. (2005). Nutrient content of whole cottonseed. *J. Dairy Sci.* 88, 1470–1477. doi: 10.3168/jds.S0022-0302(05)72815-0

Comas, L., Becker, S., Cruz, V. M., Byrne, P. F., and Dierig, D. A. (2013). Root traits contributing to plant productivity under drought. *Front. Plant Sci.* 4, 442. doi: 10.3389/fpls.2013.00442

Dabbert, T., and Gore, M. (2014). Challenges and perspectives on improving heat and drought stress resilience in cotton. *J. Cotton Sci.* 18, 393–409.

Das Choudhury, S. (2020a). "Segmentation techniques and challenges in plant phenotyping," in *Intelligent image analysis for plant phenotyping*. Eds. A. Samal and S. Das Choudhury (Boca Raton, Florida: CRC Press, Taylor & Francis Group), 69–91.

Das Choudhury, S. (2020b). "Time series modeling for phenotypic prediction and phenotype-genotype mapping using neural networks," in *European Conference on computer vision workshop*. Ed. N. L. (Computer Science (Springer), 228–243.

Das Choudhury, S., Bashyam, S., Qiu, Y., Samal, A., and Awada, T. (2018). Holistic and component plant phenotyping using temporal image sequence. *Plant Methods* 14. doi: 10.1186/s13007-018-0303-x

Das Choudhury, S., Guha, S., Das, A., Das, A. K., Samal, A., and Awada, T. (2022). Flowerphenonet: Automated flower detection from multi-view image sequences using deep neural networks for temporal plant phenotyping analysis. *Remote Sens.* 14. doi: 10.3390/rs14246252

Das Choudhury, S., and Tjahjadi, T. (2013). Gait recognition based on shape and motion analysis of silhouette contours. *Comput. Vision Image Und.* 117, 1770–1785. doi: 10.1016/j.cviu.2013.08.003

Das Choudhury, S., Vincent, S., Samal, A., Schnable, J., Liang, Z., and Yu, J.-G. (2016). "Automated vegetative stage phenotyping analysis of maize plants using visible light images," in *KDD workshop on data science for food, energy and water*(San Francisco, California, USA).

Dyrmann, M. (2015). "Fuzzy c-means based plant segmentation with distance dependent threshold," in *Proceedings of the computer vision problems in plant phenotyping (CVPPP)*, vol. 5. (BMVA Press), 1–5.11. doi: 10.5244/C.29.CVPPP.5

Gampa, S., and Quinones, R. (2020). "Data-driven techniques for plant phenotyping using hyperspectral imagery," in *Intelligent image analysis for plant phenotyping*. Eds. A. Samal and S.D. Choudhury (Boca Raton, Florida: CRC Press, Taylor & Francis Group), 175–193.

Jang, B., Kim, M., Harerimana, G., Kang, S.-u., and Kim, J. W. (2020). Bi-lstm model to increase accuracy in text classification: Combining word2vec cnn and attention mechanism. *Appl. Sci.* 10, 5841. doi: 10.3390/app10175841

Kolhar, S., and Jagtap, J. (2021). Convolutional neural network based encoder-decoder architectures for semantic segmentation of plants. *Ecol. Inf.* 64, 101373. doi: 10.1016/j.ecoinf.2021.101373

LeCun, Y., and Bengio, Y. (1995). "Convolutional networks for images, speech, and time series," in *The handbook of brain theory and neural networks*, (Cambridge, MA, USA:MIT Press) vol. 3361.

Li, L., Zhang, Q., and Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors* 14, 20078–20111. doi: 10.3390/s141120078

Maddonni, G. A., Otegui, M. E., Andrieu, B., Chelle, M., and Casal, J. J. (2002). Maize leaves turn away from neighbors. *Plant Physiol.* 130, 1181–1189. doi: 10.1104/pp.009738

Ort, D. R., Merchant, S. S., Alric, J., Barkan, A., Blankenship, R. E., Bock, R., et al. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci.* 112, 8529–8536. doi: 10.1073/pnas.1424031112

Osakabe, Y., Osakabe, K., Shinozaki, K., and Tran, L. P. (2014). Response of plants to water stress. *Front. Plant Sci.* 5, 86. doi: 10.3389/fpls.2014.00086

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybernet.* 9, 62–66. doi: 10.1109/TSMC.1979.4310076

Pettigrew, W. T. (2004). Physiological consequences of moisture deficit stress in cotton. *Crop Sci.* 44, 1265–1272. doi: 10.2135/cropsci2004.1265

Sakoe, H., and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* 26, 43–49. doi: 10.1109/TASSP.1978.1163055

Sheffield, J., Wood, E. F., and Roderick, M. L. (2012). Little change in global drought over the past 60 years. *Nature* 491, 435–438. doi: 10.1038/nature11575

Taha, M. F., Abdalla, A., ElMasry, G., Gouda, M., Zhou, L., Zhao, N., et al. (2022). Using deep convolutional neural network for image-based diagnosis of nutrient deficiencies in plants grown in aquaponics. *Chemosensors* 10. doi: 10.3390/chemosensors10020045

Townsend, T., and Sette, J. (2016). "Natural fibres and the world economy," in *Natural fibres: Advances in science and technology towards industrial applications* (Springer), 381–390.

Wang, M., Ellsworth, P. Z., Zhou, J., Cousins, A. B., and Sankaran, S. (2016). Evaluation of water-use efficiency in foxtail millet (setaria italica) using visible-near infrared and thermal spectral sensing techniques. *Talanta* 152, 531–539. doi: 10.1016/j.talanta.2016.01.062

Zheng, C., Abd-Elrahman, A., and Whitaker, V. (2021). Remote sensing and machine learning in crop phenotyping and management, with an emphasis on applications in strawberry farming. *Remote Sens.* 13, 531. doi: 10.3390/rs13030531

frontiers | Frontiers in Plant Science

TYPE Methods
PUBLISHED 17 February 2023
DOI 10.3389/fpls.2023.1122833

OPEN ACCESS

EDITED BY
Baohua Zhang,
Nanjing Agricultural University, China

REVIEWED BY
Xinyu Guo,
Beijing Research Center for Information Technology in Agriculture, China
Honghai Luo,
Shihezi University, China

*CORRESPONDENCE
Liantao Liu
liultday@126.com
Nan Wang
cmwn@163.com

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

SPECIALTY SECTION
RECEIVED 13 December 2022
ACCEPTED 07 February 2023
PUBLISHED 17 February 2023

CITATION
Yu Q, Tang H, Zhu L, Zhang W, Liu L and Wang N (2023) A method of cotton root segmentation based on edge devices. *Front. Plant Sci.* 14:1122833. doi: 10.3389/fpls.2023.1122833

A method of cotton root segmentation based on edge devices

Qiushi Yu[1], Hui Tang[1], Lingxiao Zhu[2], Wenjie Zhang[3], Liantao Liu[2*] and Nan Wang[1*]

[1]College of Mechanical and Electrical Engineering, Hebei Agricultural University, Baoding, China, [2]College of Agronomy, Hebei Agricultural University, Baoding, China, [3]College of Modern Science And Technology, Hebei Agricultural University, Baoding, China

The root is an important organ for plants to absorb water and nutrients. *In situ* root research method is an intuitive method to explore root phenotype and its change dynamics. At present, *in situ* root research, roots can be accurately extracted from *in situ* root images, but there are still problems such as low analysis efficiency, high acquisition cost, and difficult deployment of image acquisition devices outdoors. Therefore, this study designed a precise extraction method of *in situ* roots based on semantic segmentation model and edge device deployment. It initially proposes two data expansion methods, pixel by pixel and equal proportion, expand 100 original images to 1600 and 53193 respectively. It then presents an improved DeeplabV3+ root segmentation model based on CBAM and ASPP in series is designed, and the segmentation accuracy is 93.01%. The root phenotype parameters were verified through the Rhizo Vision Explorers platform, and the root length error was 0.669%, and the root diameter error was 1.003%. It afterwards designs a time-saving Fast prediction strategy. Compared with the Normal prediction strategy, the time consumption is reduced by 22.71% on GPU and 36.85% in raspberry pie. It ultimately deploys the model to Raspberry Pie, realizing the low-cost and portable root image acquisition and segmentation, which is conducive to outdoor deployment. In addition, the cost accounting is only $247. It takes 8 hours to perform image acquisition and segmentation tasks, and the power consumption is as low as 0.051kWh. In conclusion, the method proposed in this study has good performance in model accuracy, economic cost, energy consumption, etc. This paper realizes low-cost and high-precision segmentation of *in-situ* root based on edge equipment, which provides new insights for high-throughput field research and application of *in-situ* root.

KEYWORDS

***in situ* root, high-throughput phenotype, low-cost acquisition, semantic segmentation, edge equipment**

1 Introduction

Roots play a crucial role in the absorption of water and nutrients by plants, affecting plant health, environmental adaptation and productivity (Hinsinger et al., 2011; Lynch and Wojciechowski, 2015; Paez-Garcia et al., 2015). Microroots (mainly composed of fine roots and root hairs) are the main executive parts of roots. The absorption of water and nutrients accounts for more than 75% of the total absorption of roots (Nielsen et al., 2001). The dynamic changes of their own morphological characteristics (Shan and Tao, 1992) significantly affect root function and plant growth. At present, root phenotype research focuses on the accurate identification of root architecture. However, the segmentation of plant roots from the cultivation environment is vulnerable to the impact of small and medium soil particles. At the same time, it is difficult to accurately segment the edges of roots and soil, which restricts the acquisition of accurate root images.

In order to solve the above problems, scholars at home and abroad have conducted a lot of relevant research. Obtaining high-resolution images of roots in soil is the basis for accurate identification of root configuration. Traditional root acquisition methods, such as root drilling, soil column method and profile method, consume materials and manpower. Problems such as damage to root configuration and loss of small root segments are easy to occur during extraction, which cannot meet the dynamic and accurate identification of root configuration, It has been replaced by *in situ* root observation (*in situ* cultivation method and *in situ* imaging method) (Xiao et al., 2020; Liu et al., 2020b). The root *in situ* imaging method originated from the micro root canal method (Bates, 1937; Cseresnyés et al., 2021; Rajurkar et al., 2022) It refers to identifying the root image contacting the glass tube wall by inserting a glass tube into the soil. However, its disadvantages lie in poor resolution (numerical value), slow acquisition speed (time), and low degree of automation. It is difficult to achieve batch synchronization and real-time acquisition of the original root image. In addition, X-ray tomography (XCT) and nuclear magnetic resonance imaging (MRI) commonly used in medicine also provide new methods and means for the acquisition of *in situ* root images (Jahnke et al., 2009). XCT scans the root image by using the characteristics of different attenuation degrees of X-ray passing through the soil and root, and finally obtains the root image (Park et al., 2020; Scotson et al., 2021; Ferreira et al., 2022). MRI is a modern tomographic imaging technology, which mainly transmits radio frequency electromagnetic waves to obtain the MRI information of different positions of objects in the magnetic field to generate images, and uses computers to reconstruct the internal images of objects(Borisjuk et al., 2012). It has also been applied in root research (Schneider et al., 2020; Horn et al., 2021; Pflugfelder et al., 2021). However, there are still drawbacks to the above two technologies. Among them, XCT imaging takes a long time to acquire, while MRI is more suitable for acquiring large roots. Neither of them can recognize that the diameter is less than 400 μ M (Metzner et al., 2015), and both technologies have disadvantages such as high equipment cost and vulnerability to soil environment interference.

Digital equipment imaging method can dynamically collect high-resolution *in situ* root images without changing the soil environment and affecting the root growth state, which is conducive to improving the efficiency of root configuration segmentation and quantitative analysis (Hammac et al., 2011). In recent years, it has been widely reported that simple cultivation devices combined with digital equipment (smart phones, scanners, digital cameras) are used to obtain root images (Mohamed et al., 2017; Nakahata and Osawa, 2017; Nahar and Pan, 2019).

On the basis of accurately obtaining high-resolution *in situ* root images, accurate and efficient root configuration recognition is a thorny problem in current root phenotype research (Lynch, 2013). The traditional image processing methods for root recognition include traditional manual description, semi-automatic interactive recognition and automatic threshold segmentation. The manual description method has the problems of low recognition efficiency, large workload and high result error (Abramoff et al., 2004; Le Bot et al., 2010). The semi-automatic method is based on visual observation and image recognition through auxiliary software. Although semi-automatic interaction can achieve high accuracy, it is too dependent on the subjective ability of observers to distinguish roots and their own experience. The segmentation of a single complex root image takes a long time, and the efficiency is too low to achieve high flux *in situ* root image analysis. Although the fully automatic threshold method improves the efficiency of root identification, such as DIRT, GiaRoots, IJ Rhizo and EZ Rhizo can provide statistical information such as root diameter, height and density (Galkovskyi et al., 2012; Pierret et al., 2013; Das et al., 2015). However, it is difficult to eliminate the noise interference of soil background, and there are errors in root morphology identification. And most of the research focuses on obtaining the more extensive root parameters such as structure, length, diameter, etc. It is difficult to excavate more detailed morphological characteristics of micro roots.

Compared with traditional methods, root recognition based on deep learning is easier to mine multi-level characteristics of the target, and occupies a dominant position in the current root phenotype research. For example, the SegRoot platform (Wang et al., 2019) can mine multi-scale features of root images through the improved SegNet network (Badrinarayanan et al., 2017), but under fitting may occur in some cases. The ITErRoot network (Seidenthal et al., 2022) has achieved good results in root segmentation by stacking the encoding, decoding layer and residual structure of the U-shaped structure many times, but its network is too bloated and requires a high training platform. The RootNav2.0 system (Yasrab et al., 2019) is based on the encoder decoder CNN architecture and replaces the previous semi-automatic feature extraction RootNav system (Pound et al., 2013) with a multitask convolutional neural network architecture. It does not require user interaction to accurately extract the root structure, and the speed is increased by nearly 10 times, but it needs to be carried out when the root is fully visible. Through the improved UNet structure, the FaRIA platform (Narisetti et al., 2021) divides the large resolution image into 256 x 256 small images for prediction, and realizes the batch prediction of root images. The RootPainter platform (Smith et al., 2022) includes semi-automatic

and fully automatic methods. The former allows users to subjectively correct each segmented image, and the model can learn from the assigned correction, reducing the segmentation time with the segmentation process; the latter is more suitable for processing large data sets.

Edge devices include raspberry pie series developed by Raspberry Pi Foundation, jetson series developed by Nvidia Company, and orange pie series developed by Xunlong Software Company. Among them, raspberry pie is increasingly used as a low-cost, high-throughput solution for plant phenotype analysis (Jolles, 2021). For example, the "Do It Yourself" phenotyping system (Dobrescu et al., 2017) uses raspberry pie control cameras to achieve batch plant image acquisition, PYM (Valle et al., 2017) uses raspberry pie control infrared cameras to perform phenotype analysis on plant leaves, and Greenotyper (Tausen et al., 2020) uses raspberry pie control cameras and deploys depth learning to monitor plant positions. However, most of these platforms are used for phenotypic analysis of plant parts on the ground, lacking of research cases on plant underground roots. At present, mainstream platforms for root identification, such as RhizoVision Crow (Seethepalli et al., 2020), are based on desktop development and do not support deploying models to Raspberry pie.

It has been reported that the RhizoPot platform was developed by our research team in the early stage (Zhao et al., 2022) can realize high-resolution, non-destructive real-time acquisition of *in-situ* root images. In addition, the team has designed a cotton plant root segmentation method based on DeeplabV3+ and proposed the improvement strategy of the model for the research on root segmentation methods (Shen et al., 2020; Jia et al., 2021). However, the previous studies were all indoor platform development, and the equipment cost was high, lacking the exploration of portable equipment in outdoor environment. Therefore, based on the previous research, this paper designs a data augmentation scheme to expand the data set; DeeplabV3+ model is modified to connect CBAM attention mechanism with ASPP spatial pyramid pooling; The prediction strategy is modified to make it more suitable for edge devices; Deploy to Raspberry pie, and design the method of field experiment. The purpose of this paper is to design a low-cost, high-throughput *in situ* root precise identification technology by replacing the traditional GPU analysis platform with raspberry pie, and explore the possibility of its application in outdoor environment.

2 Materials and methods

2.1 Image collection

This experiment was conducted in the experimental station of Hebei Agricultural University in Baoding, Hebei Province (38.85°N, 115.30°E) in 2021. The climate of the experimental site was mild. Use Epson scanner V39 (Epson Inc., Suwa shi, Nagano, Japan) to scan root images in batches. The resolution of the collected images is set to 1200dpi and the saved format is bmp. The experimental schematic diagram and equipment are shown in Figures 1A, C respectively. Figure 1B shows the prospect of field experiments. Figures 1D–H are the relevant experimental equipment.

This paper filters and classifies the collected image set, removes incomplete and fuzzy images, and finally retains 125 complete and clear cotton roots *in situ* images, randomly selects 100 of them for network training, and the ratio of training set to verification set is 9:1. According to the image data expansion method proposed in this paper, 47873 and 1600 training set images and 5320 and 160 verification set images are finally obtained. The remaining 25 images are used as a test set to evaluate the network performance.

The image annotation is completed by an experienced agronomist using the Adobe Photoshop CC (Adobe Inc., San Jose, CA, United States) lasso tool. All pixels considered as roots are marked white and saved in a new layer. Finally, the remaining pixels are marked black. The resolution of the annotation image is 10200 pixels x 14039 pixels, and the annotation time of a single image is about 4.5 hours.

2.2 Data augmentation

The dataset format required for training DeeplabV3+ is jpg, and the image data set needs to be converted from bmp to jpg. In this paper, two image data augmentation methods are designed. In method 1, the training images are divided according to the resolution of 512 pixels x 512 pixels to ensure that the training can be carried out by resolution. At this time, the input image data set is expanded to 53193, and the training set and verification set are 47873 and 5320, respectively.

In method 2, the training input image is segmented according to the size ratio. To ensure accurate prediction of the original image, the input resolution of the whole image is set to 2048 pixels x 2048 pixels. During training, the input resolution needs to be kept at 512 pixels x 512 pixels. The ratio of training resolution to prediction resolution is 1: 16. Therefore, the image is reduced to 1/16 of the original one to ensure that the prediction of the whole image has a fixed size ratio. The annotation image also needs to go through the same processing process, and the final training set based on equal proportion is 1600 pieces, and the verification set is 160 pieces. After training, the w1 weight based on the equal proportion method and the w2 weight based on the pixel by pixel method are obtained Figure 2 shows the differences between the two methods.

2.3 Segmentation model

2.3.1 Model comparison

The root data set used in this paper is selected in turn to compare DeeplabV3+(Chen et al., 2018), PSPNet (Zhao et al., 2017), HRNet (Sun et al., 2019) and UNet (Ronneberger et al., 2015). The image segmentation results are shown in Figure 3 and the experimental data are shown in Research 1 of Table 1.

From the segmentation effect of root image (Figure 3), it can be seen that DeeplabV3+(MobilenetV2) and UNet have the best segmentation effect, while DeeplabV3+(Xception) has obvious under fitting phenomenon. The segmentation effect of HRNet

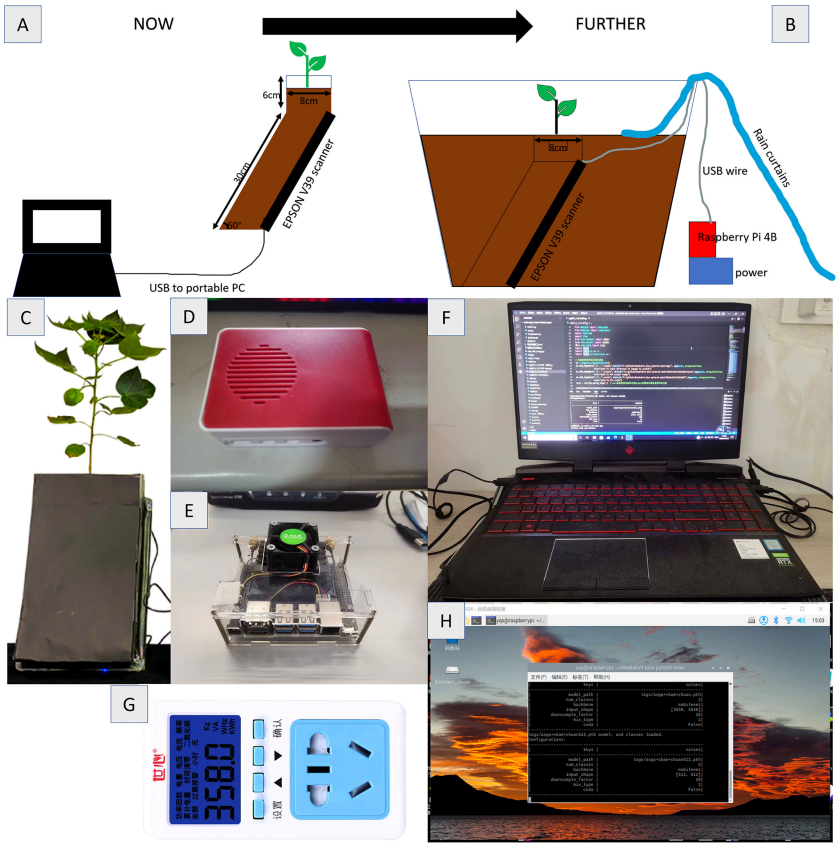


FIGURE 1
Root collection equipment and method **(A)** Schematic diagram of RhizPot **(B)** Schematic diagram of field test **(C)** RhizoPot **(D)** Raspberry Pie 4B **(E)** Jetson Nano **(F)** RTX2060 Notebook **(G)** Power Detector **(H)** Raspberry Pie Remote Desktop.

and PSPNet is very poor, and the model is not suitable for root segmentation.

According to the prediction results of root image (Research 1 of Table 1), DeeplabV3+(MobilenetV2) has the best effect, followed by UNet, DeeplabV3+(Xception), HRNet and PSPNet.

Therefore, based on the previous experimental results, this paper designs a backbone network based on DeeplabV3+ model and MobilenetV2 to train and predict the root image.

2.3.2 Model improvement

At present, the attention mechanism can significantly improve the model feature extraction ability and can be embedded in most mainstream networks without significantly increasing the model parameters and computation. Attention module includes channel attention module, space attention module, time attention module and branch attention module, and mixed attention mechanism: channel space attention mechanism and space time attention mechanism.

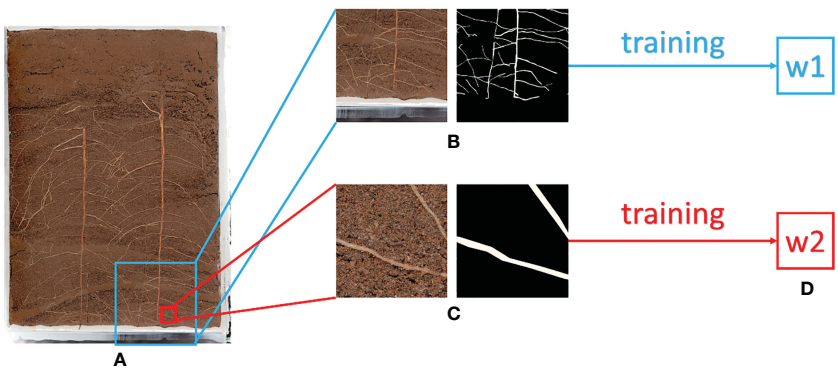


FIGURE 2
Illustration of Equal Proportion Dataset and Pixel by Pixel Dataset **(A)** Original Image **(B)** Equal Proportion Dataset **(C)** Pixel by Pixel Dataset **(D)** After Training, w1 is the weight trained by the Equal Proportion Dataset, and w2 is the weight trained by the Pixel by Pixel Dataset.

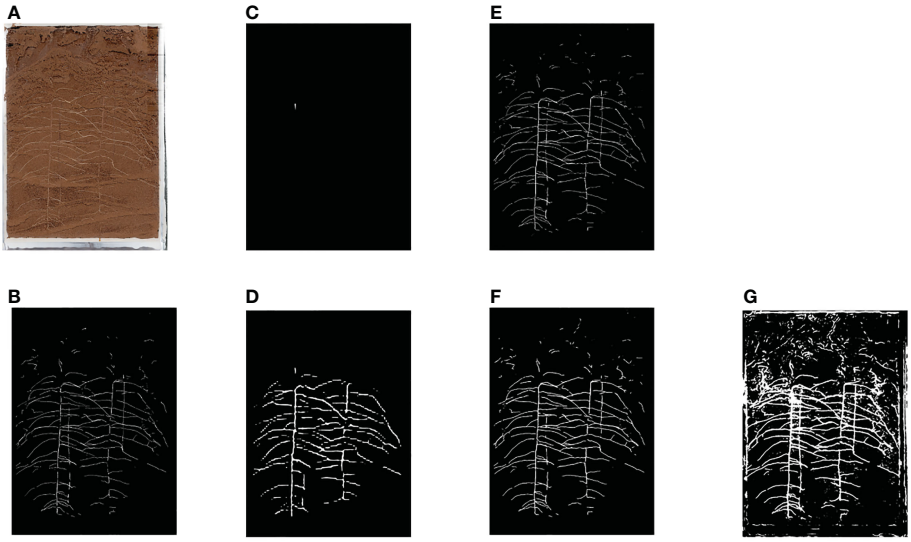


FIGURE 3
Segmentation Results of Various Network Roots **(A)** Original Image **(B)** Ground Truth **(C)** PSPNet **(D)** HRNet **(E)** UNet **(F)** DeeplabV3+(MobilenetV2) **(G)** DeeplabV3+(Xception).

CBAM (Woo et al., 2018) is an attention module in the channel spatial attention mechanism. That is, the image first goes through the channel attention mechanism (CAM) to solve the problem of "what to pay attention to", then goes through the spatial attention mechanism (SAM) to solve the problem of "where to pay attention to", and finally integrates with the original feature map to form a new feature map that emphasizes the channel and spatial features. In addition, CBAM is a lightweight attention mechanism that can be seamlessly integrated into any neural network without module overhead.

The DeeplabV3+ model designed in this paper uses the ASPP structure in the encoder part, which contains three parallel hole convolutions with expansion rates of 6, 12 and 18, which can provide a larger receiving field and capture more context information. On this basis, inspired by the deployment of the dual attention mechanism to the DeeplabV3+ network (Liu et al., 2020a). This paper tests two methods of CBAM attention mechanism deployment to the DeeplabV3 + network, namely, the series connection and parallel connection of CBAM and ASPP. The network structures are shown in Figures 4A, B respectively. Both methods use the pre training weight of the backbone network to iterate for 100 times before performance testing. Figure 5 compares the segmented images of the two methods, and the performance comparison is shown in Research 2 of Table 1. The results show that the CBAM attention mechanism in series with ASPP is better than the parallel operation.

TABLE 1 Research 1: Performance of each network partition.

Research	Network name	mIoU (%)	mPA (%)	mPrecision (%)	Precision root(%)	Precision background(%)	Recall (%)	GPU mtime (min)	Raspberry Pie mtime(min)
Research 1	PSPNet	51.95	63.8	53.93	8.43	99.43	63.8	NA	NA
	HRNet	60.23	67.94	69.09	40.12	98.06	67.94	NA	NA
	DeeplabV3+ (Xception)	67.18	70.25	91.26	86.15	96.36	70.25	NA	NA
								NA	NA
	UNet	**85.06**	**90.8**	91.83	84.2	**99.46**	**90.8**	NA	NA
	DeeplabV3+ (MobilenetV2)	75.17	78.22	**92.9**	**87.75**	98.05	78.22	NA	NA
								NA	NA
Research 2	Series	74.53	77.49	**93.01**	**88.09**	97.92	77.49	NA	NA
	Parallel	**75.77**	**79.16**	92.25	86.28	**98.23**	**79.16**	NA	NA
Research 3	Fast	85.45	**91.96**	91.19	**83.37**	99.55	**91.96**	**0.599**	**26.01**
	Normal	**85.64**	91.94	**91.46**	82.82	99.55	91.94	0.775	41.19

Research 2: Performance comparison of two methods to improve DeeplabV3+. Research 3: Performance of Fast segmentation and Normal segmentation on raspberry pie 4B and GPU platforms. NA, Not Applicable. The optimal values are written in bold font.

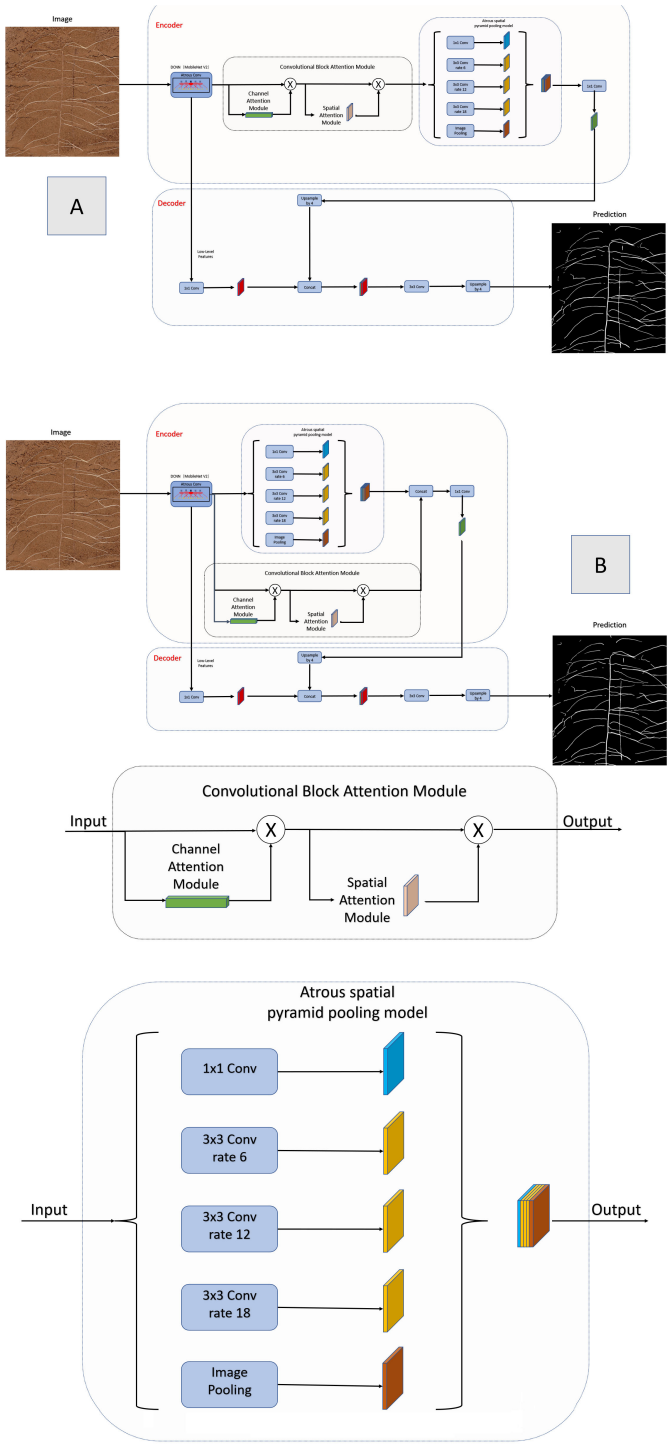


FIGURE 4
Improved DeeplabV3+ network structure **(A)** Series improvement **(B)** Parallel improvement.

2.4 Predictive policy

In the early stage of this paper, the traditional conventional strategy (Normal) of segmentation followed by splicing was tested. By dividing the original image to a specified size, network prediction was performed, and then the prediction results were spliced to obtain a complete segmentation result. The processing process is shown in Figure 6A. The test results show that although the prediction accuracy of this method is high, it takes a long time to predict after deployment to raspberry pie. The shared time of the test set image prediction process is up to 17 hours and 10 minutes. Therefore, this paper proposes an improved fast splicing and segmentation strategy (Fast).

The improved Fast method is shown in Figure 6B. Based on the results of the full image processing of the equal proportion

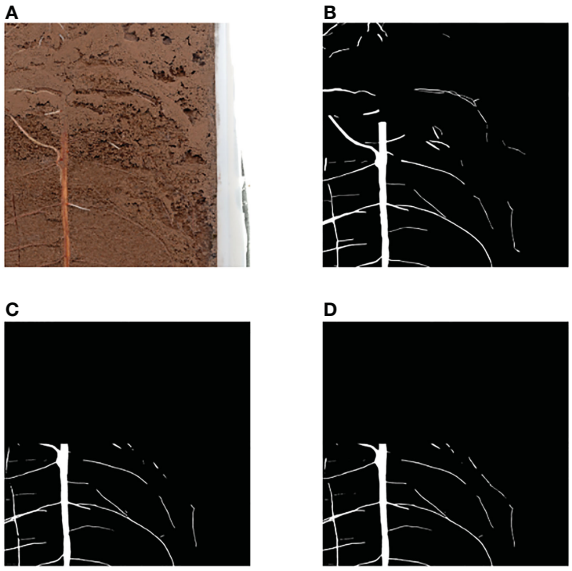


FIGURE 5
Serial and Parallel Segmentation Results **(A)** Original Image **(B)** Ground Truth **(C)** Parallel Segmentation Image **(D)** Parallel Segmentation Image.

segmentation (Figure 7C), the region is divided into foreground and background, where the foreground is the part of the region that contains roots, and the background is the part of the region that does not contain roots. For the foreground part, the region corresponding to the original image is put into the network for segmentation; for the background part, a RGB image with all gray values of 0 is directly used to replace it. Finally, combine the two into a complete segmented image (Figure 7D).

The results show that this method can save 22.71% of the time cost on GPU on average, and the segmentation accuracy only decreases 0.55% year on year, as shown in Research 3 of Table 1. Later deployed in the raspberry pie terminal, the time cost of a single picture can be saved by 36.85%.

2.5 Raspberry pie deployment

The model of the edge device selected in this article is Raspberry Pi Foundation (Cambs, United Kingdom) 4B, which contains 8G of memory, plus 32G of memory card. Raspberry Pie is an ARM based microcomputer motherboard. SD/MicroSD card is used as the memory hard disk. There are 1/2/4 USB interfaces and a 10/100 Ethernet interface (Type A has no network interface) around the card motherboard. It can connect the keyboard, mouse and network cable. It also has a TV output interface for video analog signals and an HDMI high-definition video output interface.

2.5.1 Installation of raspberry pie system

The raspberry pie system selects the raspberry pie official 64 bit system image (Raspberry Pi), sets SSH, WIFI, language and time zone through the official burning software, and then burns it into the 32G memory card. After startup, connect to Raspberry Pie *via* MobaXterm to configure corresponding functions.

2.5.2 Raspberry pie deployment batch splitter

Deploy the batch splitting program to Raspberry Pie 4B, and its settable parameters include picture address, cache address and target address. Its operation process is shown in Figure 8A.

2.5.3 Raspberry pie participates in image collection and segmentation

In order to give full play to the low power consumption advantage of Raspberry Pie 4B, this paper combines the image segmentation program with image acquisition, and can set parameters including scanner name, acquisition quantity, interval time, image storage address, cache address, and target address (the default setting of the scanner is dpi=1200, and the color mode is color). The operation process is shown in Figure 8B. Compared with batch segmentation, continuous collection and segmentation can better reflect the advantages of raspberry pie 4B. The method of timing acquisition is shown in Figure 8C.

The collection results are shown in Figure 9. The collection interval is 24 hours and the collection time is 19:00 every day. In order to show the growth process of root more clearly, the original image is cropped, and the change trend of root can be clearly observed in the image after network segmentation.

3 Results

3.1 Model evaluation

3.1.1 Model selection

This paper compares the performance of PSPNet, HRNet, UNet and DeeplabV3+ depth learning models in cotton root image segmentation. The test results show the comprehensive performance of DeeplabV3+ > UNet > HRNet > PSPNet.

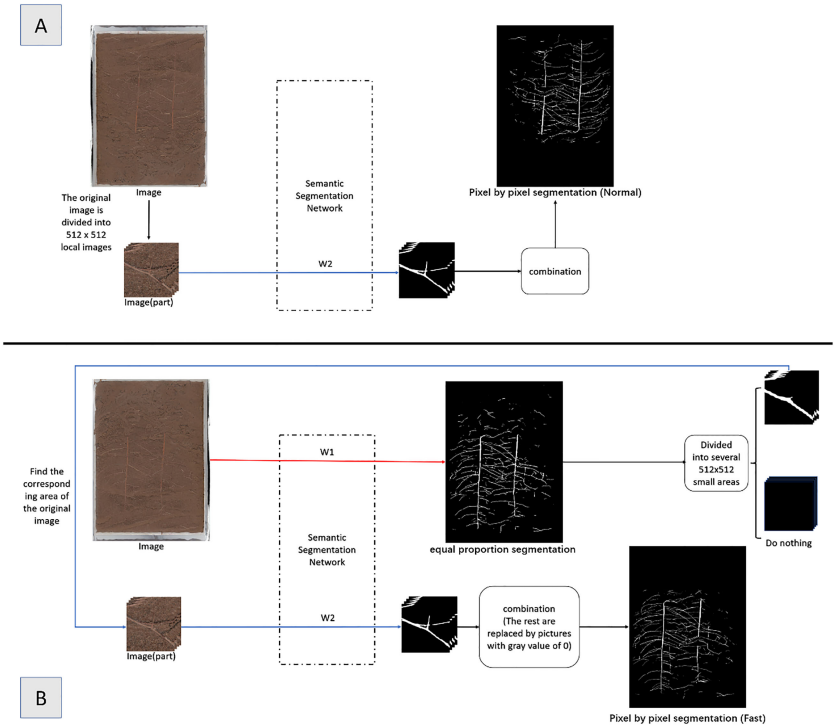


FIGURE 6
Division Method Diagram **(A)** Normal Division Method **(B)** Fast Division Method.

In DeeplabV3+, compare the backbone networks MobilenetV2 and Xception, and see Research 1 of Table 1 for performance comparison when the backbone network pre training weight is used to train iterations for 100 times. The effect of MobilenetV2 is better than that of Xoption because MobilenetV2 is lighter, the network model is smaller, the number of iterations required is less, and the prediction time is shorter. The Xception network model is larger, which is not conducive to deployment to edge devices.

3.1.2 Improvement results

This paper compares CBAM with ASPP in parallel and in series. Figure 4 for network structure and Research 2 of Table 1 for segmentation performance. The results show that in terms of segmentation accuracy, the method of CBAM in series with ASPP is the best, DeeplabV3+ is the second, and the method of CBAM in parallel with ASPP is the second.

In addition, the prediction accuracy of the improved DeeplabV3 + network is 90.15%, higher than DeeplabV3+, HRNet and PSPNet, and slightly lower than Unet. However, because the input and output sizes of the Unet network are inconsistent, resulting in burrs on the edge of the output prediction image. At the same time, the cost of the Unet network is slightly higher. If the Unet network is directly deployed to the Raspberry Pie 4B, it will not be able to segment the high-resolution root image due to memory limitations. Therefore, this paper does not use Unet as the root segmentation network.

The test results show that the improved DeeplabV3+ network can be directly deployed to the raspberry pie 4B, without pruning, and runs well with stability and reliability.

3.2 Performance evaluation

3.2.1 Predictive performance

After the network is deployed to Raspberry Pie, the average partition time of the test set is about 41.19 minutes with the Normal strategy, while the average partition time of the test set is about 26.01 minutes with the Fast strategy, a year-on-year decrease of 15.18 minutes. It takes about 9 minutes to segment images with sparse roots and 33 minutes to segment images with dense roots. Compare the segmented image with the labeled image, and the performance indicators are shown in Research 3 of Table 1.

Therefore, compared with the Normal strategy, the Fast strategy runs 36.85% faster in raspberry pie 4B on average, but the prediction accuracy is only 0.34% lower. The experimental results show that the Fast segmentation strategy proposed in this paper can replace the Normal strategy to a certain extent.

3.2.2 Prediction accuracy

In addition, this paper also uses the open-source Rhizo Vision Explorers platform (Seethepalli et al., 2021) to analyze the phenotypic data of segmented root images, mainly comparing the differences between the result images and the labeled images in root length and diameter. See Table 2 for the comparison results. The results show that the error of root phenotypic parameters such as root length and diameter obtained by Normal strategy and Fast strategy compared with the original labeled image is acceptable. However, there is a big error between the root length, diameter and the actual value of the result image obtained by the equal proportion segmentation method.

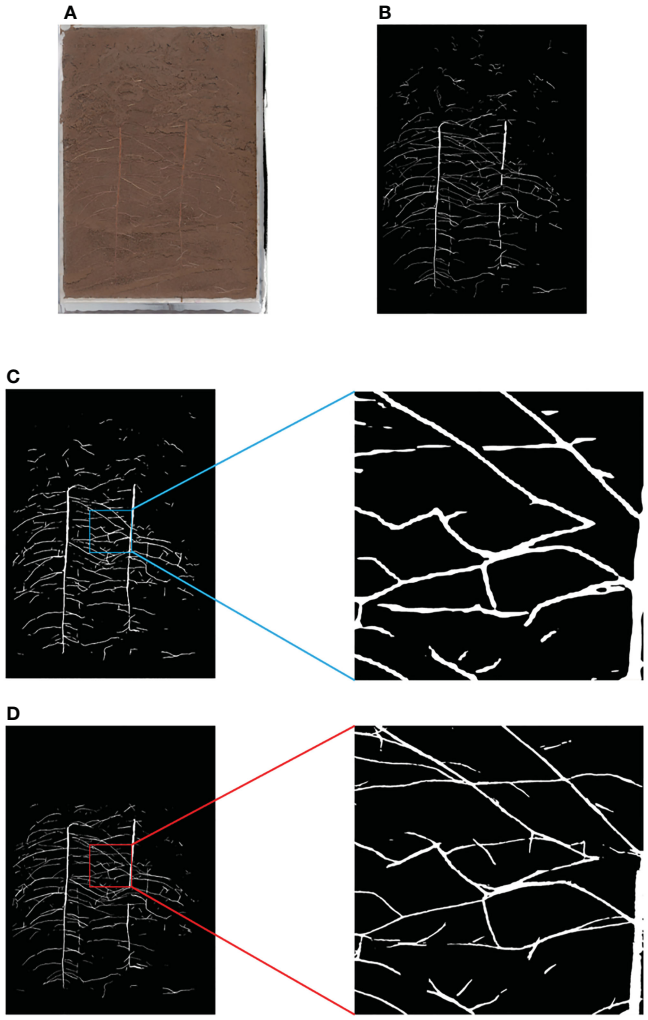


FIGURE 7
Comparison of Equal Proportion Image Segmentation and Pixel by Pixel Image Segmentation **(A)** Original Image **(B)** Ground Truth **(C)** Results and Details of Equal Proportion Segmentation **(D)** Results and Details of Pixel by Pixel Segmentation.

3.3 GPU verification

The advantages of edge devices replacing host devices lie in high portability and low power consumption. In this paper, the Raspberry pie 4B with 8G memory and HP Shadow Genie 5 (i7-9750h+RTX2060+16G memory) notebook computers are used, and the power supply is connected through the P06S-10 power detector to compare the power consumption of the two in the continuous 8-hour image acquisition and segmentation. The acquisition rate is 1 piece per hour, 8 pieces in total are collected, and the image is segmented in the acquisition window period. The experimental equipment is shown in Figure 1, and the results are shown in Table 3.

3.3.1 Power consumption verification

The power consumption of Raspberry Pie 4B is much lower than that of RTX2060 platform when collecting and segmenting the same image. At the same time, when the time interval between two root image scans exceeds 30 minutes, Raspberry pie 4B has the ability to segment within the scanning interval. The time used will not increase, but the power consumption will be greatly reduced. Therefore, it is proved that if the scanning interval is allowed, the edge devices can deploy root segmentation networks to completely replace the high cost and high energy consumption GPU analysis platform.

3.3.2 Time verification

In this paper, the Fast policy and the Normal policy are deployed on the GPU platform for comparison. The results show that the Fast policy is 22.71% faster than the ordinary policy on average, which verifies the conclusion that the Fast policy takes less time than the Normal policy.

3.3.3 Result validation

In this paper, we also carried out a comparative experiment of raspberry pie and GPU batch segmentation of root images

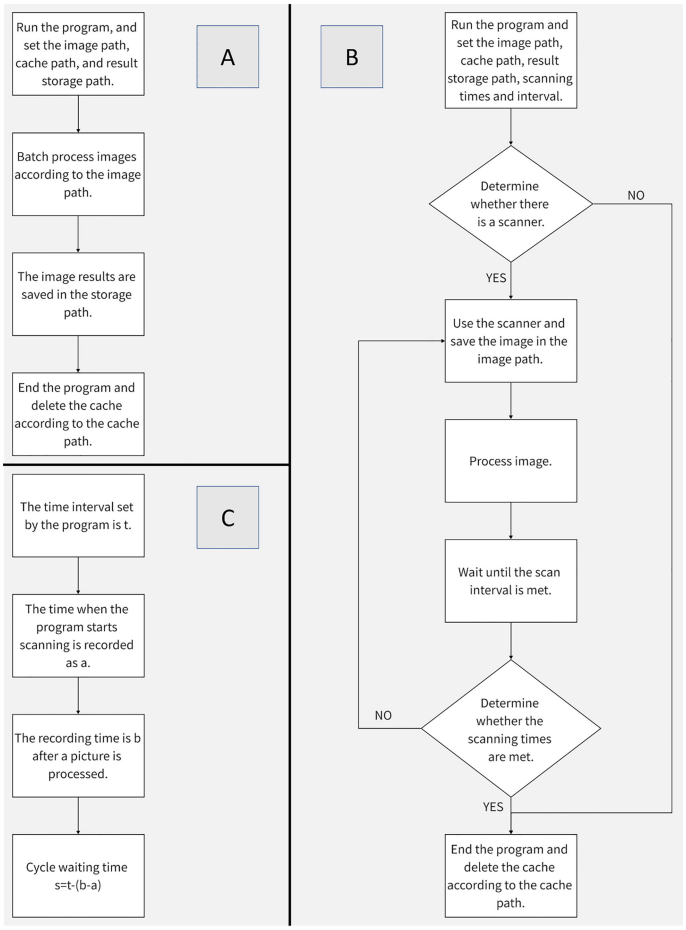


FIGURE 8
Procedure Flow Chart **(A)** Batch Segmentation Procedure Flow **(B)** Image Acquisition and Segmentation Procedure Flow **(C)** Method of Image Acquisition Interval.

simultaneously. Both of them used 25 test set images for continuous segmentation test, with the same weight. The results are shown in Table 3, which verify that the raspberry pie is completely consistent with the GPU in terms of segmentation accuracy. However, due to the limitation of 4B computing power of Raspberry pie, the total power consumption is slightly higher than that of the graphics card.

3.4 Cost evaluation

Theoretically, semantic segmentation on the Jetson Nano with GPU in this experiment will accelerate, but since the memory of the Jetson Nano is only 4G, virtual memory needs to be added for network deployment. Table 4 records the reference price of the equipment used in the experiment. Based on the data provided by

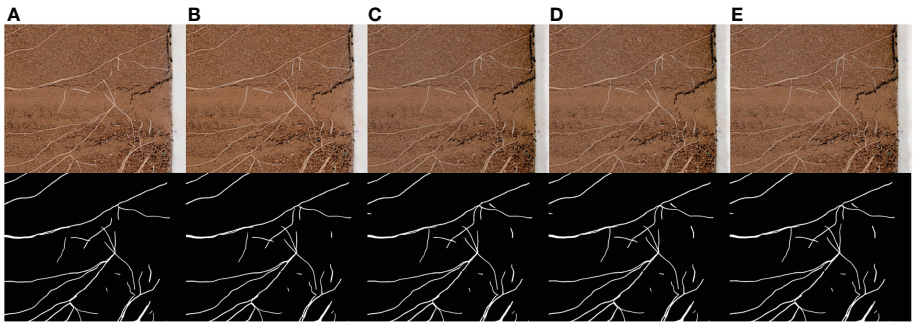


FIGURE 9
Continuous acquisition and segmentation results. **(A–E)** are the root image and network segmentation result collected continuously at a time interval of 24 hours.

TABLE 2 Comparison of equal proportion segmentation, pixel by pixel segmentation and ground truth root system.

Name	Total root length (pt)	Root length error (%)	Total root diameter (pt)	Root diameter error (%)
Ground Truth	5576876	NA	534.835	NA
Equal Proportion	6057449	8.617	837.174	56.529
Pixel by Pixel (Normal)	5659371	1.479	538.775	**0.737**
Pixel by Pixel (Fast)	5614186	**0.669**	540.198	1.003

NA, Not Applicable. The optimal values are written in bold font.

the RhizoPot platform (Zhao et al., 2022), the cost of using the RTX2060 laptop with the Tensor core to control the RhizoPot platform is approximately $1480. Compared with using the Jetson Nano to control the RhizoPot platform, the cost is reduced to $301, but the cost of using the raspberry pie 4B is only $247 (excluding the power detector). Considering the cost, the performance price ratio of raspberry pie 4B or Jetson Nano is much higher than others. Compared with the segmentation performance of raspberry pie 4B and Jetson Nano, finally, this paper selects raspberry pie 4B as the edge device for deploying root image acquisition and segmentation.

4 Discussion

4.1 Basis for model improvement

The previous results show that both CBAM and ASPP can improve the cotton image segmentation accuracy in series or in parallel, but the series method is better than the parallel method. The author believes that:

First, data processing samples are differentiated. The references are for remote sensing image data sets. This paper uses cotton root data sets. The characteristics of the two images are different. The remote sensing image is characterized by the buildings, farmland and other objects collected are basically square, while the cotton root image is irregular, similar to human blood vessels. In addition, remote sensing image segmentation usually faces multi category problems, and cotton root segmentation mainly focuses on two category problems.

Secondly, the series and parallel extraction features are differentiated. The advantage of concatenation is that after CBAM extracts multi-channel attention mechanism features, ASPP is used to sample multi-scale convolution kernel, which can more effectively extract root feature vectors in space and time. In the parallel connection method, on the premise that MobilenetV2 is used as the backbone network, the number of channels output by CBAM is 320, and the number of channels output by ASPP is 256, that is, the number of channels entering feature fusion is 576. However, in order to keep consistent with the number of channels in the network decoding part, the number of channels output by feature fusion can only be 256. Therefore, the rise of input feature fusion dimension leads to higher difficulty of classification, thus reducing the accuracy of the network.

Finally, for the optimization of the tandem method, residual structure, convolution block and pooling layer can be introduced into the attention mechanism module in the later stage to mine the characteristics of attention mechanism at multiple scales.

4.2 Forecast strategy validation

At the beginning of this experiment, the image is similar to pixel merging processing, that is, using the super pixel method (Ren et al., 2019), SLIC super pixel segmentation algorithm is introduced in the prediction, and the input image is divided into a super pixel image. The traditional DeeplabV3+ output operation of these super pixel image regions is used to obtain an accurate segmentation image. However, in actual processing, the SLIC is used to block the image, and the operation of block by block super pixel area will greatly increase the prediction time. At the same time, when the prediction image is more complex and fine, the quality of the output image will be seriously affected. At the same time, the high resolution image also limits the strategy of semantic segmentation.

There are two common methods for semantic segmentation of high-resolution images. The first method is to down sample the image and put it into the network for prediction, and then up sample the results, so that the image processing speed is fast and the context information will not be lost. However, the results of root phenotype analysis showed that the root length and diameter predicted by this method were larger than expected. The second is to use the sliding window operation to divide the image into the same area with a specified size and about 20% reserved, input the network prediction to obtain local results, and then complete the

TABLE 3 Comparison between Raspberry Pie 4B and RTX2060 Notebook for Collection and Batch Segmentation.

Method	Platform	Average split time (h)	Collection time (h)	Total time (h)	Power consumption (kWh)
Collection Segmentation	RTX2060	NA	8	8.008	1.281
	Raspberry Pie 4B	NA	8	8.496	0.051
Batch Segmentation	RTX2060	0.017	NA	0.433	0.073
	Raspberry Pie 4B	0.434	NA	10.338	0.086

NA, Not Applicable.

TABLE 4 Reference Price of Equipment Used in the Experiment.

Name	Type	Price	Other
Raspberry Pi	Raspberry Pi 4B 8G	$35	
16G TransFLash card	Kingston	$5	
RhizoPot platform		$207	Including scanner, USB cable, acrylic plate and glass sealant
Jetson Nano		$89	
RTX2060 notebook	RTX2060 + 16G Memory	$1,268	
Power detector	Worldliness	$4	
Total		$1,608	

stitching of image feature points. This method has a good degree of detail and retains the context information, but it has high time cost and is not friendly to edge devices.

The fast splicing and segmentation strategy (Fast) proposed in this paper lacks context information, so the prediction effect of scattering is poor. The equal proportion segmentation part of this method actually belongs to the first kind of common segmentation method, which contains all the context information of the image. However, as shown in Figure 7, the image processed by this method lacks details, and the root length and diameter errors are too large after root phenotype analysis. Therefore, this method only uses it as a pedal to save the computing time of edge devices.

4.3 Edge device comparison

Compare the performance of Jetson Nano and raspberry pie 4B with similar prices in *in situ* root segmentation. See Table 5 for the results. In the experiment, 25 cotton root images (from the test set) were selected for continuous and equal proportion segmentation, and the network weights used by Jetson Nano and Raspberry Pie 4B were consistent. Due to memory limitations, the Jetson Nano cannot segment 1200dpi images with a resolution of 10200 pixels x 14039 pixels. It can only compare 300dpi root images. In terms of program startup, when using cuda, the Jetson Nano can only segment up to three images consecutively, and then it will report an error that the timer has timed out. When cuda is not used, the segmentation time of the Jetson Nano for 300dpi images is much longer than that of the raspberry pie 4B.

There are three reasons why the split performance of the Jetson Nano is inferior to that of the raspberry pie 4B. First, because of the memory problem, more virtual memory needs to be configured when running programs, but the speed of virtual memory is far lower than that of running memory. The second is the CUDA problem. When running the program, the timer will timeout. The third problem is the processor. The processor model used by the Jetson Nano is Cortex-A57, which lags behind the Cortex-A72 processor of the Raspberry pie 4B. Therefore, in actual use, the performance of raspberry pie 4B is better than that of the Jetson Nano. Theoretically, the performance of the Jetson TX2 is the best (Süzen et al., 2020), but its cost is high, so it is not considered in this paper.

4.4 Outdoor deployment prospect

Because soil color is dark and soil contains more impurities in the case of natural cultivation of plants, this paper tested the image segmentation of deep soil color and obscure root, as shown in Figure 10, the network designed in this paper can still be segmented. The results show that the root segmentation network designed in this paper can carry out accurate identification of *in situ* roots in various situations, and because of the portability and mobility of raspberry pie, it can be deployed outdoors in the field for experiments. The schematic diagram of field experiments is

TABLE 5 Comparison between raspberry pie 4B and jetson nano (the processing method used for comparison is equal proportion segmentation, 25 sheets).

Name	Split time (300DPI)(ms)	Split time (600DPI)(ms)	Split time (1200DPI) (ms)	Power consumption (300DPI)(kWh)	Power consumption (600DPI)(kWh)	Power consumption (1200DPI)(kWh)
Raspberry Pi 4B	1747842ms	1834684ms	2060141ms	0.004	0.004	0.005
Jetson nano (CUDA)	CUDA error	CUDA error	Memory error	NA	NA	NA
Jetson nano (no CUDA)	17960733ms	18582792ms	Memory error	0.021	0.031	NA

NA, Not Applicable.

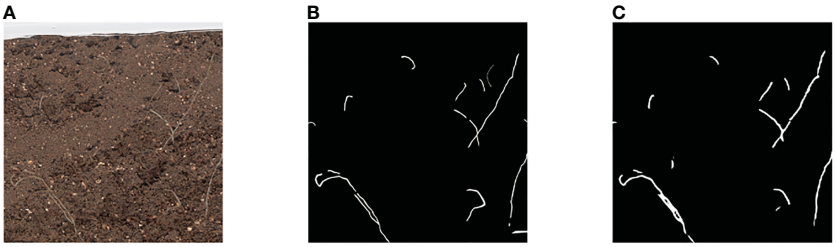


FIGURE 10
Segmentation result of root image with darker soil color **(A)** Original image **(B)** Ground Truth **(C)** Segmentation result.

shown in Figure 1B. At the same time, raspberry pie can be used as a continuous working platform for image segmentation. Compared with GPU as the control, the Raspberry pie has lower power consumption, saves GPU, and enables GPU to perform more computationally demanding tasks.

In the future, this paper will consider adding Nvidia neural computing stick to raspberry pie, which can theoretically improve the segmentation speed of raspberry pie. For outdoor experiments, considering the limited storage capacity of the SD card used by Raspberry Pie, consider adding cloud storage in the future, and upload the collected images obtained by the Raspberry Pie control scanner and the identification images processed by the root segmentation network to the cloud synchronously. This can not only save the limited storage space of Raspberry Pie, but also download scanning and processing images directly from the cloud, replacing the transmission of removable storage devices, improving the work efficiency, and laying a foundation for the development of high-throughput outdoor root phenotype research.

At present, image labeling and training are classified into two categories: root and non-root. When performing root phenotypic analysis, only the whole picture can be analyzed, and the taproot and lateral root cannot be analyzed separately. In the future, this paper will consider using transfer learning to update the categories of taproot and lateral root on the basis of the current situation, which will have more agronomic significance.

In the process of root segmentation, small particles in the soil will have an impact on the segmentation results, resulting in root breakage in the segmentation results. The future goal of our experimental group is to reconstruct the root system by generating an antagonistic network and analyze the reconstructed root system.

5 Conclusion

This paper proposes a method to deploy semantic segmentation model to edge devices, and improves DeeplabV3+ model to segment image edges better. At the same time, we propose an image segmentation strategy which can save time and has both image details and context information. The improved DeeplabV3+ deployment in Raspberry Pie 4B shows good performance. Compared with the deployment of the split network on the GPU platform, the cost of deployment in Raspberry Pie is as low as $247, and the power consumption of 8-hour acquisition and segmentation is as low as 0.051kWh. Considering the time and cost, the accuracy of the improved model is 91.19%, and the errors of the root length and diameter are 0.669% and 1.003% respectively. The effect is similar to that of the GPU, and it is more portable than computers. It can be deployed outdoors for field experiment analysis. It can be seen that if time permits, edge devices can replace laptops to complete batch collection and segmentation of plant root images. In this paper, based on the edge equipment, the segmentation of root phenotype is effectively explored, which provides a favorable basis for the study of root phenotype from the experimental environment to the field and outdoors.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/jiwd123/imporved_deeplabv3plus.

Author contributions

LZ, LL, and NW initiated and designed the research. QY, HT, LZ, and WZ performed the experiments and collected the data. QY wrote the code and tested the methods. QY and HT analyzed the data and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The study was funded by the Top-notch Talent Plan Program of the Education Department of Hebei Province (BJ2021058), and State Key Laboratory of North China Crop Improvement and Regulation (NCCIR2021ZZ-23).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1122833/full#supplementary-material

References

Abramoff, M. D., Magalhães, P. J., and Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics Int.* 11 (7), 36–42.

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (12), 2481–2495. doi: 10.1109/TPAMI.2016.2644615

Bates, G. H. (1937). A device for the observation of root growth in the soil. *Nature* 139 (3527), 966–967. doi: 10.1038/139966b0

Borisjuk, L., Rolletschek, H., and Neuberger, T. (2012). Surveying the plant's world by magnetic resonance imaging. *Plant J.* 70 (1), 129–146. doi: 10.1111/j.1365-313X.2012.04927.x

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the Computer Vision - ECCV 2018. ECCV 2018. Lecture Notes in Computer Science* V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss Eds. (Cham: Springer), Vol. 11211. doi: 10.1007/978-3-030-01234-2_49

Cseresnyés, I., Kelemen, B., Takács, T., Füzy, A., Kovács, R., Megyeri, M., et al. (2021). Electrical capacitance versus minirhizotron technique: A study of root dynamics in wheat–pea intercrops. *Plants* 10 (10), 1991. doi: 10.3390/plants10101991

Das, A., Schneider, H., Burridge, J., Ascanio, A. K. M., Wojciechowski, T., Topp, C. N., et al. (2015). Digital imaging of root traits (DIRT): A high-throughput computing and collaboration platform for field-based root phenomics. *Plant Methods* 11 (1), 51. doi: 10.1186/s13007-015-0093-3

Dobrescu, A., Scorza, L. C. T., Tsaftaris, S. A., and McCormick, A. J. (2017). A "Do-It-Yourself" phenotyping system: Measuring growth and morphology throughout the diel cycle in rosette shaped plants. *Plant Methods* 13 (1), 95. doi: 10.1186/s13007-017-0247-6

Ferreira, T. R., Pires, L. F., and Reichardt, K. (2022). 4D X-ray computed tomography in soil science: an overview and future perspectives at Mogno/Sirius. *Braz. J. Phys.* 52 (2), 33. doi: 10.1007/s13538-021-01043-x

Galkovskyi, T., Mileyko, Y., Bucksch, A., Moore, B., Symonova, O., Price, C. A., et al. (2012). GiA roots: software for the high throughput analysis of plant root system architecture. *BMC Plant Biol.* 12 (1), 116. doi: 10.1186/1471-2229-12-116

Hammac, W. A., Pan, W. L., Bolton, R. P., and Koenig, R. T. (2011). High resolution imaging to assess oilseed species' root hair responses to soil water stress. *Plant Soil* 339 (1), 125–135. doi: 10.1007/s11104-010-0335-0

Hinsinger, P., Brauman, A., Devau, N., Gérard, F., Jourdan, C., Laclau, J.-P., et al. (2011). Acquisition of phosphorus and other poorly mobile nutrients by roots. where do plant nutrition models fail? *Plant Soil* 348 (1), 29. doi: 10.1007/s11104-011-0903-y

Horn, J., Zhao, Y., Wandel, N., Landl, M., Schnepf, A., and Behnke, S. (2021). "Robust skeletonization for plant root structure reconstruction from MRI," in *2020 25th International Conference on Pattern Recognition (ICPR)*. (Milan, Italy: IEEE), 10689–10696. doi: 10.1109/ICPR48806.2021.9413045

Jahnke, S., Menzel, M. I., Van Dusschoten, D., Roeb, G. W., Bühler, J., Minwuyelet, S., et al. (2009). Combined MRI–PET dissects dynamic changes in plant structures and functions. *Plant J.* 59 (4), 634–644. doi: 10.1111/j.1365-313X.2009.03888.x

Jia, K. A., Llbc, D., Fz, E., Chen, S. A., Nan, W., and Ls, A. (2021). Semantic segmentation model of cotton roots *in-situ* image based on attention mechanism. *Comput. Electron. Agric.* 189, 106370. doi: 10.1016/j.compag.2021.106370

Jolles, J. W. (2021). Broad-scale applications of the raspberry pi: A review and guide for biologists. *Methods Ecol. Evolution* 12 (9), 1562–1579. doi: 10.1111/2041-210X.13652

Le Bot, J., Serra, V., Fabre, J., Draye, X., Adamowicz, S., and Pagès, L. (2010). DART: A software to analyse root system architecture and development from captured images. *Plant Soil* 326 (1), 261–273. doi: 10.1007/s11104-009-0005-2

Liu, X., Gu, H., Han, J., Jiang, H., and Duan, S. (2020b). Research progress of ground penetrating radar and electrical capacitance for in-situ non-destructive measurement of crop roots. *Trans. Chin. Soc. Agric. Eng.* 36 (20), 226–237.

Liu, W., Shu, Y., Tang, X., and Liu, J. (2020a). Remote sensing image segmentation using dual attention mechanism Deeplabv3+ algorithm. *Trop. Geogr.* 40 (2), 303–313. doi: 10.13284/j.cnki.rddl.003229

Lynch, J. P. (2013). Steep, cheap and deep: an ideotype to optimize water and n acquisition by maize root systems. *Ann. Bot.* 112 (2), 347–357. doi: 10.1093/aob/mcs293

Lynch, J. P., and Wojciechowski, T. (2015). Opportunities and challenges in the subsoil: pathways to deeper rooted crops. *J. Exp. Bot.* 66 (8), 2199–2210. doi: 10.1093/jxb/eru508

Metzner, R., Eggert, A., van Dusschoten, D., Pflugfelder, D., Gerth, S., Schurr, U., et al. (2015). Direct comparison of MRI and X-ray CT technologies for 3D imaging of root systems in soil: Potential and challenges for root trait quantification. *Plant Methods* 11 (1), 17. doi: 10.1186/s13007-015-0060-z

Mohamed, A., Monnier, Y., Mao, Z., Lobet, G., Maeght, J.-L., Ramel, M., et al. (2017). An evaluation of inexpensive methods for root image acquisition when using rhizotrons. *Plant Methods* 13 (1), 11. doi: 10.1186/s13007-017-0160-z

Nahar, K., and Pan, W. L. (2019). High resolution *in situ* rhizosphere imaging of root growth dynamics in oilseed castor plant (Ricinus communis l.) using digital scanners. *Modeling Earth Syst. Environ.* 5 (3), 781–792. doi: 10.1007/s40808-018-0564-4

Nakahata, R., and Osawa, A. (2017). Fine root dynamics after soil disturbance evaluated with a root scanner method. *Plant Soil* 419 (1), 467–487. doi: 10.1007/s11104-017-3361-3

Narisetti, N., Henke, M., Seiler, C., Junker, A., Ostermann, J., Altmann, T., et al. (2021). Fully-automated root image analysis (faRIA). *Sci. Rep.* 11 (1), 16047. doi: 10.1038/s41598-021-95480-y

Nielsen, K. L., Eshel, A., and Lynch, J. P. (2001). The effect of phosphorus availability on the carbon economy of contrasting common bean (Phaseolus vulgaris l.) genotypes. *J. Exp. Bot.* 52 (355), 329–339. doi: 10.1093/jexbot/52.355.329

Paez-Garcia, A., Motes, C. M., Scheible, W.-R., Chen, R., Blancaflor, E. B., and Monteros, M. J. (2015). Root traits and phenotyping strategies for plant improvement. *Plants* 4 (2), 334–355. doi: 10.3390/plants4020334

Park, J., Seo, D., and Kim, K. W. (2020). X-Ray computed tomography of severed root wounds of prunus serrulata and zelkova serrata. *For. Pathol.* 50 (4), e12622. doi: 10.1111/efp.12622

Pflugfelder, D., Kochs, J., Koller, R., Jahnke, S., Mohl, C., Pariyar, S., et al. (2021). The root system architecture of wheat establishing in soil is associated with varying elongation rates of seminal roots: Quantification using 4D magnetic resonance imaging. *J. Exp. Bot.* 73 (7), 2050–2060. doi: 10.1093/jxb/erab551

Pierret, A., Gonkhamdee, S., Jourdan, C., and Maeght, J.-L. (2013). IJ_Rhizo: an open-source software to measure scanned images of root samples. *Plant Soil* 373 (1), 531–539. doi: 10.1007/s11104-013-1795-9

Pound, M. P., French, A. P., Atkinson, J. A., Wells, D. M., Bennett, M. J., and Pridmore, T. (2013). RootNav: Navigating images of complex root architectures. *Plant Physiol.* 162 (4), 1802–1814. doi: 10.1104/pp.113.221531

Rajurkar, A. B., McCoy, S. M., Ruhter, J., Mulcrone, J., Freyfogle, L., and Leakey, A. D. B. (2022). Installation and imaging of thousands of minirhizotrons to phenotype root systems of field-grown plants. *Plant Methods* 18 (1), 39. doi: 10.1186/s13007-022-00874-2

Ren, F., He, X., Wei, Z., Lv, Y., and Li, M. (2019). Sematic segmentation based on DeepLabV3+ and superpixel optimization. *Optics Precis. Eng.* 27 (12), 2722–2729. (in Chinese).

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional networks for biomedical image segmentation" in *Lecture Notes in Computer Science* Eds N. Navab, J. Hornegger, W. Wells and A. Frangi (Cham: Springer) 9351, 234–241. doi: 10.1007/978-3-319-24574-4_28

Schneider, H. M., Postma, J. A., Kochs, J., Pflugfelder, D., Lynch, J. P., and van Dusschoten, D. (2020). Spatio-temporal variation in water uptake in seminal and nodal root systems of barley plants grown in soil. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.01247

Scotson, C. P., van Veelen, A., Williams, K. A., Koebernick, N., McKay Fletcher, D., and Roose, T. (2021). Developing a system for *in vivo* imaging of maize roots containing iodinated contrast media in soil using synchrotron XCT and XRF. *Plant Soil* 460 (1), 647–665. doi: 10.1007/s11104-020-04784-x

Seethepalli, A., Dhakal, K., Griffiths, M., Guo, H., Freschet, G. T., and York, L. M. (2021). RhizoVision explorer: Open-source software for root image analysis and measurement standardization. *AoB Plants* 13 (6), plab056. doi: 10.1093/aobpla/plab056

Seethepalli, A., Guo, H., Liu, X., Griffiths, M., Almtarfi, H., Li, Z., et al. (2020). RhizoVision crown: An integrated hardware and software platform for root crown phenotyping. *Plant Phenomics* 2020, 3074916. doi: 10.34133/2020/3074916

Seidenthal, K., Panjvani, K., Chandnani, R., Kochian, L., and Eramian, M. (2022). Iterative image segmentation of plant roots for high-throughput phenotyping. *Sci. Rep.* 12 (1), 16563. doi: 10.1038/s41598-022-19754-9

Shan, J., and Tao, D. (1992). Overseas researches on tree fine root. *Chin. J. Ecol.* 11 (4), 46–49.

Shen, C., Liu, L., Zhu, L., Kang, J., and Shao, L. (2020). High-throughput *in situ* root image segmentation based on the improved DeepLabv3+ method. *Front. Plant Sci.* 11, 576791. doi: 10.3389/fpls.2020.576791

Smith, A. G., Han, E., Petersen, J., Olsen, N. A. F., Giese, C., Athmann, M., et al. (2022). RootPainter: deep learning segmentation of biological images with corrective annotation. *New Phytol.* 236 (2), 774–791. doi: 10.1111/nph.18387

Sun, K., Xiao, B., Liu, D., and Wang, J. (2019). "Deep high-resolution representation learning for human pose estimation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Long Beach, CA, USA: IEEE), 5686–5696. doi: 10.1109/CVPR.2019.00584

Süzen, A. A., Duman, B., and Şen, B. (2020). "Benchmark analysis of jetson TX2, jetson nano and raspberry PI using deep-CNN," in *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. (Ankara, Turkey: IEEE), 1–5. doi: 10.1109/HORA49412.2020.9152915

Tausen, M., Clausen, M., Moeskjær, S., Shihavuddin, A., Dahl, A. B., Janss, L., et al. (2020). Greenotyper: Image-based plant phenotyping using distributed computing and deep learning. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.01181

Valle, B., Simonneau, T., Boulord, R., Sourd, F., Frisson, T., Ryckewaert, M., et al. (2017). PYM: A new, affordable, image-based method using a raspberry pi to phenotype plant leaf area in a wide diversity of environments. *Plant Methods* 13 (1), 98. doi: 10.1186/s13007-017-0248-5

Wang, T., Rostamza, M., Song, Z., Wang, L., McNickle, G., Iyer-Pascuzzi, A. S., et al. (2019). SegRoot: A high throughput segmentation method for root image analysis. *Comput. Electron. Agric.* 162, 845–854. doi: 10.1016/j.compag.2019.05.017

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*. (Cham: Springer), 3–19. doi: 10.1007/978-3-030-01234-2_1

Xiao, S., Liu, L., Zhang, Y., Sun, H., Bai, Z., Zhang, K., et al. (2020). Review on new methods of *in situ* observation of plant micro-roots and interpretation of root images. *J. Plant Nutr. Fertilizers* 26 (2), 370–385. (in Chinese).

Yasrab, R., Atkinson, J. A., Wells, D. M., French, A. P., Pridmore, T. P., and Pound, M. P. (2019). RootNav 2.0: Deep learning for automatic navigation of complex plant root architectures. *Gigascience* 8 (11), giz123. doi: 10.1093/gigascience/giz123

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). "Pyramid scene parsing network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Honolulu, HI, USA: IEEE), 6230–6239. doi: 10.1109/CVPR.2017.660

Zhao, H., Wang, N., Sun, H., Zhu, L., Zhang, K., Zhang, Y., et al. (2022). RhizoPot platform: A high-throughput *in situ* root phenotyping platform with integrated hardware and software. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1004904

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 February 2023
DOI 10.3389/fpls.2023.1073530

OPEN ACCESS

EDITED BY
Mansour Ghorbanpour,
Arak University, Iran

REVIEWED BY
Alireza Sanaeifar,
Zhejiang University, China
Seyed Mehdi Talebi,
Arak University, Iran
Behnam Asgari Lajayer,
University of Tabriz, Iran

*CORRESPONDENCE
Shizhuang Weng
weng_1989@126.com
Linsheng Huang
linsheng0808@163.com

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

RECEIVED 18 October 2022
ACCEPTED 14 February 2023
PUBLISHED 28 February 2023

CITATION
Weng S, Ma J, Tao W, Tan Y, Pan M, Zhang Z, Huang L, Zheng L and Zhao J (2023) Drought stress identification of tomato plant using multi-features of hyperspectral imaging and subsample fusion.
Front. Plant Sci. 14:1073530.
doi: 10.3389/fpls.2023.1073530

Drought stress identification of tomato plant using multi-features of hyperspectral imaging and subsample fusion

Shizhuang Weng*, Junjie Ma, Wentao Tao, Yujian Tan, Meijing Pan, Zixi Zhang, Linsheng Huang*, Ling Zheng and Jinling Zhao

National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, Anhui University, Hefei, China

Drought stress (DS) is one of the most frequently occurring stresses in tomato plants. Detecting tomato plant DS is vital for optimizing irrigation and improving fruit quality. In this study, a DS identification method using the multi-features of hyperspectral imaging (HSI) and subsample fusion was proposed. First, the HSI images were measured under imaging condition with supplemental blue lights, and the reflectance spectra were extracted from the HSI images of young and mature leaves at different DS levels (well-watered, reduced-watered, and deficient-watered treatment). The effective wavelengths (EWs) were screened by the genetic algorithm. Second, the reference image was determined by ReliefF, and the first four reflectance images of EWs that are weakly correlated with the reference image and mutually irrelevant were obtained using Pearson's correlation analysis. The reflectance image set (RIS) was determined by evaluating the superposition effect of reflectance images on identification. The spectra of EWs and the image features extracted from the RIS by LeNet-5 were adopted to construct DS identification models based on support vector machine (SVM), random forest, and dense convolutional network. Third, the subsample fusion integrating the spectra and image features of young and mature leaves was used to improve the identification further. The results showed that supplemental blue lights can effectively remove the high-frequency noise and obtain high-quality HSI images. The positive effect of the combination of spectra of EWs and image features for DS identification proved that RIS contains feature information pointing to DS. Global optimal classification performance was achieved by SVM and subsample fusion, with a classification accuracy of 95.90% and 95.78% for calibration and prediction sets, respectively. Overall, the proposed method can provide an accurate and reliable analysis for tomato plant DS and is hoped to be applied to other crop stresses

KEYWORDS

hyperspectral imaging, drought stress, tomato, multi-features, subsample fusion

1 Introduction

Tomato (*Solanum lycopersicum* L.) is a popular and important vegetable crop cultivated in more than 100 countries and regions (Xia et al., 2021). The tomato fruit possesses antiaging and cancer-preventing effects and is also valuable for human health as it contains natural antioxidants, such as lycopene, carotene, and vitamins, as well as organic acids, such as malic acid and citric acid (Eid et al., 2020; Mukhtar et al., 2020). During cultivation, tomato is inevitably subject to many biotic and abiotic stresses. Drought stress (DS) is the main factor affecting its growth and development. The rational utilization of water resources is one of the topics of global universal concern. How to identify DS degree accurately and optimize irrigation reasonably must be explored for the sustainable development of agriculture. Different molecular, biochemical, physiological, morphological, and ecological traits of plants are impaired under DS conditions (Seleiman et al., 2021), resulting in wilted leaves, small stem diameter, and reduced photosynthetic efficiency. Furthermore, DS can affect the concentration of nutrients, such as sugars, acids, and proteins, in tomato fruit and lead to a decline in yield and quality (Chen et al., 2014; Hao et al., 2019). The detection of tomato plant DS can assist in providing timely irrigation, ensure normal plant growth, improve fruit quality, and reduce economic loss (Moharana and Dutta, 2019).

Plants rely on leaves for photosynthesis and respiration to provide energy for themselves and exchange gases with the outside world (Flexas et al., 2006; Haworth et al., 2018; Rad et al., 2022). The leaf effectively summarizes the stress-driven perturbations of the plant's physiological status (Melandri et al., 2021). Therefore, plant DS is usually characterized by the appearance, temperature, and optical properties of leaves. Visual analysis, canopy temperature, thermal imaging, machine vision, and spectroscopic techniques are commonly used to analyze the DS degree of plants. Visual analysis, which relies on professional and experienced inspectors, is convenient and nondestructive but susceptible to subjective interference (Weng et al., 2021). Canopy temperature and thermal imaging can quantify the complex relationship between temperature and stress degree without needing physical contact, but they are affected by the aliasing of plants and soil background information (Ni et al., 2015; Han et al., 2016). The low cost, noncontact, and rapid acquisition of the leaf's external features are the major advantages of machine vision, however, the lack of information about the internal composition and structure of the leaf limits its identification accuracy (Taghizadeh et al., 2011; Pandey et al., 2017). In recent years, spectroscopic techniques, such as near-infrared spectroscopy and reflectance spectroscopy, have been widely used in plant DS assessment because of their simplicity, speediness, and zero reagent consumption, these techniques provide information on the stretching vibrations of hydrogen-containing functional groups, such as C—H, N—H, S—H, and O—H (Steidle Neto et al., 2017; Li P, et al., 2020; Das et al., 2021; Raddi et al., 2022). Nevertheless, the techniques cannot precisely locate the leaf on the designated plant. They may also be influenced by other plants. The lack of spatial information reflecting the color, texture, shape, and position of the leaf also limits the improvement of accuracy.

As the combination of two sensor modalities, namely, imaging and spectroscopy, hyperspectral imaging (HSI) is a promising nondestructive detection method that can provide spatial and internal information, such as the composition and molecular structure of analytes (Mishra et al., 2019; Sun et al., 2021). HSI technology is widely used in analyzing DS in plants. Chen et al. established the machine learning model using HSI to monitor the drought degree of tea seedlings under DS (Chen et al., 2021). Zhou et al. tested the application of hyperspectral reflectance as a high-throughput phenotyping approach for the early identification of DS in citrus trees by conducting a greenhouse experiment (Zhou et al., 2021). HSI has also been used to explore the physiological processes of DS in plants, which is essential for selecting drought-tolerant genotypes and promoting breeding research (Asaari et al., 2019). HSI has an excellent performance in many analyses but may hardly obtain a good recognition effect using only spectral information. For instance, the nondestructive detection of healthy leaves and leaves infected with grapevine leafroll disease based on the spectra from HSI obtained a classification accuracy of 60.74% to 89.93% in the first four phenological stages (Gao et al., 2020).

Recent studies have attempted to combine spectra and image features to gain adequate information and improve the application effects of HSI (Wang et al., 2015; Ru et al., 2019). Compared with the accuracy of identifying yellow rust in wheat leaves using the spectra alone, the accuracy of identifying yellow rust in wheat leaves using spectra and texture features was increased by 7.3%(Guo et al., 2020). Combining spectra, texture features, and morphological features can improve the accuracy by 2% and 1.3% for the germ side and endosperm side, respectively (Yang et al., 2015). The texture and morphological features of images were extracted in previous studies through statistical analysis methods, such as gray level co-occurrence matrix (GLCM) and morphological parameter calculation, whereas these methods are complex, time-consuming, dependent on spatial scale, and subject to prior information (Sachar and Kumar, 2021). In recent years, deep learning has demonstrated its excellent feature extraction ability and has been widely used, especially in the imaging field (Yu et al., 2020).

The leaf state in a single growth stage can hardly represent an accurate expression of plant stress because of limited multidimensional and heterogeneous information. Borraz-Martínez et al. examined young and adult leaves at the spectral level and found differential pieces of information between them (Borraz-Martinez et al., 2019). Subsample fusion, the integration of information from leaves at various growth stages, was explored to determine tomato plant DS.

In addition, illumination conditions considerably influence the image quality in HSI experiments. As a stable and diffuse light source, the halogen lamp is often used as an illumination unit in HSI. Nevertheless, the available light amount of the halogen lamp is low in the visible region, resulting in HSI images with a poor signal-to-noise ratio (SNR). Mahlein et al. indicated that supplemental visible light can alleviate this problem (Mahlein et al., 2015).

The lack of light energy at 400–500 nm may be the main factor leading to high-frequency noise in the visible region. The spatial information may provide some help in accurately analyzing the plant DS. Moreover, the fusion of multiple types of leaf samples may enhance the judgment of plant physiological status. Herein, tomato

plant DS was detected using multi-features of HSI and subsample fusion (Figure 1). In particular, the objectives of the present study were to (1) analyze the impact of blue lights on HSI image quality, (2) combine the spectra of the effective wavelengths (EWs) and image features extracted by LeNet-5 for DS analysis, (3) explore the effect of subsample fusion in DS analysis, and (4) develop the identification models of DS using dense convolutional network (DenseNet).

2 Materials and methods

2.1 Experimental design and irrigation treatments

The experiment was conducted in a greenhouse situated in the Laboratory of National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, Anhui University, China. The red cherry tomato seedlings (50 days old) were purchased from the local market and transplanted into 18 pots (approximately 3 L each and one plant per pot) with ordinary soil on July 22, 2021. The available nitrogen, phosphorus, and potassium levels in the soil were determined by distillation method, spectrophotometric method, and flame photometric method, respectively. The fertilizer conversion ratio required to grow the tomato between the field and pot were calculated. Based on the calculation results, 1.603 g urea, 0.773 g phosphorus pentoxide, and 0.985 g potassium sulfate were added to the soil to promote tomato plant growth. The temperature and the relative humidity in the greenhouse were set at 24°C and 68%, respectively. The initial soil relative humidity (SRH) was maintained at 60%–80% by applying stored rainwater suitably. Appropriate and uniform light was also provided. After the plant roots were fixed, watering treatment was halted to reduce moisture. SRH was monitored daily by a temperature and humidity sensor. The samples of the three water treatments were obtained over time: (a) well-watered treatment: 60%–80% SRH; (b) reduced-watered treatment: 40%–60% SRH; (c) deficient-watered treatment: 20%–40% SRH.

2.2 HSI system and data acquisition

The HSI images of the leaves on tomato plants were collected using the HSI system in the visible/near-infrared range from 400 nm to 1000 nm (Figures 2; S1). The system consists of an indoor measuring platform with an area of 0.45 × 0.45 m^2, a Headwall Nano-Hyperspec (Headwall Photonics Inc., Bolton, MA, USA)

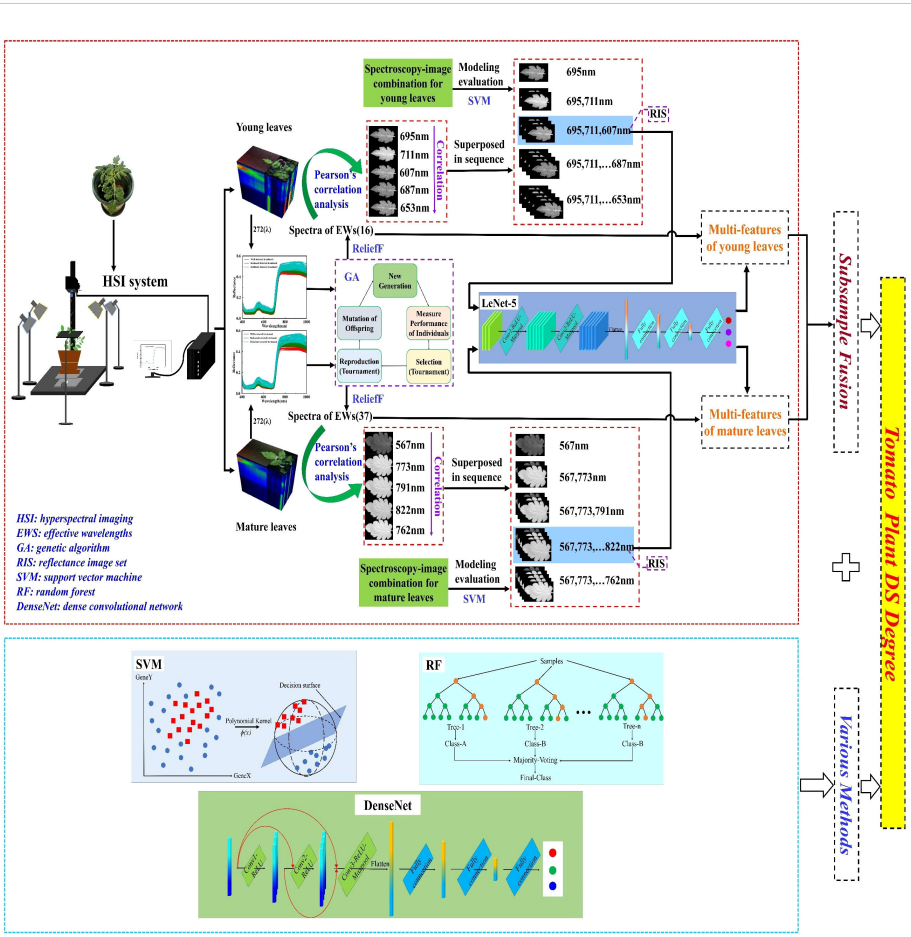


FIGURE 1
Flowchart of the determination of tomato plant DS degree.

10.3389/fpls.2023.1073530

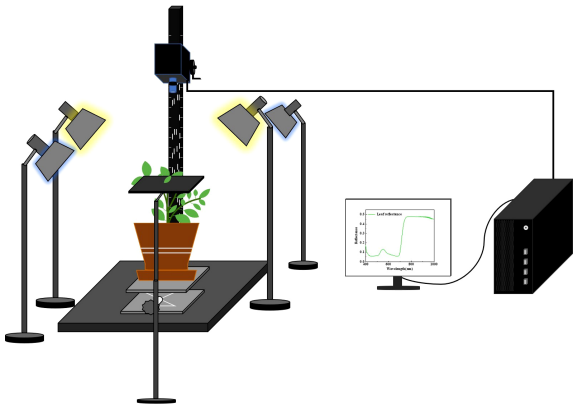


FIGURE 2
HSI system for tomato leaf data acquisition.

push-broom sensor that offers 272 spectral bands and 640 spatial pixels, a lighting unit with two halogen lamps (75 W) as the main radiation and two blue lights (3 W) as the auxiliary radiation, and a computing unit. Two halogen lamps and two blue lamps were placed on both sides and diagonal of the sample to ensure proper and uniform illumination. All lights were preheated for 30 min before imaging to reduce the influence of light intensity changes over time on the experiment (Ma et al., 2020). During data acquisition, the parameters of the HSI system were set as follows: exposure time, 60 ms; frame period, 65 ms; scanning speed, 0.543 deg/s. The distance between the lens and the sample was set at 30 cm by controlling the height of the lifting platform. For the image calibration, white and dark reference images were acquired by scanning a standard white board with 98% reflectance and covering the lens before collecting the HSI images of leaves. The correction formula is as follows:

$$I_c = \frac{I_r - I_d}{I_w - I_d} \quad (1)$$

where I_c is the corrected image, I_r is the measured raw leaf image, and I_w a I_d are the white and dark reference images, respectively.

After the parameters of the HSI spectrometer were set, the first three leaves on the branch of the plant canopy were considered young leaves, and the last three leaves on the branch below the plant's stem were mature leaves. Three young leaves (brightly colored with luster and toughness) and three mature leaves (dull colored without luster) were selected from each plant for HSI measurement. A total of 630 samples were obtained, including 315 images of young leaves and 315 images of mature leaves (Figure 3; Table 1).

2.3 Selection of EWs

After image correction, the region of interest (ROI) of the leaf was obtained by using the threshold segmentation method. The average value for all pixels within the ROI was calculated as the reflectance. The extracted reflectance spectra have high dimension

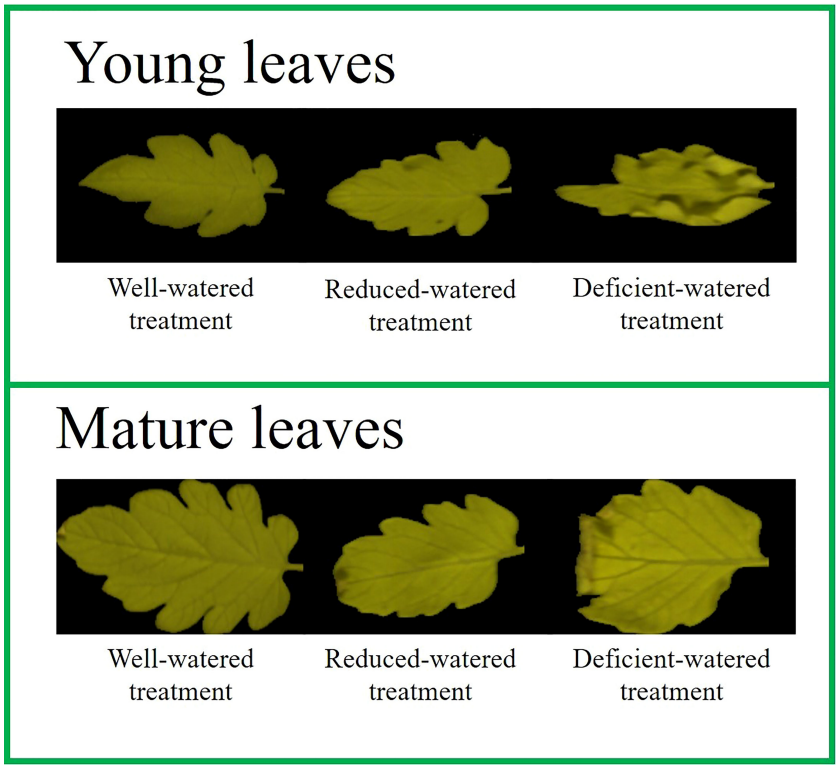


FIGURE 3
Color composite images of tomato leaves under different water treatments.

TABLE 1 Sample distribution under different water treatments.

Datasets	Number of tomato leaf images		
	Well-watered	Reduced-watered	Deficient-watered
Young leaves	108	105	102
Mature leaves	108	105	102

and multicollinearity (Ng et al., 2019). A small number of variables can reduce the influence of noncorrelated variables, raise computational efficiency, and improve model performance. The genetic algorithm (GA), a variable selection method, was used to select EWs. The support vector machine (SVM) classifier was used as an evaluator. The algorithm parameters, such as initialized population, number of iterations, crossover probability, and variance probability, were set to 100, 400, 0.5, and 0.1, respectively. The fivefold crossover validation was employed to seek the global optimal solution in the descendants, and the mean value of cross-validation accuracy was used as the fitness function in this study. The EWs were determined by the following steps: (1) After the first application of GA, an accuracy value of the test set was obtained and used as the reference value. (2) GA was executed in a loop, and the number of the loop was set to 1000. If the run result was greater than the reference value, the result was designated as the new reference value. The loop was exited, and step (2) was repeated. (3) If GA was executed 1000 times continuously without obtaining a good result, the previously acquired feature subset would be considered the EWs. Table S1 shows the selected EWs.

GA is an adaptive global probability search and optimization algorithm (Song et al., 2021) that utilizes selection, exchange, and mutation operations to retain the variables with high objective function values and delete the variables with low objective function values by continuous genetic iterations based on the biological evolution mechanism in nature. Thus, the optimal combination of variables was obtained.

2.4 Image features

The reflectance image of EW with the highest weight value given by ReliefF (Key et al., 2022) was regarded as the reference image. The reference image contains specific and crucial information. Other reflectance images with a weak correlation with the reference image can provide complementary information. Therefore, Pearson's correlation analysis (Yang B, et al., 2021) was used to calculate the correlation between the reflectance images of EWs. Moreover, a threshold ranging from −0.3 to +0.3 was set. The reference image was placed in a defined container. Its correlation with the reflectance images of the remaining EWs (indexed sequentially) was analyzed until a reflectance image that met the threshold condition was found and added to the container. The correlation between the reflectance images of residual EWs and the reflectance images in the container was further calculated. A set of mutually unrelated reflectance images were obtained by threshold filtering. Then, the first four reflectance images with a weak correlation with the reference image were selected. Five reflectance images were superposed one by one according to the correlation ranking. The reflectance image set (RIS) was determined by analyzing the superposition effect of reflectance images. The RIS demonstrates high specificity, sensitivity, and convenience for further processing. Finally, the image features were extracted from RIS by LeNet-5. The relevant parameters are shown in Table S2.

Convolutional neural networks (CNNs), which use local connection and weight sharing to reduce the training parameters and computational complexity, can extract useful features quickly and accurately (Yang W, et al., 2021). CNNs are composed of convolution, pooling, and full connection layers. The convolution layer continuously learns the different characteristics of the input data. The pooling layer keeps the most important features while reducing the feature dimension to avoid overfitting. The full connection layer maps the resulting feature maps into a feature vector and generates a probability vector belonging to each class to achieve classification (Fazari et al., 2021; Weng et al., 2022). LeNet-5, a classical CNN, consists of two convolution layers, two pooling layers, two full connection layers, and one output layer (Priyadharshini et al., 2019). Image features were extracted by LeNet-5 without the final activation function.

2.5 Model construction

2.5.1 Conventional machine learning methods

SVM is a supervised machine learning algorithm. It is often used to solve classification or regression problems owing to its excellent generalization ability. It maps data to a high-dimensional space through nonlinear transformation (defining an appropriate kernel function); then, it constructs the optimal separated hyperplane in the high-dimensional space to transform a nonlinear problem into a linear problem (Huang et al., 2019). Random forest (RF) can strongly prevent overfitting and resist noise, as it combines a large number of decision trees and averages the results of all the decision trees to determine the final classification type (Yang H, et al., 2021).

2.5.2 Dense convolution neural networks

DenseNet, a mainstream learning method, was used to construct depth recognition models, enhancing feature transmission, encouraging feature reuse, and mitigating the gradient disappearance phenomenon (Li G, et al., 2020). It mainly comprises dense connecting blocks. Each layer in a dense block obtains additional inputs from all preceding layers and passes its feature maps to all subsequent layers, which can derive gradients directly from the loss function and the original input signal, leading to implicit deep supervision (Lawal, 2021). The spectra and image features were input as a 1D vector, and only one dense block was used in this study.

2.6 Performance evaluation

The HSI images of the tomato leaves that underwent three kinds of water treatments were divided into the calibration set and prediction set with a ratio of 7:3. In this study, the calibration set accuracy (ACC_C), prediction set accuracy (ACC_P), precision, recall, and F1-score of the prediction set, were used to evaluate model performance. All methods, including variable selection, image feature extraction, and model construction, were performed in Python 3.7.0. All programs were run on a computer with an Intel Core i7-3770 CPU, a main frequency of 3.40 GHz, and PyCharm software.

3 Results and discussion

3.1 Spectral analysis

Blue lights (3 W) were used as supplementary lighting in this study. Figure 4A displays the spectral differences in the 400–1000 nm range, and the spectral curve without blue lights was affected by noise within the 400–500 nm range. The deviation plot (Figure 4B) shows considerable high-frequency noise in the spectrum without blue lights in the 400–500 nm range, indicating that the auxiliary lighting of blue lights can effectively improve image quality. It is also evidenced by the results of the parallel experiment in Table S3.

The reflectance spectra for all the available young and mature leaves of tomato plants under different water treatments with the supplementary illumination of blue lights are shown in Figures 4C, D respectively. The low reflectance in the 450–700 nm range was due to the strong absorption of chlorophylls A and B for blue and red lights. A convex peak was present at 560 nm, and a high degree of tomato plant DS indicated a high reflectance value. This finding suggests that DS would reduce the concentration of photosynthetic pigments and weaken the light absorption capacity. The red edge phenomenon near 730 nm and the reflectance at a high level within 780–1000 nm were controlled by the internal structure of the leaf. The tomato plants with a serious stress degree had great blue shift distance for spectra. Plant self-protection mechanisms, such as leaf dehydration, stomatal closure, and leaf curling, would be activated under DS. However, this physiological state feedback was delayed relative to the spectral response. In brief, spectral changes determine the feasibility of HSI in identifying tomato plant DS.

3.2 DS identification based on spectra

SVM, RF, and DenseNet were used to develop the recognition models of tomato plant DS based on the full spectra and the spectra of EWs of young and mature leaves (Table 2). The parameter settings of models are shown in Table S4. In terms of the full spectra of the young leaves, SVM, RF, and DenseNet had ACC_C values of 90.90%, 83.63%, and 95.45%, respectively, and ACC_P values of 87.36%,74.73%, and 87.36%, respectively. Similarly, SVM gained the optimal recognition result for mature leaves with ACC_C =

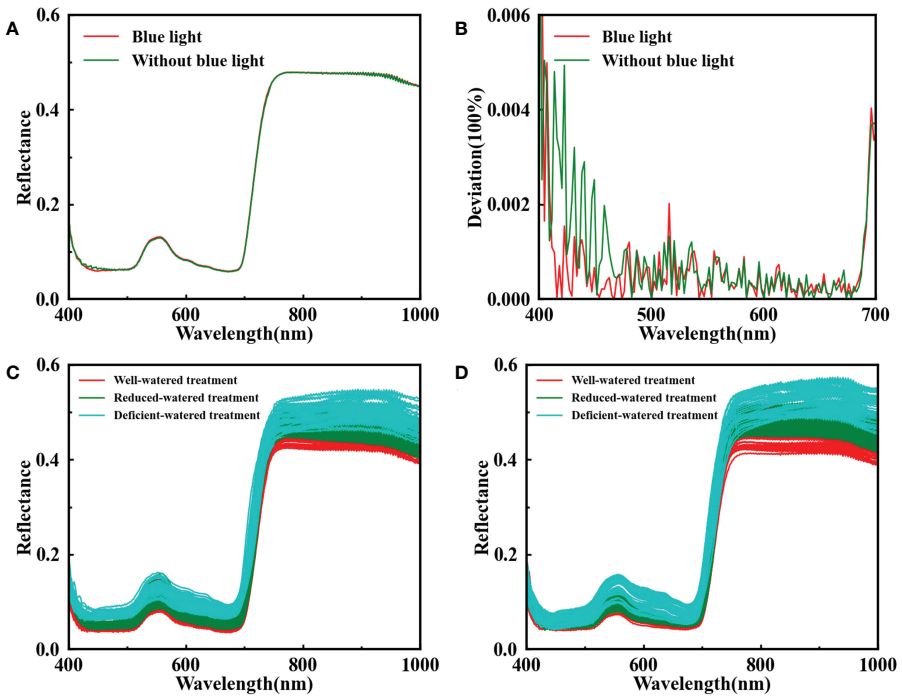


FIGURE 4
Reflectance spectra of the leaf with blue lights (red line) and without blue lights (green line) in the range of 400–1000 nm **(A)**; the deviation calculated by subtracting between the raw spectrum and the spectrum smoothed with Savitzky-Golay at 400–700 nm **(B)**; reflectance spectra for all available young leaves **(C)** and mature leaves **(D)** under different water treatments.

TABLE 2 Classification results of tomato plant DS based on full spectra and spectra of EWs.

Data types	Categories	Methods/Accuracy (%)					
		SVM		RF		DenseNet	
		ACC_C	ACC_P	ACC_C	ACC_P	ACC_C	ACC_P
Full spectra	Young leaves	90.90	87.36	83.63	74.73	95.45	87.36
	Mature leaves	94.09	91.57	93.18	86.31	97.27	88.42
Spectra of EWs	Young leaves	95.45	90.52	81.81	73.68	90.00	88.42
	Mature leaves	95.45	92.63	93.63	86.31	98.63	89.47

EWs, effective wavelengths; ACC_C, calibration set accuracy; ACC_P, prediction set accuracy; SVM, support vector machine; RF, random forest; DenseNet, dense convolutional network.

94.09% and ACC_P = 91.57%. The reason may be that the decision boundary of SVM was suited for the data distribution of tomato plant DS. With a powerful feature representation capability, DenseNet can accurately distinguish the DS levels of the tomato plant by combining low-level and deep features. RF performed slightly worse than the two other methods, it may be insensitive to tomato plant DS data.

Selecting the important variables beneficial to the learning algorithm can reduce the difficulty of the learning task and increase the interpretability of models. The spectra of EWs selected by GA were used to analyze tomato plant DS, and the ACC_P of the optimal SVM model improved by 3.16% for young leaves and 1.06% for mature leaves compared with the use of full spectra (Table 2). However, a slight deterioration in the identification result of RF was observed in young leaves because of a sharp decrease in the number of wavelengths. A total of 16 EWs for young leaves and 37 EWs for mature leaves were selected by GA, as shown in Figure 5. Specially, the EWs of young leaves all appear near the peak and valley in the visible region, meaning that young leaves are sensitive to changes in pigment concentration under DS. The EWs of mature leaves spread across the entire spectral range. Overall, the changes in the photosynthetic pigments and cell structure in the leaf were important indicators for evaluating tomato plant DS. The spectral information representing tomato plant DS differed between young and mature leaves. Therefore, the information on young and mature leaves can complement each other, probably facilitating DS analysis.

3.3 DS identification using spectra and image features

The direct use of leaf reflectance images with 272 wavelengths in HSI to identify tomato plant DS would consume a huge amount of time and reduce accuracy because of the high redundancy of HSI images. Knowing the correlation between reflectance images may be advantageous for extracting and synthesizing valid information and discarding useless information. Therefore, one young leaf sample and one mature leaf sample were randomly selected, and the correlation coefficient matrices of the EWs' reflectance images were calculated using Pearson correlation analysis. The heat map indicated the weak and strong correlations between the reflectance images of EWs, Further screening of the RIS helped remove collinearity variables (Figure 6). The RIS was determined by modeling analysis based on SVM for the spectra of EWs and image features extracted by LeNet-5 from different reflectance image combinations obtained by increasing the number of images in a sequence according to the correlation ranking of the first four reflectance images with weak correlation with the reference image (Table S5). The final RIS included the images of 607, 695, and 711 nm for young leaves and 567, 773, 791, and 822 nm for mature leaves. As shown in Figures 6A and 6B, the determined RIS was marked in the heat map. Any reflectance image in the RIS was weakly correlated with the other reflectance images, with correlation coefficients ranging from −0.3 to +0.3. Then, the image features extracted from RIS by LeNet-5 were combined with the spectra of EWs, called spectroscopy-image combination,

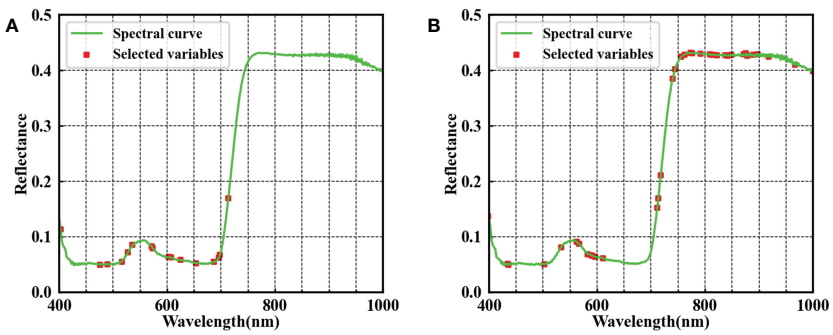


FIGURE 5
EWs selected by the GA of young **(A)** and mature **(B)** leaves.

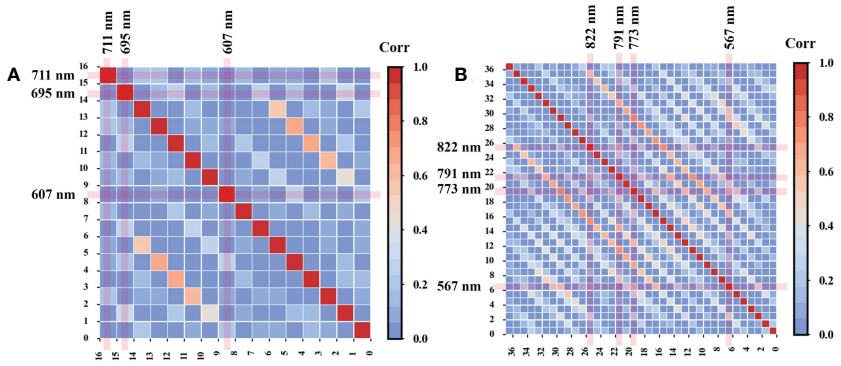


FIGURE 6
Pearson's correlation between the reflectance images of EWs in young **(A)** and mature **(B)** leaves.

to estimate the DS level (Table 3). The addition of new features resulted in different responses for the models. The accuracy of all models, except RF, was improved for young leaves, possibly because of the weak adaptability of RF to heterogeneous features and the presence of nonnegligible noise in image features. In regard to the mature leaves, the classification accuracy was improved by 1%–2% after the spectroscopy-image combination. The parameters of SVM, DenseNet, and RF are illustrated in Table S6. In general, the image features can replenish the missing spatial information, and spectroscopy-image combination offers an accurate stress analysis.

3.4 Effect of subsample fusion on DS analysis

The tomato leaves at different growth stages showed varied characteristics in color, composition, and appearance. The spectra and image features of young and mature leaves would provide the multilevel information of tomato plant DS. After subsample fusion (Table 4), a global optimal result of ACC_C = 95.90% and ACC_P = 95.78% was achieved by SVM. DenseNet also showed a strong identification ability of tomato plant DS. The detailed parameter settings of the models are shown in Table S7. The precision, recall, and F1-score values for the tomato plants with reduced-watered treatment were slightly lower than those in the two other water treatments. The reason is that reduced-watered treatment was in the intermediate state during the whole process of DS and easy to be misclassified into other classes. The confusion matrices of SVM, RF, and DenseNet (Figures 7A, C, E) showed that the tomato plants with reduced-watered treatment were easily misclassified as well-watered or deficient-watered tomato plants owing to inapparent differences in characteristics. Furthermore, the receiver operating characteristic (ROC) curves clearly showed the strong specificity and high sensitivity of SVM, followed by DenseNet and RF. The dispersed distributions of the ROC curves under the three water treatments indicated the presence of variances recognizing different classes (Figures 7B, D, F). Subsample fusion could augment the information difference of inter-class samples to promote further the effective identification of tomato plant DS. SVM performed well in identifying tomato plant DS with excellent accuracy and satisfactory robustness.

This study aims to provide a definite management decision for the rapid identification of tomato plant DS based on the multi-features of HSI and subsample fusion, which has significance in agricultural crop management and production practices (Table 5).

3.5 Discussion

3.5.1 Feasibility of HSI in plant DS

Traditional approaches for assessing plant DS include canopy temperature (Taghvaeian et al., 2014), chlorophyll fluorescence (Kautz et al., 2014), and thermal imaging (Jose Blaya-Ros et al., 2020). However, these measurements have low sensitivity and poor accuracy in practical applications, limiting their utility for rapid detection in the field. Furthermore, the concentration of substances in the leaves had already changed before DS caused changes in leaf traits. Recently, HSI has been widely used as a mainstream, rapid, and nondestructive measurement method in agriculture to obtain plant biological information reflecting metabolic changes (Zhao

TABLE 3 Classification results of tomato plant DS based on spectroscopy-image combination.

Strategy	Categories	Methods/Accuracy (%)					
		SVM		RF		DenseNet	
		ACC_C	ACC_P	ACC_C	ACC_P	ACC_C	ACC_P
Spectra of EWs + image features	Young leaves	94.09	91.57	80.90	70.52	90.45	89.47
	Mature leaves	96.36	93.68	99.09	87.36	98.18	90.52

10.3389/fpls.2023.1073530

TABLE 4 Classification results of tomato plant DS based on subsample fusion.

Methods	Classes	Accuracy (%)	Prediction Set		
			Precision (%)	Recall (%)	F1-score (%)
SVM	Well-watered	ACC_C=95.90 ACC_P= 95.78	97.06	97.06	97.06
	Reduced-watered		93.75	93.75	93.75
	Deficient-watered		96.55	96.55	96.55
RF	Well-watered	ACC_C= 96.81 ACC_P= 88.42	88.24	88.24	88.24
	Reduced-watered		84.38	84.38	84.38
	Deficient-watered		93.10	9310	93.10
DenseNet	Well-watered	ACC_C= 97.27 ACC_P= 94.73	94.12	94.12	94.12
	Reduced-watered		93.55	90.62	92.06
	Deficient-watered		96.67	100.00	98.31

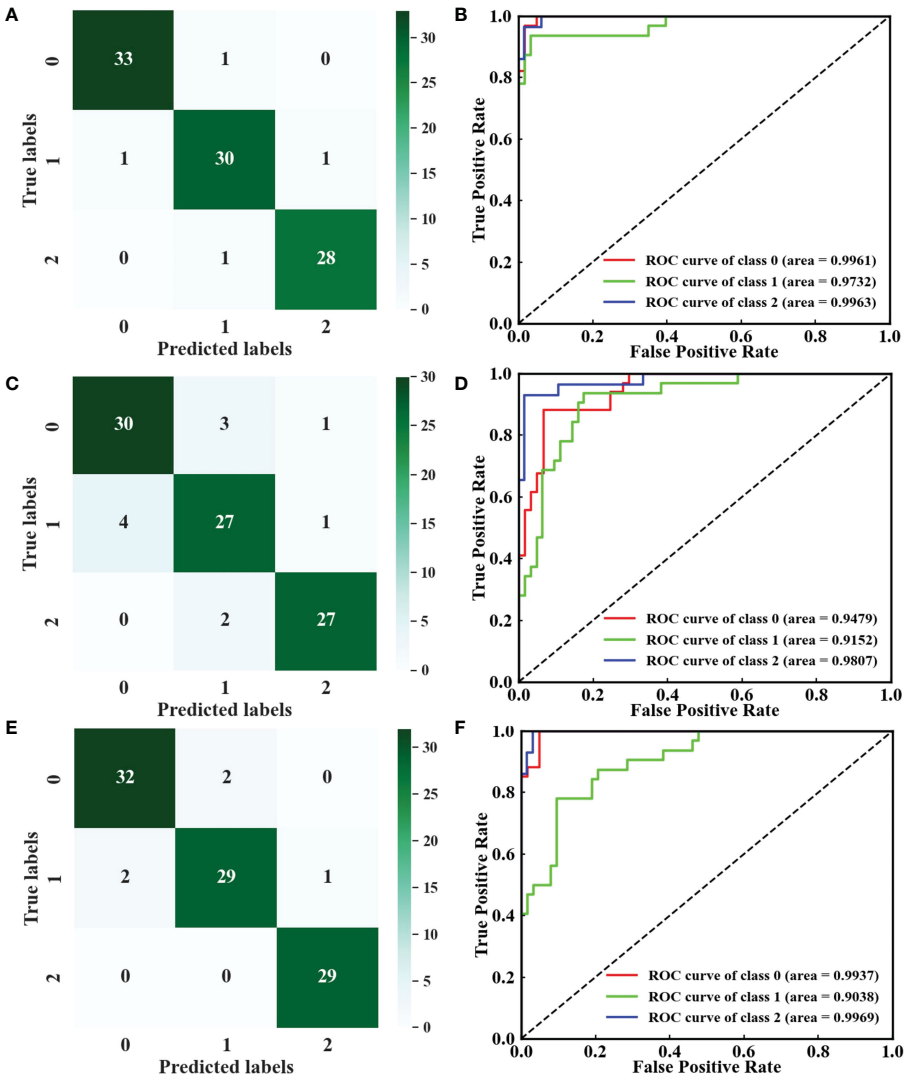


FIGURE 7
Confusion matrices and ROC curves of **(A, B)** SVM, **(C, D)** RF, and **(E, F)** DenseNet. Classes 0–2: well-watered, reduced-watered, and deficient-watered tomato plants, respectively.

TABLE 5 Significance of results.

Points	Significance
Supplement blue light	Suppress high-frequency noise and improve imaging quality effectively
Variable selection	Reduce computational cost and optimize classification performance
Spectroscopy-image combination	Synthesize multidimensional (spectral and spatial) information and improve information utilization
Subsample fusion	Integrate heterogeneous information and enhance classification model recognition ability

et al., 2020). Elvanidi et al. applied HSI to the detection of changes in the spectral reflectance of tomato plants under varying irrigation regimes to estimate plant water status under a controlled environment (Elvanidi et al., 2018), demonstrating the feasibility of HSI for nondestructive observation in tomato plant DS.

3.5.2 Utilization of spatial information

However, the utilization of spectra alone cannot show the imaging advantage of HSI in many spectral technologies, and the lack of spatial information leads to unsatisfactory results. Spectroscopy-image combination can fully use the information in the HSI images and effectively avoid information loss. Few previous studies have discussed the spectroscopy-image combination to identify DS in tomato plants. The superiority of the spectroscopy-image combination in this study may be due to the fusion of internal and external attributes in tomato leaves. However, most studies only applied wavelength selection methods to gain RIS (Weng et al., 2021), ignoring the relevancy degree between reflectance images. Wavelength selection methods rely on the reflectance values obtained by averaging all the pixels of the reflectance image. Correlation analysis directly considers the global information of the image. RIS was screened by comprehensively considering the spectral attributes and correlations between the images in this study. Simultaneously, image features were generally extracted by image statistical methods, such as GLCM, local binary pattern, and color moment (Xie et al., 2015; Lu et al., 2019). Sachar et al. summarize some methods for leaf image feature extraction without mentioning deep neural networks (Sachar and Kumar, 2021). Feature extraction using deep neural networks was proved feasible by Yang et al. and Zheng et al. (Yang W, et al., 2021; Zheng et al., 2022). Thus, LeNet-5 was used to extract image features automatically using the flexible network structure, avoiding complex mathematical analysis. The classification accuracy values of the spectroscopy-image combination models were ACC_C = 94.09% and ACC_P = 91.57% for young leaves and ACC_C = 96.36% and ACC_P = 93.68% for mature leaves, which were better than those of the model based on the spectra or the image features alone (Table S8). During the DS analysis of the plant, the possible reasons for the poor results obtained by relying only on images to distinguish DS degree are the insignificant differences in the appearance of tomato leaves with different stress levels, the influence of noise, and the low spatial resolution of the HSI images. Combining spectral and spatial information would be an effective solution for the first reason. Then, the noise can be removed by suitable supplemental lighting or by developing algorithms. In addition, super-resolution image reconstruction techniques are considered for enhancing the spatial resolution and improving the sensitivity of DS response.

3.5.3 Fusion strategy

The leaves at different growth stages of the same plant have different responses to DS. Other studies may point to the differences between young and mature leaves, but the data fusion approach for promoting DS analysis has rarely been applied. Subsample fusion, integrating the spectra and image features of young and mature leaves, can provide the differential information of multiple types of samples and reduce generalization error. The classification accuracy of the subsample fusion model was further improved, with ACC_C = 95.90% and ACC_P = 95.78%. Consequently, subsample fusion can supply holistic information about the tomato plant and enhance DS expression to fulfill the accurate analysis of tomato plant DS. Manually distinguishing the leaves at different growth stages is a laborious and time-consuming task. The algorithms must be developed to distinguish leaves at different growth stages automatically by texture, color, and morphological information from an image.

3.5.4 Effects of blue light on imaging

Besides, HSI systems using halogen lamps tend to have inferior SNR in the blue region (400–500 nm) of the electromagnetic spectrum, because a minimal amount of light is divided into the visible region for the HSI spectrometer with high spectral resolution (Mahlein et al., 2015). Previous works determined that low light intensity causes a dark current noise effect in the spectral profile of HSI images (Manea and Calin, 2015; Zhang et al., 2019). Moreover, the spectral curves in the 400–500 nm range, which depend on the amount of light absorbed by leaf pigments (chlorophyll, carotenoids, and anthocyanins), can reflect the physiological health information of plants (Zhao et al., 2016). DS affects metabolic reactions and the synthesis of photosynthetic pigments in the plant, and the response mechanism suggests that the visible light region is particularly sensitive to DS (Ihuoma and Madramootoo, 2019; Ihuoma and Madramootoo, 2020). High-frequency noise is suppressed by adding blue lights, and SNR is ameliorated effectively.

3.5.5 Challenges, improvements, and developments

The detection of tomato plant DS still faces many challenges. For example, Susič et al. and Žibrat et al. indicated that both root-knot nematodes (biotic stress) and water deficiency (abiotic stress) lead to similar drought symptoms in plants (Susic et al., 2018; Zibrat

et al., 2019). In the actual planting environment, the accurate differentiation of the biotic stress with similar symptoms of DS can contribute to the prevention of misidentification and inappropriate preventive measures affecting plant survival rate. Alordzinu et al. reported that plant responses to water stress are articulated by various physiological and biophysical changes and soil properties, they also assessed the water stress of tomato plants at different growth periods under different soils (Alordzinu et al., 2021). The mechanisms of resistance to tomato plant DS at different growth periods need to be further determined. Soils faced with widely cultivated tomato plants will be the focus of our future research. The specific effects of nutrients in the soil on the plant also need to be explored. Burnett et al. considered that spectroscopy can detect DS by investigating the potential biochemical changes before visual differences are observed, and metabolic responses to DS can be detected by HSI (Burnett et al., 2021). We should pay attention to the changes in the internal components of tomato leaves under DS, such as chlorophyll, soluble protein, catalase and so on. Moreover, it was shown that chlorophyll content, protein content decreased under DS, and the deeper the stress degree is, the lower the content is. The enzymatic activity of superoxide dismutase is significantly enhanced under severe DS, whereas catalase has a slight enhancement (Zgallai et al., 2006). We will attempt to estimate quantitatively the leaf metabolite concentrations pointing to tomato plant DS and achieve the early identification of DS. Two different types of leaves may be insufficient for estimating the physiological state of the plant system. The stress-induced changes in the physiological, biochemical, and molecular attributes of different plant organs need to be investigated. In this way, the health status, stress tolerance, and complex adaptation mechanisms of plant can be comprehensively assessed.

Additionally, some other aspects also need to be improved. (1) With the application of HSI in a greenhouse or a field, the uncontrollable lighting, complex background (soil, weeds, etc.), and mutual interference between plants can lead to incorrect DS analysis. Therefore, image correction and background segmentation should be considered. (2) In our work, the deep learning recognition models exhibited a barely prominent performance owing to the lack of training samples. Increasing the sampling data may improve the identification. Simultaneously, tomato plants with diverse stresses, including diseases and pests, various varieties, and different growth periods need to be further researched. Novel modeling algorithms deserve to be developed to accommodate heterogeneous samples and optimize classification performance. (3) Hyperspectral sensors can be mounted on the unmanned aerial vehicle (UAV) to perform large-scale DS detection and timely management. The trajectory, cycle, and height for the flight of the UAV, as well as the speed and range for the lens scanning of the HSI spectrometer, should be further explored.

During plant protection management, it is necessary to consider not only improving the accuracy of DS identification in plants, but also raising the adaptability of the plants to DS. Plant growth depends on the absorption of water from the soil and its transfer from roots to other plant parts. Therefore, understanding drought-induced changes to root anatomical traits is important to enhance plant drought adaptation (Alagoz et al., 2022). Different plant growth regulators respond differently to plant DS. Studying the effects of different plant growth regulators on plant physiological and biochemical processes can also help to promote the drought resistance of plants (Ghassemi et al., 2018). Along with the research on DS, the influence of fertilizer application on plants growth and biodiversity should also be discussed (Li et al., 2022). The use of chemical fertilizers remains controversial, so finding alternatives to chemical fertilizers for ecological sustainability is one of the pressing issues in modern agriculture (Sun et al., 2022). Overall, modern agricultural management aims to improve plant resistance while identifying stresses accurately and intervening scientifically.

4 Conclusion

In this work, the identification of tomato plant DS was performed using the multi-features of HSI and subsample fusion. The addition of blue lights removed the high-frequency noise in the 400–500 nm region. The reflectance spectra extracted from the HSI images showed that reflectance increased with the severity of tomato plant DS. Moreover, the image features extracted from the RIS by LeNet-5 positively affected the improvement of the model performance. Spectroscopy-image combination obtained good results with ACC_C = 94.09% and ACC_P = 91.57% for young leaves and ACC_C = 96.36% and ACC_P = 93.68% for mature leaves, which are superior to the identification accuracy values of the modeling by spectra or image features alone. Moreover, the classification accuracy of the subsample fusion model was further improved with ACC_C = 95.90% and ACC_P = 95.78%. In summary, the multi-features of HSI and the subsample fusion yielded an accurate identification of tomato plant DS under the supplementary illumination of blue lights. Applying HSI in complex environments, adding sample types and sample size, optimizing modeling algorithms, and utilizing of UAV equipped with an HSI spectrometer should be considered in future explorations to establish a stable, precise, and comprehensive classification model for various stress types and stress degrees.

Data availability statement

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

Author contributions

SW: Methodology, Conceptualization, Supervision, Funding acquisition. JM: Methodology, Validation, Formal analysis, Investigation, Visualization, Writing-original draft. WT: Writing–review & editing, Visualization. YT: Software, Investigation. MP: Investigation. ZZ: Validation, Formal analysis. LH: Resources, Methodology. LZ: Methodology, Supervision. JZ: Methodology, Validation. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (Nos. 32001421, 31971789), the Key Research and Development Program of Anhui Province (Nos. 202004a06020032, 202104a06020025), and Innovation and the Entrepreneurship team project of Lu'an Branch, Anhui Institute of Innovation for Industrial Technology (LAY-2020-CXG-008).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1073530/full#supplementary-material

References

Alagoz, S. M., Zahra, N., Kamrani, M. H., Lajayer, B. A., Nobaharan, K., Astatkie, T., et al. (2022). Role of root hydraulics in plant drought tolerance. *J. Plant Growth Regul* 1-16. doi: 10.1007/s00344-022-10807-x

Alordzinu, K. E., Li, J. H., Lan, Y. B., Appiah, S. A., Al Aasmi, A., Wang, H., et al. (2021). Ground-based hyperspectral remote sensing for estimating water stress in tomato growth in sandy loam and silty loam soils. *Sensors* 21 (17), 5705. doi: 10.3390/s21175705

Asaari, M. S. M., Mertens, S., Dhondt, S., Inze, D., Wuyts, N., and Scheunders, P. (2019). Analysis of hyperspectral images for detection of drought stress and recovery in maize plants in a high-throughput phenotyping platform. *Comput. Electron. Agric.* 162, 749–758. doi: 10.1016/j.compag.2019.05.018

Borraz-Martinez, S., Boque, R., Simo, J., Mestre, M., and Gras, A. (2019). Development of a methodology to analyze leaves from prunus dulcis varieties using near infrared spectroscopy. *Talanta* 204, 320–328. doi: 10.1016/j.talanta.2019.05.105

Burnett, A. C., Serbin, S. P., Davidson, K. J., Ely, K. S., and Rogers, A. (2021). Detection of the metabolic response to drought stress using hyperspectral reflectance. *J. Exp. Bot.* 72 (18), 6474–6489. doi: 10.1093/jxb/erab255

Chen, S., Gao, Y., Fan, K., Shi, Y., Luo, D., Shen, J., et al. (2021). Prediction of drought-induced components and evaluation of drought damage of tea plants based on hyperspectral imaging. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.695102

Chen, J., Kang, S., Du, T., Guo, P., Qiu, R., Chen, R., et al. (2014). Modeling relations of tomato yield and fruit quality with water deficit at different growth stages under greenhouse condition. *Agric. Water Management* 146 (1), 131–148. doi: 10.1016/j.agwat.2014.07.026

Das, B., Sahoo, R. N., Pargal, S., Krishna, G., Verma, R., Viswanathan, C., et al. (2021). Evaluation of different water absorption bands, indices and multivariate models for water-deficit stress monitoring in rice using visible-near infrared spectroscopy. *Spectrochimica Acta Part a-Molecular Biomolecular Spectrosc.* 247, 119104. doi: 10.1016/j.saa.2020.119104

Eid, M. A. M., Abdel-Salam, A. A., Salem, H. M., Mahrous, S. E., Seleiman, M. F., Alsadon, A. A., et al. (2020). Interaction effects of nitrogen source and irrigation regime on tuber quality, yield, and water use efficiency of solanum tuberosum l. *Plants-Basel* 9 (1), 110. doi: 10.3390/plants9010110

Elvanidi, A., Katsoulas, N., Ferentinos, K. P., Bartzanas, T., and Kittas, C. (2018). Hyperspectral machine vision as a tool for water stress severity assessment in soilless tomato crop. *Biosyst. Eng.* 165, 25–35. doi: 10.1016/j.biosystemseng.2017.11.002

Fazari, A., Pellicer-Valero, O. J., Gomez-Sanchis, J., Bernardi, B., Cubero, S., Benalia, S., et al. (2021). Application of deep convolutional neural networks for the detection of anthracnose in olives using VIS/NIR hyperspectral images. *Comput. Electron. Agric.* 187, 106252. doi: 10.1016/j.compag.2021.106252

Flexas, J., Bota, J., Galmes, J., Medrano, H., and Ribas-Carbo, M. (2006). Keeping a positive carbon balance under adverse conditions: responses of photosynthesis and respiration to water stress. *Physiologia Plantarum* 127 (3), 343–352. doi: 10.1111/j.1399-3054.2006.00621.x

Gao, Z., Khot, L. R., Naidu, R. A., and Zhang, Q. (2020). Early detection of grapevine leafroll disease in a red-berried wine grape cultivar using hyperspectral imaging. *Comput. Electron. Agric.* 179 (1), 105807. doi: 10.1016/j.compag.2020.105807

Ghassemi, S., Farhangi-Abriz, S., Faegi-Analou, R., Ghorbanpour, M., and Lajayer, B. A. (2018). Monitoring cell energy, physiological functions and grain yield in field-grown mung bean exposed to exogenously applied polyamines under drought stress. *J. Soil Sci. Plant Nutr.* 18 (4), 1108–1125. doi: 10.4067/S0718-95162018005003102

Guo, A., Huang, W., Ye, H., Dong, Y., Ma, H., Ren, Y., et al. (2020). Identification of wheat yellow rust using spectral and texture features of hyperspectral images. *Remote Sens.* 12 (9), 1419. doi: 10.3390/rs12091419

Han, M., Zhang, H., DeJonge, K. C., Comas, L. H., and Trout, T. J. (2016). Estimating maize water stress by standard deviation of canopy temperature in thermal imagery. *Agric. Water Manage.* 177, 400–409. doi: 10.1016/j.agwat.2016.08.031

Hao, S., Cao, H., Wang, H., and Pan, X. (2019). The physiological responses of tomato to water stress and re-water in different growth periods. *Scientia Hortic.* 249, 143–154. doi: 10.1016/j.scienta.2019.01.045

Haworth, M., Marino, G., and Centritto, M. (2018). An introductory guide to gas exchange analysis of photosynthesis and its application to plant phenotyping and precision irrigation to enhance water use efficiency. *J. Water Climate Change* 9 (4), 786–808. doi: 10.2166/wcc.2018.152

Huang, L., Wu, Z., Huang, W., Ma, H., and Zhao, J. (2019). Identification of fusarium head blight in winter wheat ears based on fisher's linear discriminant analysis and a support vector machine. *Appl. Sciences-Basel* 9 (18), 3894. doi: 10.3390/app9183894

Ihuoma, S. O., and Madramootoo, C. A. (2019). Sensitivity of spectral vegetation indices for monitoring water stress in tomato plants. *Comput. Electron. Agric.* 163, 104860. doi: 10.1016/j.compag.2019.104860

Ihuoma, S. O., and Madramootoo, C. A. (2020). Narrow-band reflectance indices for mapping the combined effects of water and nitrogen stress in field grown tomato crops. *Biosyst. Eng.* 192, 133–143. doi: 10.1016/j.biosystemseng.2020.01.017

Jose Blaya-Ros, P., Blanco, V., Domingo, R., Soto-Valles, F., and Torres-Sanchez, R. (2020). Feasibility of low-cost thermal imaging for monitoring water stress in young and mature sweet cherry trees. *Appl. Sciences-Basel* 10 (16), 5461. doi: 10.3390/app10165461

Kautz, B., Noga, G., and Hunsche, M. (2014). Sensing drought- and salinity-imposed stresses on tomato leaves by means of fluorescence techniques. *Plant Growth Regul.* 73 (3), 279–288. doi: 10.1007/s10725-014-9888-x

Key, S., Baygin, M., Demir, S., Dogan, S., and Tuncer, T. (2022). Meniscal tear and ACL injury detection model based on AlexNet and iterative ReliefF. *J. Digital Imaging* 35 (2), 200–212. doi: 10.1007/s10278-022-00581-3

Lawal, O. M. (2021). Development of tomato detection model for robotic platform using deep learning. *Multimedia Tools Appl.* 80 (17), 26751–26772. doi: 10.1007/s11042-021-10933-w

Li, J., Charles, L. S., Yang, Z., Du, G., and Fu, S. (2022). Differential mechanisms drive species loss under artificial shade and fertilization in the alpine meadow of the Tibetan plateau. *Front. Plant Science* 13, 63. doi: 10.3389/fpls.2022.832473

Li, G., Zhang, C., Lei, R., Zhang, X., Ye, Z., and Li, X. (2020). Hyperspectral remote sensing image classification using three-dimensional-squeeze-and-excitation-DenseNet (3D-SE-DenseNet). *Remote Sens. Lett.* 11 (2), 195–203. doi: 10.1080/2150704x.2019.1697001

Li, P., Zhang, X., Li, S., Du, G., Jiang, L., Liu, X., et al. (2020). A rapid and nondestructive approach for the classification of different-age citri reticulatae pericarpium using portable near infrared spectroscopy. *Sensors* 20 (6), 1586. doi: 10.3390/s20061586

Lu, B., Sun, J., Yang, N., Wu, X.-h., and Zhou, X. (2019). Prediction of tea diseases based on fluorescence transmission spectrum and texture of hyperspectral image. *Spectrosc. Spectral Anal.* 39 (8), 2515–2521. doi: 10.3964/j.issn.1000-0593(2019)08-2515-07

Ma, D., Maki, H., Neeno, S., Zhang, L., Wang, L., and Jin, J. (2020). Application of non-linear partial least squares analysis on prediction of biomass of maize plants using hyperspectral images. *Biosyst. Eng.* 200, 40–54. doi: 10.1016/j.biosystemseng.2020.09.002

Mahlein, A.-K., Hammersley, S., Oerke, E.-C., Dehne, H.-W., Goldbach, H., and Grieve, B. (2015). Supplemental blue LED lighting array to improve the signal quality in hyperspectral imaging of plants. *Sensors* 15 (6), 12834–12840. doi: 10.3390/s150612834

Manea, D., and Calin, M. A. (2015). Hyperspectral imaging in different light conditions. *Imaging Sci. J.* 63 (4), 214–219. doi: 10.1179/1743131x15y.0000000001

Melandri, G., Thorp, K. R., Broeckling, C., Thompson, A. L., Hinze, L., and Pauli, D. (2021). Assessing drought and heat stress-induced changes in the cotton leaf metabolome and their relationship with hyperspectral reflectance. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.751868

Mishra, P., Karami, A., Nordon, A., Rutledge, D. N., and Roger, J.-M. (2019). Automatic de-noising of close-range hyperspectral images with a wavelength-specific shearlet-based image noise reduction method. *Sensors Actuators B-Chemical* 281, 1034–1044. doi: 10.1016/j.snb.2018.11.034

Moharana, S., and Dutta, S. (2019). Estimation of water stress variability for a rice agriculture system from space-borne hyperion imagery. *Agric. Water Manage.* 213, 260–269. doi: 10.1016/j.agwat.2018.10.001

Mukhtar, T., Rehman, S. U., Smith, D., Sultan, T., Seleiman, M. F., Alsadon, A. A., et al. (2020). Mitigation of heat stress in solanum lycopersicum l. by ACC-deaminase and exopolysaccharide producing bacillus cereus: Effects on biochemical profiling. *Sustainability* 12 (6), 2159. doi: 10.3390/su12062159

Ng, W., Minasny, B., Malone, B. P., Sarathjith, M. C., and Das, B. S. (2019). Optimizing wavelength selection by using informative vectors for parsimonious infrared spectra modelling. *Comput. Electron. Agric.* 158, 201–210. doi: 10.1016/j.compag.2019.02.003

Ni, Z., Liu, Z., Huo, H., Li, Z.-L., Nerry, F., Wang, Q., et al. (2015). Early water stress detection using leaf-level measurements of chlorophyll fluorescence and temperature data. *Remote Sens.* 7 (3), 3232–3249. doi: 10.3390/rs70303232

Pandey, P., Ge, Y., Stoerger, V., and Schnable, J. C. (2017). High throughput *In vivo* analysis of plant leaf chemical properties using hyperspectral imaging. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.01348

Priyadharshini, R. A., Arivazhagan, S., Arun, M., and Mirnalini, A. (2019). Maize leaf disease classification using deep convolutional neural networks. *Neural Computing Appl.* 31 (12), 8887–8895. doi: 10.1007/s00521-019-04228-3

Rad, S. A., Dehghanian, Z., Lajayer, B. A., Nobaharan, K., and Astatkie, T. (2022). Mitochondrial respiration and energy production under some abiotic stresses. *J. Plant Growth Regul.* 41 (8), 3285–3299. doi: 10.1007/s00344-021-10512-1

Raddi, S., Giannetti, F., Martini, S., Farinella, F., Chirici, G., Tani, A., et al. (2022). Monitoring drought response and chlorophyll content in quercus by consumer-grade, near-infrared (NIR) camera: a comparison with reflectance spectroscopy. *New Forests* 53 (2), 241–265. doi: 10.1007/s11056-021-09848-z

Ru, C., Li, Z., and Tang, R. (2019). A hyperspectral imaging approach for classifying geographical origins of rhizoma atractylodis macrocephalae using the fusion of spectrum-image in VNIR and SWIR ranges (VNIR-SWIR-FuSI). *Sensors* 19 (9), 2045. doi: 10.3390/s19092045

Sachar, S., and Kumar, A. (2021). Survey of feature extraction and classification techniques to identify plant through leaves. *Expert Syst. Appl.* 167, 114181. doi: 10.1016/j.eswa.2020.114181

Seleiman, M. F., Al-Suhaibani, N., Ali, N., Akmal, M., Alotaibi, M., Refay, Y., et al. (2021). Drought stress impacts on plants and different approaches to alleviate its adverse effects. *Plants-Basel* 10 (2), 259. doi: 10.3390/plants10020259

Song, D., Gao, D., Sun, H., Qiao, L., Zhao, R., Tang, W., et al. (2021). Chlorophyll content estimation based on cascade spectral optimizations of interval and wavelength characteristics. *Comput. Electron. Agric.* 189, 106413. doi: 10.1016/j.compag.2021.106413

Steidle Neto, A. J., Lopes, D. C., Pinto, F. A. C., and Zolnier, S. (2017). Vis/NIR spectroscopy and chemometrics for non-destructive estimation of water and chlorophyll status in sunflower leaves. *Biosyst. Eng.* 155, 124–133. doi: 10.1016/j.biosystemseng.2016.12.008

Sun, J., Jia, Q., Li, Y., Zhang, T., Chen, J., Ren, Y., et al. (2022). Effects of arbuscular mycorrhizal fungi and biochar on growth, nutrient absorption, and physiological properties of maize (Zea mays l.). *J. Fungi* 8 (12), 1275. doi: 10.3390/jof8121275

Sun, Y., Pessane, I., Pan, L., and Wang, X. (2021). Hyperspectral characteristics of bruised tomatoes as affected by drop height and fruit size. *Lwt-Food Sci. Technol.* 141, 110863. doi: 10.1016/j.lwt.2021.110863

Susic, N., Zibrat, U., Sirca, S., Strajnar, P., Razinger, J., Knapic, M., et al. (2018). Discrimination between abiotic and biotic drought stress in tomatoes using hyperspectral imaging. *Sensors Actuators B-Chemical* 273, 842–852. doi: 10.1016/j.snb.2018.06.121

Taghizadeh, M., Gowen, A. A., and O'Donnell, C. P. (2011). Comparison of hyperspectral imaging with conventional RGB imaging for quality evaluation of agaricus bisporus mushrooms. *Biosyst. Eng.* 108 (2), 191–194. doi: 10.1016/j.biosystemseng.2010.10.005

Taghvaeian, S., Comas, L., DeJonge, K. C., and Trout, T. J. (2014). Conventional and simplified canopy temperature indices predict water stress in sunflower. *Agric. Water Manage.* 144, 69–80. doi: 10.1016/j.agwat.2014.06.003

Wang, L., Liu, D., Pu, H., Sun, D.-W., Gao, W., and Xiong, Z. (2015). Use of hyperspectral imaging to discriminate the variety and quality of rice. *Food Analytical Methods* 8 (2), 515–523. doi: 10.1007/s12161-014-9916-5

Weng, S., Chu, Z., Wang, M., Han, K., Zhu, G., Liu, C., et al. (2022). Reflectance spectroscopy with operator difference for determination of behenic acid in edible vegetable oils by using convolutional neural network and polynomial correction. *Food Chem.* 367, 130668. doi: 10.1016/j.foodchem.2021.130668

Weng, S., Han, K., Chu, Z., Zhu, G., Liu, C., Zhu, Z., 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

Xia, J. A., Zhang, W., Zhang, W., Yang, Y., Hu, G., Ge, D., et al. (2021). A cloud computing-based approach using the visible near-infrared spectrum to classify greenhouse tomato plants under water stress. *Comput. Electron. Agric.* 181. doi: 10.1016/j.compag.2020.105966

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, 16564. doi: 10.1038/srep16564

Yang, X., Hong, H., You, Z., and Cheng, F. (2015). Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. *Sensors* 15 (7), 15578–15594. doi: 10.3390/s150715578

Yang, H., Li, F., Wang, W., and Yu, K. (2021). Estimating above-ground biomass of potato using random forest and optimized hyperspectral indices. *Remote Sens.* 13 (12). doi: 10.3390/rs13122339

Yang, B., Ma, J., Yao, X., Cao, W., and Zhu, Y. (2021). Estimation of leaf nitrogen content in wheat based on fusion of spectral features and deep features from near infrared hyperspectral imagery. *Sensors* 21 (2). doi: 10.3390/s21020613

Yang, W., Nigon, T., Hao, Z., Paiao, G. D., Fernandez, F. G., Mulla, D., et al. (2021). Estimation of corn yield based on hyperspectral imagery and convolutional neural network. *Comput. Electron. Agric.* 184. doi: 10.1016/j.compag.2021.106092

Yu, C., Li, F., Chang, C.-I., Cen, K., and Zhao, M. (2020). "*Deep 2D convolutional neural network with deconvolution layer for hyperspectral image classification*," in *Paper presented at the 7th International Conference on Communications, Signal Processing, and Systems (CSPS)* Lecture Notes in Electrical Engineering (Dalian, Peoples R China: Springer-Verlag Singapore Pte Ltd152 Beach Road, #21-01/04 Gateway East, Singapore, 189721, Singapore).

Zgallai, H., Steppe, K., and Lemeur, R. (2006). Effects of different levels of water stress on leaf water potential, stomatal resistance, protein and chlorophyll content and certain anti-oxidative enzymes in tomato plants. *J. Integr. Plant Biol.* 48 (6), 679–685. doi: 10.1111/j.1744-7909.2006.00272.x

Zhang, L., Maki, H., Ma, D., Sanchez-Gallego, J. A., Mickelbart, M. V., Wang, L., et al. (2019). Optimized angles of the swing hyperspectral imaging system for single corn plant. *Comput. Electron. Agric.* 156, 349–359. doi: 10.1016/j.compag.2018.11.030

Zhao, Y.-R., Li, X., Yu, K.-Q., Cheng, F., and He, Y. (2016). Hyperspectral imaging for determining pigment contents in cucumber leaves in response to angular leaf spot disease. *Sci. Rep.* 6, 27790. doi: 10.1038/srep27790

Zhao, T., Nakano, A., Iwaski, Y., and Umeda, H. (2020). Application of hyperspectral imaging for assessment of tomato leaf water status in plant factories. *Appl. Sciences-Basel* 10 (13). doi: 10.3390/app10134665

Zheng, L., Bao, Q., Weng, S., Tao, J., Zhang, D., Huang, L., et al. (2022). Determination of adulteration in wheat flour using multi-grained cascade forest-related models coupled with the fusion information of hyperspectral imaging. *Spectrochimica Acta Part a-Molecular Biomolecular Spectrosc.* 270. doi: 10.1016/j.saa.2021.120813

Zhou, J.-J., Zhang, Y.-H., Han, Z.-M., Liu, X.-Y., Jian, Y.-F., Hu, C.-G., et al. (2021). Evaluating the performance of hyperspectral leaf reflectance to detect water stress and estimation of photosynthetic capacities. *Remote Sens.* 13 (11), 2160. doi: 10.3390/rs13112160

Zibrat, U., Susic, N., Knapic, M., Sirca, S., Strajnar, P., Razinger, J., et al. (2019). Pipeline for imaging, extraction, pre-processing, and processing of time-series hyperspectral data for discriminating drought stress origin in tomatoes. *MethodsX* 6, 399–408. doi: 10.1016/j.mex.2019.02.022

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 April 2023
DOI 10.3389/fpls.2023.1124939

OPEN ACCESS

EDITED BY
Yijun Yan,
Robert Gordon University, United Kingdom

REVIEWED BY
Shanwen Zhang,
Xijing University, China
Zhenyu Fang,
Northwestern Polytechnical University, China
Dianlong You,
University of Technology Sydney, Australia

*CORRESPONDENCE
Taosheng Xu
taosheng.x@gmail.com

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

RECEIVED 15 December 2022
ACCEPTED 10 March 2023
PUBLISHED 18 April 2023

CITATION
Hu Y, Meng A, Wu Y, Zou L, Jin Z and Xu T (2023) Deep-agriNet: a lightweight attention-based encoder-decoder framework for crop identification using multispectral images.
Front. Plant Sci. 14:1124939.
doi: 10.3389/fpls.2023.1124939

Deep-agriNet: a lightweight attention-based encoder-decoder framework for crop identification using multispectral images

Yimin Hu[1,2], Ao Meng[1], Yanjun Wu[2,3], Le Zou[1], Zhou Jin[2] and Taosheng Xu[2]*

[1]School of Big Data And Artificial Intelligence, Hefei University, Hefei, China, [2]Institute of Intelligent Machines, Hefei Institutes of Physical Science, Chinese Academy of Science, Hefei, China, [3]Science Island Branch, University of Science and Technology of China, Hefei, China

The field of computer vision has shown great potential for the identification of crops at large scales based on multispectral images. However, the challenge in designing crop identification networks lies in striking a balance between accuracy and a lightweight framework. Furthermore, there is a lack of accurate recognition methods for non-large-scale crops. In this paper, we propose an improved encoder-decoder framework based on DeepLab v3+ to accurately identify crops with different planting patterns. The network employs ShuffleNet v2 as the backbone to extract features at multiple levels. The decoder module integrates a convolutional block attention mechanism that combines both channel and spatial attention mechanisms to fuse attention features across the channel and spatial dimensions. We establish two datasets, DS1 and DS2, where DS1 is obtained from areas with large-scale crop planting, and DS2 is obtained from areas with scattered crop planting. On DS1, the improved network achieves a mean intersection over union (mIoU) of 0.972, overall accuracy (OA) of 0.981, and recall of 0.980, indicating a significant improvement of 7.0%, 5.0%, and 5.7%, respectively, compared to the original DeepLab v3+. On DS2, the improved network improves the mIoU, OA, and recall by 5.4%, 3.9%, and 4.4%, respectively. Notably, the number of parameters and giga floating-point operations (GFLOPs) required by the proposed Deep-agriNet is significantly smaller than that of DeepLab v3+ and other classic networks. Our findings demonstrate that Deep-agriNet performs better in identifying crops with different planting scales, and can serve as an effective tool for crop identification in various regions and countries.

KEYWORDS

multispectral image, crop identification, feature extraction, encoder-decoder, lightweight, DeepLab v3+

1 Introduction

Timely identification of large-scale crops is vital for agricultural production, which can provide an important basis for yield estimation, structure adjustment and optimization of agricultural management (Becker-Reshef et al., 2010). The traditional identification methods of farm crops are mainly based on statistical statement, but the outdated method restricts identification efficiency and increases labor costs (Tan et al., 2020). Recently, a variety of automated detecting technologies have been proposed in crop identification and achieved lots of successful applications (Waldhoff et al., 2017; Longato et al., 2019; Xu et al., 2019). Remote sensing, as a large-scale non-contact monitoring technology, plays an extremely important role in modern agriculture (Shi et al., 2019). Identification of farm crops in remote sensing images in large-scale farmland can obtain the spatial location information of farmland and the ground attachment. The related information helps agricultural administrators to figure out the distribution and planting structure of regional species from a macro perspective, thereby formulate more accurate and efficient agricultural policies.

Crop identification based on remote sensing has been a research theme of considerable interest, which is of great value in the field of precision agriculture. With the development of image processing and artificial intelligence, the technologies of crop identification can be summarized into three streams. In the first stream, the traditional remote sensing feature extraction is mainly based on spectral, spatial, and temporal features (Zhang et al., 2016; Qiong et al., 2017; Sun et al., 2019b). Tian et al. (2021a) analyzed spectral characteristics and vegetation indices at each growth stage of crops and used reasonable thresholds to screen these parameters and successfully identified winter wheat and garlic planting areas. The result shows that varying vegetation indices could effectively distinguish crops with different spectral characteristics. Li et al., (2015) used the Stepwise Discriminant Analysis (SDA) method for feature selection from the Landsat MODIS Enhanced time series data and screened out 10 optimal features for crop classification. In the second stream, Machine learning methods are widely used in the field of large-scale crop identification due to their heuristic learning strategy and accelerated training mechanism (Jia et al., 2019; Zhang et al., 2019; Tian et al., 2021b). Zheng et al. (2015) applied Support Vector Machines (SVMs) to time-series Landsat images of Arizona to test its ability to discriminate between multiple crop types in a complex cropping system. Han et al. (2022) extracted relevant features of corn lodging regions and proposed the SMOTE-ENN-XGBoost model based on the Synthetic Minority Oversampling Technique (SMOTE) and Edited Nearest Neighbor (ENN) methods, which showed an F1 score of 0.930 and a recall rate of 0.899 on the lodging detection test set. With the proposal of the Convolutional Neural Network (CNN), deep learning leads the third stream for crop identification using remote-sensing images (Kuwata and Shibasaki, 2015; Wang et al., 2020; Yuan et al., 2020). Yu et al. (2022) improved the U-Net network by introducing the Involution operator and Dense block module and proposed a wheat lodging evaluation method based on UAV multispectral images. Kussul et al. (2017) proposed a multi-level Deep Learning (DL) method using multi-temporal land cover and crop type classification to identify crops in a heterogeneous environment, achieving a target accuracy of 85% for major crops. The proposal of attentional mechanisms has dramatically advanced the field of deep learning. Naturally, this ingenious mechanism has also been widely used for crop identification with great success (Jin et al., 2021). Wang et al. (2022) proposed a novel architecture called Coupled CNN and Transformer Network (CCTNet), which combines the local details of CNN and the global context of the transformer to achieve a 72.97% mIoU score on the Barley remote-sensing dataset. Lu et al. (2022) proposed a deep neural network with Dual Attention and Scale Fusion (DASFNet) to extract farmland from GF-2 images of southern Xinjian. The result shows that the dual attention mechanism module can correct the shape and boundary of the fields effectively.

The above methods show excellent performances in their respective datasets. However, these datasets are mainly derived from areas where crops are grown on a large scale. In fact, the vast majority of China's regions are planted discretely, and the plots under this type of planting are relatively tiny, making it difficult for existing networks to achieve high-precision crop identification. In addition, a high-precision network is often accompanied by a considerable amount of parameter calculation, which is difficult to be applied to low-end agricultural equipment. Therefore, a highly accurate and lightweight neural network is urgently needed for agricultural production.

In this paper, we proposed a lightweight attention-based encoder-decoder framework for crop identification, and summarize the contributions of this paper as follows:

- Designed a lightweight network structure with much smaller parameters and floating-point of operations than DeepLab v3+ and other classical networks.
- Got an excellent identification accuracy, which can reach more than 98% accuracy in large-scale plots and more than 97% accuracy in small-scale plots.
- We also built two datasets corresponding to regular large-scale plots and irregular small-scale plots to test the performance of the Deep-agirNet in different environments.

2 Materials and methods

2.1 Study area

As the most important winter crop in China, especially in the Yangtze River basin, winter wheat and canola have similar planting cycles, generally sown in September to October and harvested in April to May of the following year. Given that winter wheat and winter canola are important components of the agricultural economy, it is significant to know the distribution of these two crops for agricultural production and policy making. In this paper, two representative regions in the middle and lower reaches of the Yangtze River in China are selected as study areas, and their geographical locations are shown in Figure 1:

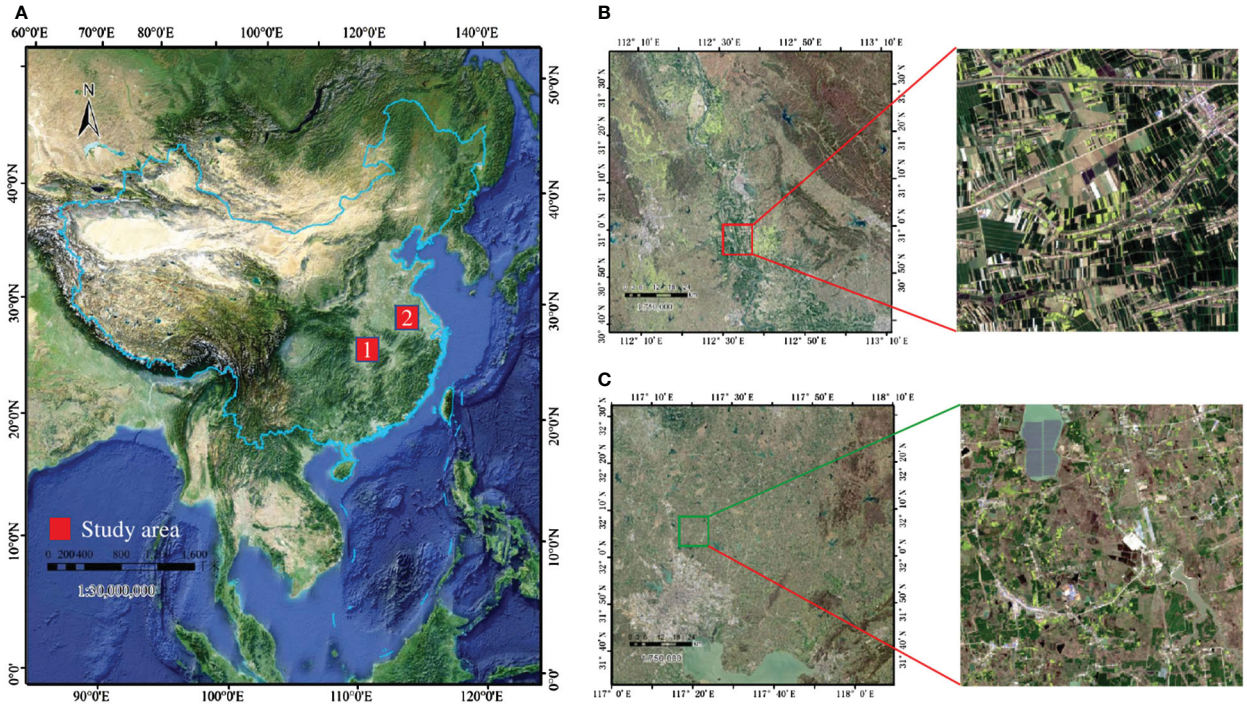


FIGURE 1
(A) Locations of the study areas in China (red), **(B)** Sentinel-2 image of the first study region, **(C)** Sentinel-2 image of the second study region.The red hollow rectangle shows the crop cultivation in the two regions.

The first rectangular study area T49RFQ is bounded by longitudes 112°0′ to 113°10′ E and latitudes 30°40′ to 31°40′ N which mainly belongs to Jingmen City, Hubei Province. Jingmen City is planted on a large scale with regular and continuously distributed plots, which makes it easy to plant and harvest. The second rectangular study area, T50SNA, is bounded by longitudes 117°0′ to 112°10′ E and latitudes 31°30′ to 32°30′ N which mainly belongs to Hefei, Anhui Province. Hefei is planted discretely, with small and scattered plot sizes and low land use.

2.2 Remote sensing images processing

Sentinel 2 is a high-resolution multispectral imaging satellite built by the European Space Corporation and consisted of the "twin" satellites Sentinel 2A and Sentinel 2B. The remote sensing images taken by the Sentinel satellites contain 13 bands with different spatial resolutions (10m, 20m, 60m). In this study, all bands except Band 1 (Coastal aerosol), Band 9 (Water vapor), and Band 10 (SWIR-cirrus), which have the lowest spatial resolution, were screened and excluded, and a bilinear interpolation algorithm was applied to Band 5, 6, 7 (Red edge), Band 8b (Narrow NIR) and Band 11, 12 (SWIR) are resampled to a spatial resolution of 10m, and then these bands are fused to obtain a 10-channel remote sensing image with 10m spatial resolution.

Since crops of different planting scales will have different Digital Number (DN) distributions in remote sensing images, as shown in Figure 2. To reduce the errors caused by DN, percentage linear stretching is adopted in this study for each band of T50SNA:

$$result = \frac{DN - min_{in}}{max_{in} - min_{in}} \times (max_{out} - min_{out}) + min_{out} \quad (1)$$

here max_{in} and min_{in} represent the maximum and minimum of DN of the stretched image, then max_{out} and min_{out} represent the set maximum and minimum of DN, respectively. Specifically, the maximum and minimum values of DN of T49RFQ are setas the upper and lower limits of pixel values. Then, the other image is linearly stretched to the set range so that the DN of the two images is finally distributed in the same range to reduce the deviation.

2.3 A lightweight encoder-decoder network based on DeepLab v3+

Since the first time used fully convolutional neural networks (Long et al., 2015) for end-to-end segmentation of natural images, semantic segmentation tasks for pixel-level classification have achieved leap-forward development. The vast majority of state-of-the-art (SOTA) segmentation networks, such as U-Net (Ronneberger et al., 2015), PSPNet (Zhao et al., 2017), DeepLab (Chen et al., 2017), and HRNetv2 (Sun et al., 2019a) are built based on encoder-decoder architecture. As an excellent semantic segmentation model with the encoder-decoder structure, DeepLab v3+ (Chen et al., 2018) is widely used in the field of semantic segmentation of remote sensing images. The part of the decoder includes Atrous Spatial Pyramid Pooling (ASPP) and an improved Xception module, where the ASPP module can control the size of the perceptual field by adjusting the expansion coefficient to capture

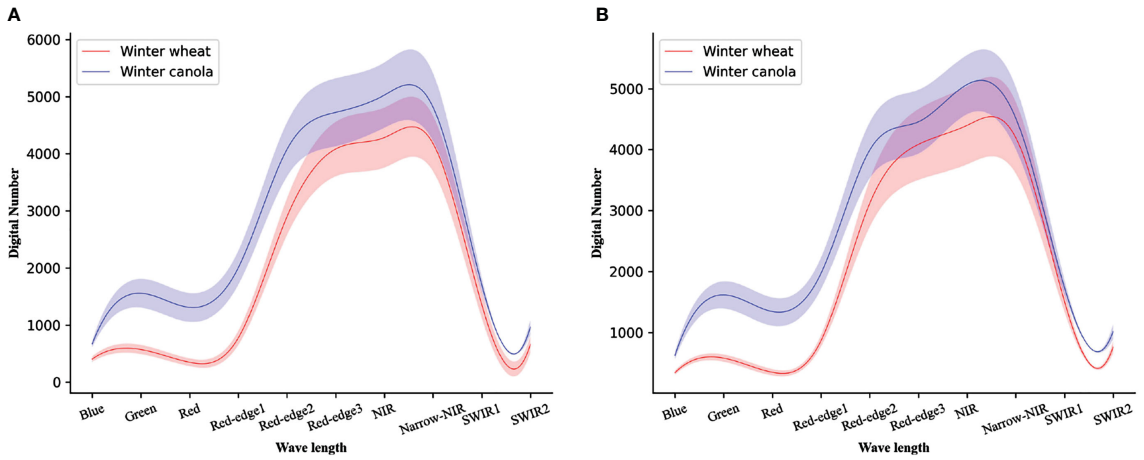


FIGURE 2
The Digital Number distribution plot of a multispectral of **(A)** the first study region of T49RFQ, **(B)** the second study region of T50SNA. Solid lines represent means and shaded areas represent one standard deviation from the mean.

the features at different scales. Then, two quadruple upsampling are used in the decoder part, where the first upsampling concatenates the low-dimensional features from the decoder and encoder to make features fusing, and the second upsampling restores the concatenated result to the same scale of inputs and classifies each pixel to obtain the segmentation result finally.

Despite the excellent performance of DeepLab v3+, it is hard to accept for agricultural production due to the large parameters. Considering a network serving agricultural production must balance accuracy and parameters, we made a lightweight improvement to DeepLab v3+ and named the improved network as Deep-agriNet.

As shown in Figure 3, the improvement of the network is mainly reflected in the following parts: In the encoder part, we chose ShuffleNet v2 (Ma et al., 2018), an advanced lightweight network architecture, as the feature extractor of Deep-agriNet. The design of ShuffleNet v2 is based on four network design criteria:

- Keeping the numbers of input and output channels equal minimizes memory access cost.
- A large group number used in group convolution increases computational cost.
- Complex network structure (abuse of branches and basic units) reduces the degree of network parallelism.
- The costs of element-wise operations cannot be neglected either.

The operation of channel split was used in the basic shuffle unit of ShuffleNet v2. Then, it divided input channels evenly into two branches to replace the group convolution. As shown in Figure 4A, one branch of the basic shuffle unit does nothing to reduce network computation, and the other branch maintains the same number of channels in each convolution. The shuffle unit for spatial down sampling, as shown in Figure 4B, removes the channel split and doubles the number of output channels compared to the input

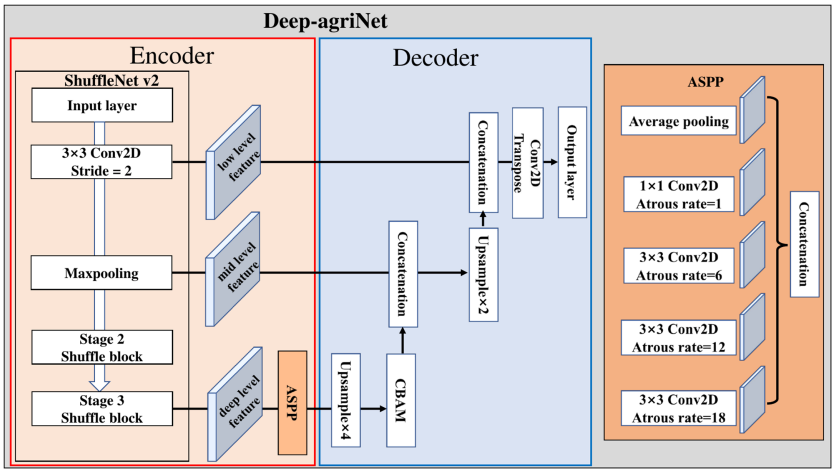


FIGURE 3
The framework of the Deep-agriNet.

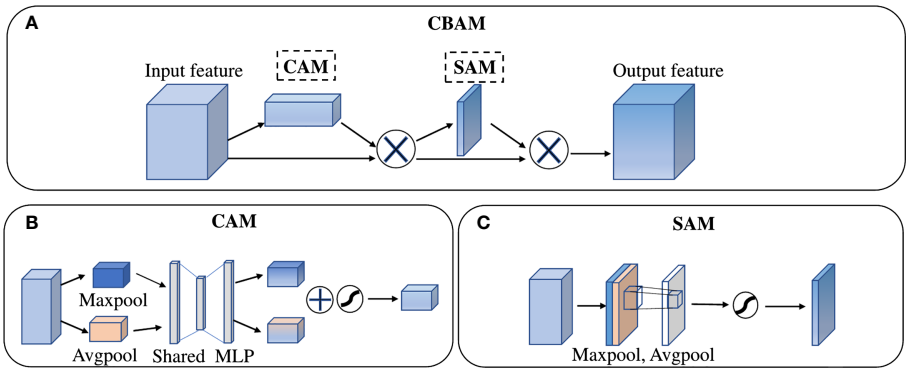


FIGURE 4
The structure of **(A)** Convolutional Block Attention Module, **(B)** Channel Attention Module, **(C)** Spatial Attention Module.

channels. In addition, the outputs of both shuffle unit are no longer an add operation between elements but a concatenation, which can fuse the extracted features or output information instead of simple superposition. Finally, the results of concatenation are shuffled at the end of the basic unit by the channel shuffle operation to increase the information exchange between channels, thus improving the network performance.

In the decoder part, a Convolutional Block Attention Module (CBAM) (Woo et al., 2018) was added to the DeepLab v3+ decoding module. As a "plug-and-play" lightweight convolutional attention module, CBAM is composed of a Channel Attention Module (CAM) and a Spatial Attention Module (SAM) in series, as shown in the Figure 5.

In the channel attention module, both the operation of average-pooling and max-pooling are used simultaneously to generate average-pooled features and max-pooled features. Then, the two kinds of features are forwarded to a Multiple-layer Perceptron (MLP) to share feature. The output features of MLP are merged by element-wise summation and then activated by a sigmoid function to generate the channel attention feature maps $M_c(F)$. In short, the detailed operation to obtain channel attention can be computed as:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (2)$$

In the spatial attention module, the output features of CAM are taken as the input feature of SAM. Firstly, twice operations of pooling based on channels are used to aggregate channel information and generate average-pooled features and max-pooled features. Then, these features are concatenated and convolved by a standard convolution layer and produce the spatial attention feature map $M_s(F)$. In short, the detailed operation to obtain spatial attention can be computed as:

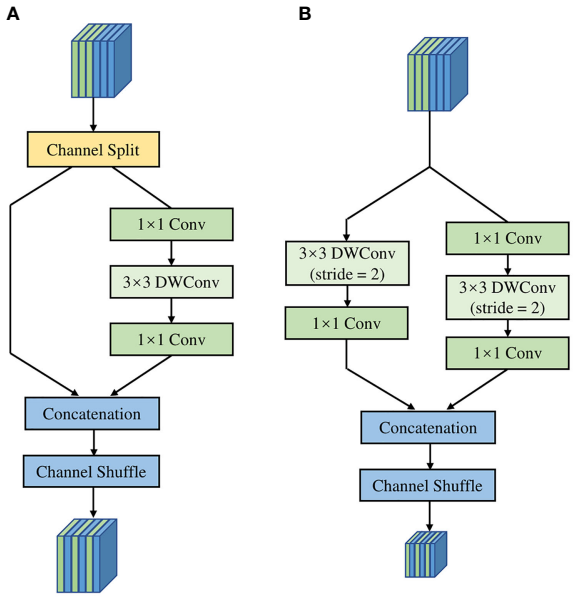


FIGURE 5
The structure of **(A)** basic unit, **(B)** unit for spatial down sampling.

$$M_c(F) = \sigma(f^{7\times7}[AvgPool(F); MaxPool(F)]) \quad (3)$$

Where σ is an activation function of sigmoid, F denotes the input feature, and f^{7x7} denotes a convolution operation with the filter size of 7 x 7.

In addition, some simple but effective adjustments are applied to the network improvement. Specifically: (a) The channels of input layer of the encoder were modified to ten layers because of the multispectral remote-sensing images containing more feature information than the traditional 3-channel RGB images (Zhao et al., 2022). (b) It is worth noting that the continuous large-scale upsampling is not conducive to obtain a satisfactory segmentation result, so we replaced the second 4-fold upsampling in thedecoder section with a 2-fold upsampling and a transpose convolution (Luo et al., 2021).

To extract richer multi-level features from the encoder, Deep-agriNet defined the low-level, mid-level and deep-level features in different scales to represent the extracted features by the Conv1, Maxpooling and Stage3 in the ShuffleNet v2, respectively. The specific process of this network could be described as below. Firstly, our network took 10-bands remote sensing images of 512 x 512 pixels as input and then processed by a 3 x 3 convolution layer with a stride of 2x 2 in the encoder module to obtain the low-level features of 256 x 256 pixels. These low-level features are then passed through a max-pooling layer to obtain mid-level features of 128 x 128 pixels. With the forward transmission of data, these intermediate features are down-sampled by multiple Shuffle blocks to obtain the deep-level features of 32 x 32 pixels. To obtain multi-scale fusion features, the deep-level features flow to an ASPP module of the decoder, where the input features are processed in parallel and concatenated by dilated convolutions with different dilation rates to capture the multi-scale information. Subsequently, the ASPP output information is passed through a single quadrupling up-sampling layer and a CBAM module in turn, resulting in 128*times*128 pixels feature with channel and spatial attention. These features are concatenated with the mid-level features in the encoder to reduce the loss of detail caused by multiple convolutions. After a 2-fold up-sampling, the concatenated features arerestored to 256 x 256 pixels and then concatenated with the low-level features in the encoder to fuse different level features from low to deep. Finally, the fused results are processed by transposed convolution to obtain the predicted image of pixel-level classification with 512 x 512 pixels.

2.4 Data acquisition

In this study, the remote sensing images of the T49RFQ area were cropped according to the size of 512 x 512 pixels, and 380 small-size patches were obtained in total. Since some patches did not contain crops or the cropped area is extremely tiny, we filtered and removed the data where the crop coverage was less than 30% of the total area. Finally, 100 patches were retained as dataset DS1. The same treatment was used on the T50SNA region to build dataset DS2, aiming to verify whether large-scale cropping affects the identification performance of the network. Then, we used the ArcMap program to annotate each pixel of the patches. When a pixel belongs to winter canola, its value is assigned as 1; when a pixel belongs to winter wheat, its value is assigned as 2, and in the rest cases, its value is assigned as 0. Finally, we obtained the same number of single-channel images corresponding to the patches on the dataset and used it as the annotation of the dataset. Prior to training, the dataset was split into training set and validation set randomly with a 7:3 ratio, which could reduce the imbalance of data and improve the network's generalization ability.

2.5 Model training

During the training process, Back Propagation (BP) algorithm (LeCun et al., 2015) and Adaptive Moment Estimation (Adam) (Kingma and Ba, 2014) algorithm were adopted to speed up and optimize the convergence rate of Deep-agriNet. Since Deep-agriNet is a multi-class crop identification network, the multi-class cross-entropy loss function was used to calculate the loss between the predicted result and the true value of each epoch:

$$Loss(y, \hat{y}) = -\sum_{i=1}^{N} y_i \log \hat{y_i} \quad (4)$$

where y_i is the true value of a category whose value is 0 or 1, $\hat{y}_i$ is the predicted probability of the category whose value is distributed between 0 and 1,and N represents the category contained by the sample. Usually, the Learning Rate decays gradually during training, so we adopted the Polynomial Learning Rate Policy (Mishra and Sarawadekar, 2019) to dynamically adjust the learning rate:

$$lr = base\_lr \times (1 - \frac{epoch}{max\_epoch})^{power} \quad (5)$$

where lr is the dynamic learning rate, *base_lr* is the baseline learning rate, *epoch* is the current number of iterations, *max_epoch* is the maximum number of iterations, and *power* is the power of the polynomial.

To prevent overfitting during training, we employed the operations of rotation, mirroring, and adding noise to augment the dataset to improve the generalization ability and robustness of the model. Meanwhile, the same operations were applied to the annotation as well.

2.6 Evaluation metrics

To evaluate the performance of Deep-agriNet, overall accuracy (OA), mean intersection over union (mIoU), and recall as evaluation metrics were used in this experiment. mIoU is one of the most basic metrics to evaluate the performance of semantic segmentation, and it represents the average of the ratio of the intersection and union of the predicted and true values for all classes:

$$mIoU = \frac{1}{N}\sum_{i=1}^{N}\frac{P_i \cap G_i}{P_i \cup G_i} \quad (6)$$

where N is all categories of the sample including background. P and G are predicted and true pixels of a sample, respectively. OA represents the proportion of correctly classified pixels to all sample

pixels. Recall represents the proportion of correctly classified pixels to all positive sample pixels:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$recall = \frac{TP}{TP + FN} \tag{8}$$

where *TP* is True Positive, indicating correct classification of pixels and positive predicted outcomes, *FP* is False Positive, meaning that the negative pixel is divided into positive samples, *TN* is True Negative, indicating the real background area is identified as the background area, and *FN* is False Negative, which represents the positive pixel is divided into negative samples.

Additionally, we introduced some metrics to evaluate the lightness of the network. The parameter is a commonly used evaluation metric for lightness of a network, which can measure the complexity of a model and the consumption of memory in computation. Theformula of parameter is shown as follow:

$$parameter = K \times K \times C_{in} \times C_{out} \tag{9}$$

where the K x K means the size of kernel, and the C_{in} and C_{out} represents the number of input channels and output channels, respectively. In addition, the FLOPs which stands for floating-point of operations is a measure of network complexity. The FLOPs can be computed as:

$$FLOPs = K \times K \times C_{in} \times H_{out} \times W_{out} \times C_{out} \tag{10}$$

where H_{out} and W_{out} represents the height and width of the output feature map. In this paper, we used the giga floating-point operations (GFLOPs, 10^9 x *FLOPs*) to measure the complexity of network.

2.7 Hyperparameters and environment setting

To obtain more effective hyperparameters, we set the base learning rate to 0.0005, 0.001, 0.005, and 0.01. and batch size to 4, 8, 16, and 32, respectively. After several training sessions, the best results were obtained when the base learning rate was 0.001 and the batch size was 4. This experiment was trained on the Linux platform, and the deep learning framework used was Google's open-source TensorFlow, and the GPU used for training was 24GB Nvidia GeForce GTX3090Ti.

3 Results

3.1 The comparison of lightweight between Deep-agriNet and other methods

To verify the effectiveness and superiority of the Deep-agriNet in terms of lightness, we calculated the parameter, GFLOPs and Inference Time(IT) for Deep-agriNet and other methods, and the results are shown in Table 1. From this table, it can be seen that HRNetv2 has the most parameters at 65.94M, while U-Net has the most GFLOPs at 450.64. In comparison, Deep-agriNet has significant advantages in evaluation metrics, parameters and GFLOPs, which are only 3.89M and 47.5. Moreover, the IT of Deep-agriNet is roughly comparable to that of U-Net (2.4 vs 1.8s).

TABLE 1 The lightweight metrics for Deep-agriNet and other methods.

Method	Backbone	Parameters(M)	GFLOPs	IT(s)
U-Net	VGG-16	24.9	450.64	**1.8**
PSPNet	ResNet-50	46.77	116.5	2.7
DeepLab v3+	Xception	54.2	103.16	3.1
HRNetv2	HRNetv2-W48	65.94	169.94	9.4
Deep-agriNet	ShuffleNetv2	**3.89**	**47.5**	2.4

The bold values indicate the highest scores in the experiments.

In addition, we adopted a scatter plot better visualize the trade-off between accuracy and complexity and clarify the superiority of the proposed model Deep-agriNet. In the scatter plot, the x-axis indicates OA and the y-axis indicates GFLOPs. As shown in Figure 6, Deep-agriNet expresses the higher accuracy and lower GFLOPs than the other benchmark methods which means more lightweight and accurate of the proposed network.

3.2 The performance of Deep-agriNet for crop identification

The loss function curve can reflect the robustness and accuracy of the network. The smoother the curve is, the better the robustness of the model, and the smaller the loss value, the higher the accuracy of the model. The annotated DS1 was fed into Deep-agriNet and other methods for training. After 50 epochs, the loss convergence curves are shown in Figure 7. It clearly demonstrates that the cross-entropy loss tends to decrease with increasing epochs. According to the results in Figure 7, it can be seen that the Deep-agriNet based on

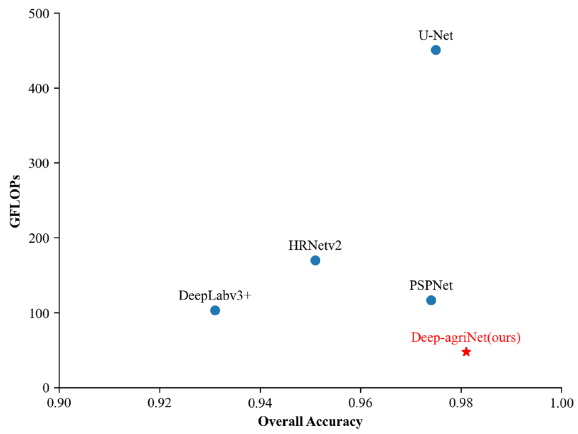


FIGURE 6
The scatter plot of accuracy and complexity of the networks on DS1.

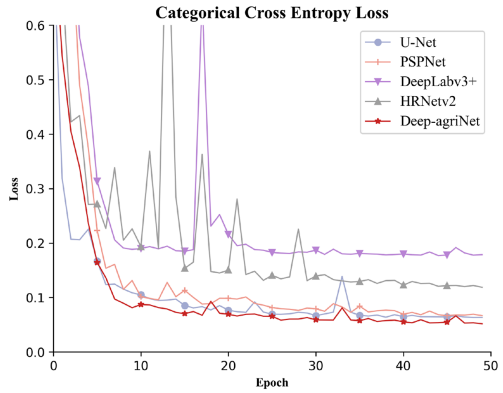


FIGURE 7
The convergence curves of Deep-agriNet and other methods.

DeepLab v3+ has more stable performance and higher accuracy in the training process.

To further analyze the performance of Deep-agriNet, this experiment compared these methods in terms of more evaluation metrics on accuracy, and the specific experimental results are shown in Table 2. From this table, it can be seen that Deep-agriNetperforms best in all aspects, where mIoU, OA, and recall is 0.972, 0.981, and 0.980. This results are significant improvement of 7.0%, 5.0% and 5.7% over the original DeepLab v3+, and slightly better than the next best performer U-Net by 0.8%, 0.6%,and 0.6%.

The Figure 8 shows the identification results of Deep-agriNet and other methods for winter wheat and winter canola on DS1, where the yellow markers represent the winter canola planting area, the green markers represent the winter wheat planting area, and the gray markers represent the background. To show the prediction results of different methods more clearly, the marked patch1, patch2 and patch3 in the figure are enlarged to observe the details of the images. Comparing the original images and theprediction of multiple methods, the crop planting areas identified by U-Net, PSPNet and Deep-agriNet are highly consistent with the original images. Both the paths in the fields and the edges of the plots can be predicted with clarity. In contrast, the identification results of DeepLab v3+ and HRNetv2 are much worse. In patch1, the majority of the roads are not identified, and in patch2 and patch3, almost all the plots have different degrees of missing boundaries.

To further demonstrate the predictive capability of Deep-agriNet on the area with irregular small-scale plots, we trained and validated it on DS2. As shown in Table 3, Deep-agriNet still hold the best results in the respect of mIoU, OA, and recall with 0.961, 0.974, and 0.973, respectively. Figure 9 shows the prediction results of different methods for winter wheat and winter canola on DS2. From the local magnification results of patch1, patch2 and patch3, Deep-agriNet still shows excellent performance on background identification and can predict the roads in the plots clearly. However, compared with the results on DS1, it can be obviously found that the model is less effective in irregular plots prediction and there is a slight phenomenon of boundaries missing. Comparing the identification results in Figures 8, 9, Deep-agriNet performs better for crop identification with different planting scales.

TABLE 2 Comparison of the identification results of different methods on DS1.

Method	mIoU	OA	Recall
U-Net	0.964	0.975	0.974
PSPNet	0.962	0.974	0.973
DeepLab v3+	0.902	0.931	0.923
HRNetv2	0.930	0.951	0.949
Deep-agriNet	**0.972**	**0.981**	**0.980**

The bold values indicate the highest scores in the experiments.

3.3 Ablation study

To validate the role of CBAM in this network, The Deep-agriNet without CBAM was used as the baseline, and the two networks were trained with the same hyperparameters such as baseline learning rate and training epochs. As shown in Table 4, the network with CBAM is slightly improved in all aspects, including 0.8%, 0.6% and 0.6% for mIoU, OA and recall respectively on DS1. In addition, the parameters of the network only increased by 0.2M after adding CBAM. Figure 10 shows the identification results of winter wheat and winter canola on DS1 before and after adding CBAM. As shown in the figure, CBAM was able to focus attention on the areas where winter wheat and winter canola were mixed and clearly identified plots of several pixel widths. In contrast, the network without CBAM was only able to identify fuzzy outlines but was unable to identify cross-planted plots.

4 Discussion

4.1 Effects comparison between DeepLab v3+ and improved network

In this study, DeepLab v3+ was used as the base crop identification network, and a series of improvements were made on its basis. Finally, Deep-agriNet, the improved network, was applied to spring crop identification. Firstly, the backbone of DeepLab v3+, Xception, was replaced with ShuffleNet v2, a more advanced feature extractor. Based on this improvement, the identification accuracy was significantly improved, with mIoU, OA and recall improving by 6.3%, 4.5% and 5.1% on DS1. Meanwhile, the number of parameters and GFLOPs were also greatly optimized, much smaller than DeepLab v3+ and other methods. These performance improvements are mainly attributed to the following two factors: (a) the channel split method proposed by ShuffleNet v2 makes the input channels to be divided into two, with one part being passed down directly and the other part participating in the convolution operation, and finally the two parts are reassembled to reuse features. (b) ShuffleNet v2 transforms the elementwise add operation in depthwise convolution into a concatenation operation and replaces the grouped convolution with the ordinary convolution to greatly

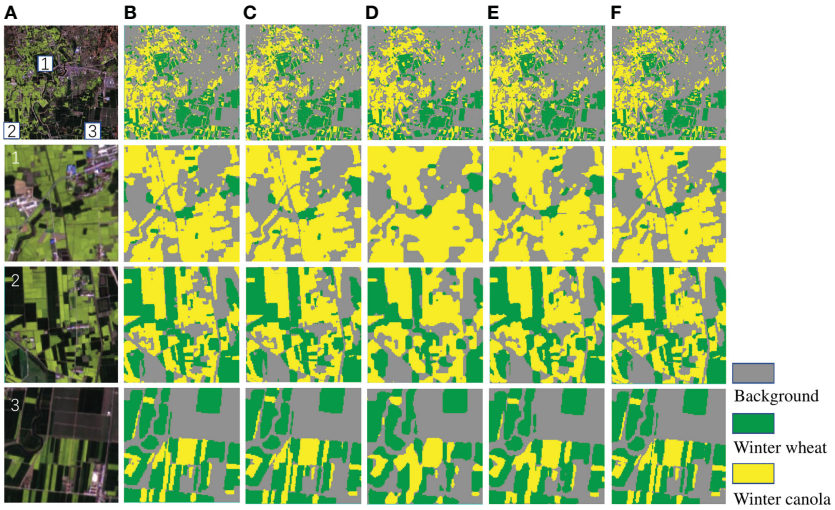


FIGURE 8
The original images and clipped regions of the experimental area on DS1 and the prediction results of different methods. **(A)** original image, **(B)** prediction result of U-Net, **(C)** prediction result of PSPNet, **(D)** prediction result of DeepLab v3+, **(E)** prediction result of HRNetv2, **(F)** prediction result of Deep-agriNet.

TABLE 3 Comparison of the identification results of different methods on DS2.

Method	mIoU	OA	Recall
U-Net	0.956	0.967	0.963
PSPNet	0.954	0.965	0.965
DeepLab v3+	0.907	0.935	0.929
HRNetv2	0.921	0.945	0.943
Deep-agriNet	**0.961**	**0.974**	**0.973**

The bold values indicate the highest scores in the experiments.

TABLE 4 Comparison of the identification results of Deep-agriNet before and after adding CBAM.

Method	mIoU	OA	Recall	Parameters
Deep-agriNet without CBAM	0.964	0.975	0.974	3.87
Deep-agriNet	0.972	0.981	0.980	3.89

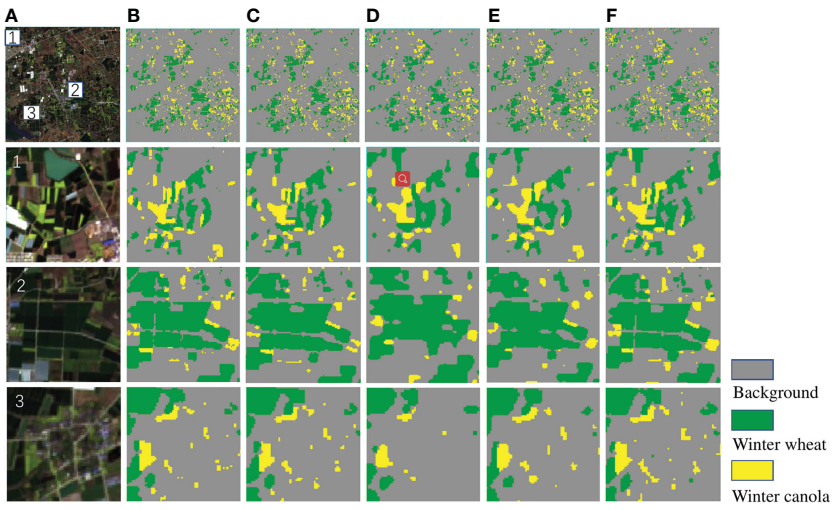


FIGURE 9
The original images and clipped regions of the experimental area on DS2 and the prediction results of different methods. **(A)** original image, **(B)** prediction result of U-Net, **(C)** prediction result of PSPNet, **(D)** prediction result of DeepLab v3+, **(E)** prediction result of HRNetv2, **(F)** prediction result of Deep-agriNet.

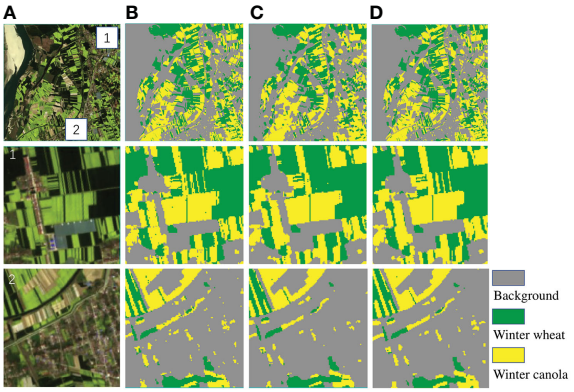


FIGURE 10
The original images and clipped regions of the experimental area on DS1 and the prediction results of the network with CBAM. **(A)** original image, **(B)** labeled image, **(C)** Deep-agriNet without CBAM, **(D)** Deep-agriNet (with CBAM).

reduce the amount of computation. Then, A CBAM module was added between the encoder module and the decoder module in this study. Based on this improvement, the performance of mIoU, OA and recall has been improved slightly at the cost of a small computational cost. Specifically, on DS1, mIoU, OA, and recall are improved by 0.7%, 0.5%, and 0.6%, respectively, and the number of parameters hardly increased. As shown in Figure 10, although the improvement of evaluation metrics is quite small, the identification performance is improved considerably, and the phenomenon of missing edges and misidentified mixed-species regions is significantly improved compared with that before the improvement.

4.2 Result analysis of different areas

To investigate whether large-scale planting will affect the network performance, two regions, T49RFQ and T50SNA, were selected for this experiment, and Deep-agriNet was used to train DS1 and DS2 corresponding to the two regions. As shown in the Tables 2, 3, Deep-agriNet has a better identification effect on the T49RFQ region of large-scale planting. Compared with the training results of DS2, the mIoU, OA and recall of DS1 increased by 0.9%, 0.7% and 0.7%, respectively. The author believes that the attention mechanism can capture the context dependence, and the data in DS1 has stronger spatial continuity. Even after multiple feature extraction, there is still a strong context dependence, which is beneficial to the decoder to infer the category of surrounding pixels through this dependence, and therefore the identification effect is improved.

5 Conclusions

In this paper, we proposed an improved lightweight network based on DeepLab v3+ and apply the network to spring crop identification. An advanced feature extractor, ShuffleNet v2, was used in this network to replace the backbone of DeepLab v3+. In addition, a CBAM combining channel and spatial attention mechanisms was added at the end of the encoder. In the decoder part of the original network, a 4-fold upsampling was modified to two adjacent 2-fold upsampling. To verify the performance of Deep-agriNet, two datasets with different planting scales were constructed for experiments. The experimental results show that the Deep-agriNet exhibits better performance on both datasets, and the parameters of the Deep-agriNet are only one-fourteenth of the original network. The Deep-agriNet can be applied not only for spring crop identification but also extended to other agricultural projects, such as crop yield prediction or crop disaster detection. However, to achieve this goal, further research on related work is needed to improve the algorithm so that the quantification of crop acreage can be achieved. In future work, we will try to use more advanced networks and larger agricultural datasets to meet more kinds of crop precision identification needs, and strive toserve our research results more effectively in the agricultural field.

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

YH, AM, LZ, and TX conceived the idea and designed the network. AM, LZ, ZJ, and YH contributed to collecting the dataset. YH and AM wrote the code, validated the method, and wrote the paper. TX, ZJ and YW revised the paper. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Key Research and Development Program of China [2021YFD2000205].

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

References

Becker-Reshef, I., Justice, C., Sullivan, M., Vermote, E., Tucker, C., Anyamba, A., et al. (2010). Monitoring global croplands with coarse resolution earth observations: The global agriculture monitoring (glam) project. *Remote Sens* 2 (6), 1589–609. doi: 10.3390/rs2061589

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). DeepLab: Semantic image segmentation with deep convolutional nets and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (4), 834–848. doi: 10.1109/TPAMI.2017.2699184

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*. Munich, Germany: Springer, 801–818.

Han, L., Yang, G., Yang, X., Song, X., Xu, B., Li, Z., et al. (2022). An explainable xgboost model improved by smote-enn technique for maize lodging detection based on multi-source unmanned aerial vehicle images. *Comput. Electron. Agric.* 194, 106804. doi: 10.1016/j.compag.2022.106804

Jia, X., Khandelwal, A., Mulla, D. J., Pardey, P. G., and Kumar, V. (2019). Bringing automated, remote-sensed, machine learning methods to monitoring crop landscapes at scale. *Agric. Economics* 50, 41–50. doi: 10.1111/agec.12531

Jin, X.-B., Zheng, W.-Z., Kong, J.-L., Wang, X.-Y., Zuo, M., Zhang, Q.-C., et al. (2021). Deep-learning temporal predictor *via* bidirectional self-attentive encoder–decoder framework for iot-based environmental sensing in intelligent greenhouse. *Agriculture* 11, 802. doi: 10.3390/agriculture11080802

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv*. 1412.6980. doi: 10.48550/arXiv.1412.6980

Kussul, N., Lavreniuk, M., Skakun, S., and Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sens. Lett.* 14, 778–782. doi: 10.1109/LGRS.2017.2681128

Kuwata, K., and Shibasaki, R. (2015). "Estimating crop yields with deep learning and remotely sensed data," in *2015 IEEE international geoscience and remote sensing symposium (IGARSS)* (Milan, Italy: IEEE), 858–861.

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

Li, Q., Wang, C., Zhang, B., and Lu, L. (2015). Object-based crop classification with landsat-modis enhanced time-series data. *Remote Sens.* 7, 16091–16107. doi: 10.3390/rs71215820

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. Boston, MA, USA: IEEE, 3431–3440.

Longato, D., Gaglio, M., Boschetti, M., and Gissi, E. (2019). Bioenergy and ecosystem services trade-offs and synergies in marginal agricultural lands: A remote-sensing-based assessment method. *J. Cleaner Production* 237, 117672. doi: 10.1016/j.jclepro.2019.117672

Lu, R., Wang, N., Zhang, Y., Lin, Y., Wu, W., and Shi, Z. (2022). Extraction of agricultural fields *via* dasfnet with dual attention mechanism and multi-scale feature fusion in south xinjiang, china. *Remote Sens.* 14, 2253. doi: 10.3390/rs14092253

Luo, X., Tong, X., and Hu, Z. (2021). An applicable and automatic method for earth surface water mapping based on multispectral images. *Int. J. Appl. Earth Observation Geoinformation* 103, 102472. doi: 10.1016/j.jag.2021.102472

Ma, N., Zhang, X., Zheng, H.-T., and Sun, J. (2018). "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European conference on computer vision (ECCV)*. Munich, Germany: Springer, 116–131.

Mishra, P., and Sarawadekar, K. (2019). "Polynomial learning rate policy with warm restart for deep neural network," in *TENCON 2019-2019 IEEE region 10 conference (TENCON)* (Kochi, India: IEEE), 2087–2092.

Qiong, H., Wu, W.-b., Qian, S., Miao, L., Di, C., Yu, Q.-y., et al. (2017). How do temporal and spectral features matter in crop classification in heilongjiang province, china? *J. Integr. Agric.* 16, 324–336. doi: 10.1016/S2095-3119(15)61321-1

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional networks for biomedical image segmentation," in *International conference on medical image computing and computer-assisted intervention* (Munich, Germany: Springer), 234–241.

Shi, C., Zhang, J., and Teng, G. (2019). Mobile measuring system based on labview for pig body components estimation in a large-scale farm. *Comput. Electron. Agric.* 156, 399–405. doi: 10.1016/j.compag.2018.11.042

Sun, Y., Luo, J., Wu, T., Zhou, Y., Liu, H., Gao, L., et al. (2019b). Synchronous response analysis of features for remote sensing crop classification based on optical and sar time-series data. *Sensors* 19, 4227. doi: 10.3390/s19194227

Sun, K., Zhao, Y., Jiang, B., Cheng, T., Xiao, B., Liu, D., et al. (2019a). High-resolution representations for labeling pixels and regions. *arXiv preprint arXiv*. 1904.04514. doi: 10.48550/arXiv.1904.04514

Tan, C., Zhang, P., Zhang, Y., Zhou, X., Wang, Z., Du, Y., et al. (2020). Rapid recognition of field-grown wheat spikes based on a superpixel segmentation algorithm using digital images. *Front. Plant Sci.* 11, 259. doi: 10.3389/fpls.2020.00259

Tian, H., Wang, Y., Chen, T., Zhang, L., and Qin, Y. (2021a). Early-season mapping of winter crops using sentinel-2 optical imagery. *Remote Sens.* 13, 3822. doi: 10.3390/rs13193822

Tian, Y., Yang, C., Huang, W., Tang, J., Li, X., and Zhang, Q. (2021b). Machine learning-based crop recognition from aerial remote sensing imagery. *Front. Earth Sci.* 15, 54–69. doi: 10.1007/s11707-020-0861-x

Waldhoff, G., Lussem, U., and Bareth, G. (2017). Multi-data approach for remote sensing-based regional crop rotation mapping: A case study for the rur catchment, germany. *Int. J. Appl. Earth observation geoinformation* 61, 55–69. doi: 10.1016/j.jag.2017.04.009

Wang, S., Chen, W., Xie, S. M., Azzari, G., and Lobell, D. B. (2020). Weakly supervised deep learning for segmentation of remote sensing imagery. *Remote Sens.* 12, 207. doi: 10.3390/rs12020207

Wang, H., Chen, X., Zhang, T., Xu, Z., and Li, J. (2022). Cctnet: Coupled cnn and transformer network for crop segmentation of remote sensing images. *Remote Sens.* 14, 1956. doi: 10.3390/rs14091956

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*. Munich, Germany: Springer, 3–19.

Xu, H., Wang, Y., Guan, H., Shi, T., and Hu, X. (2019). Detecting ecological changes with a remote sensing based ecological index (rsei) produced time series and change vector analysis. *Remote Sens.* 11, 2345. doi: 10.3390/rs11202345

Yu, J., Cheng, T., Cai, N., Lin, F., Zhou, X.-G., Du, S., et al. (2022). Wheat lodging extraction using improved_unet network. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1009835

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., et al. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* 241, 111716. doi: 10.1016/j.rse.2020.111716

Zhang, J., He, Y., Yuan, L., Liu, P., Zhou, X., and Huang, Y. (2019). Machine learning-based spectral library for crop classification and status monitoring. *Agronomy* 9, 496. doi: 10.3390/agronomy9090496

Zhang, X., Sun, Y., Shang, K., Zhang, L., and Wang, S. (2016). Crop classification based on feature band set construction and object-oriented approach using hyperspectral images. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 9, 4117–4128. doi: 10.1109/JSTARS.2016.2577339

Zhao, J., Kumar, A., Banoth, B. N., Marathi, B., Rajalakshmi, P., Rewald, B., et al. (2022). Deep-learning-based multispectral image reconstruction from single natural color rgb image–enhancing uav-based phenotyping. *Remote Sens.* 14, 1272. doi: 10.3390/rs14051272

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. Honolulu, HI, USA: IEEE, 2881–2890.

Zheng, B., Myint, S. W., Thenkabail, P. S., and Aggarwal, R. M. (2015). A support vector machine to identify irrigated crop types using time-series landsat ndvi data. *Int. J. Appl. Earth Observation Geoinformation* 34, 103–112. doi: 10.1016/j.jag.2014.07.002

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 05 June 2023
DOI 10.3389/fpls.2023.1121012

OPEN ACCESS

EDITED BY
Baohua Zhang,
Nanjing Agricultural University, China

REVIEWED BY
Ali Parsaeimehr,
Delaware State University, United States
Jakub Nalepa,
Silesian University of Technology, Poland

*CORRESPONDENCE
Dmitry Kurouski
dkurouski@tamu.edu

RECEIVED 10 December 2022
ACCEPTED 18 May 2023
PUBLISHED 05 June 2023

CITATION
Farber C, Shires M, Ueckert J, Ong K and Kurouski D (2023) Detection and differentiation of herbicide stresses in roses by Raman spectroscopy.
Front. Plant Sci. 14:1121012.
doi: 10.3389/fpls.2023.1121012

Detection and differentiation of herbicide stresses in roses by Raman spectroscopy

Charles Farber[1], Madalyn Shires[2], Jake Ueckert[2], Kevin Ong[2] and Dmitry Kurouski[1,3,4]*

[1]Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX, United States, [2]Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX, United States, [3]Department of Biomedical Engineering, Texas A&M University, College Station, TX, United States, [4]Department of Molecular and Environmental Plant Science, Texas A&M University, College Station, TX, United States

Herbicide application is a critical component of modern horticulture. Misuse of herbicides can result in damage to economically important plants. Currently, such damage can be detected only at symptomatic stages by subjective visual inspection of plants, which requires substantial biological expertise. In this study, we investigated the potential of Raman spectroscopy (RS), a modern analytical technique that allows sensing of plant health, for pre-symptomatic diagnostics of herbicide stresses. Using roses as a model plant system, we investigated the extent to which stresses caused by Roundup (Glyphosate) and Weed-B-Gon (2, 4-D, Dicamba and Mecoprop-p (WBG), two of the most commonly used herbicides world-wide, can be diagnosed at pre- and symptomatic stages. We found that spectroscopic analysis of rose leaves enables ~90% accurate detection of Roundup- and WBG-induced stresses one day after application of these herbicides on plants. Our results also show that the accuracy of diagnostics of both herbicides at seven days reaches 100%. Furthermore, we show that RS enables highly accurate differentiation between the stresses induced by Roundup- and WBG. We infer that this sensitivity and specificity arises from the differences in biochemical changes in plants that are induced by both herbicides. These findings suggest that RS can be used for a non-destructive surveillance of plant health to detect and identify herbicide-induced stresses in plants.

KEYWORDS

Raman spectroscopy, herbicides, roses, PLS-DA, glyphosate, Weed-B-Gon

Highlights

- We report an innovative laser-based approach for non-invasive diagnostics of herbicide damage in plants.
- Our Raman-based technique enables ~90% accurate detection of Roundup- and WBG-induced stresses one day after application of these herbicides on plants.
- Our results also show that the accuracy of diagnostics of both herbicides at seven days reaches 100%.

Introduction

Application of herbicides is a cornerstone of modern horticulture. However, many ornamental plant species are sensitive to herbicides and are commonly damaged from non-target exposure by mechanisms such as drift (Henry et al., 2004; Mehdizadeh et al., 2021). Moreover, herbicide-induced damages in some plant species, such as roses (*Rosa* spp.), have visual similarities with the symptoms induced by biotic stresses. For instance, chlorosis of leaves and shortened internodes caused by herbicide application on roses is often misdiagnosed as Rose Rosette Disease (RRD), a devastating disease that affects plants in Europe and the U.S (Farber et al., 2019a). Substantial biological and horticultural expertise is required to detect and identify symptoms of herbicide exposure; such expertise typically requires years of experience and training. Therefore, it becomes very important to develop a confirmatory analytical method that can be used to detect symptoms of herbicide exposure, as well as disentangle biotic and abiotic stresses in ornamental plants (Tataridas et al., 2022).

Roundup and WBG are some of the most used herbicides worldwide. In the US alone, between 1974 and 2014, the use of herbicides with glyphosate as their active ingredient, such as Roundup, increased almost 200-fold, from 635,000 kg to 125 billion kg (Benbrook, 2016). Glyphosate acts by inhibiting 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase, a key enzyme in the shikimate pathway, which is responsible for synthesis of aromatic amino acids in plants (Sherwani et al., 2015). The effect of glyphosate is primarily evident at the sites of new growth (root and shoot meristem). Symptoms of glyphosate exposure, such as chlorosis and necrosis of tip tissues, can be observed across the entire plant at about 5 to 10 days after the herbicide application. Because products of the shikimate pathway are closely associated with a host defense, glyphosate also increases plant susceptibility to pathogens (Hammerschmidt, 2018)

While glyphosate has generic targeting, WBG is highly specific for dicots. WBG is a formulation of three different synthetic auxins: 2,4-dichlorophenoxyacetic acid (2,4-D), mercoprop-p acid (MCPP), and dicamba acid. Auxins are a family of plant hormones which are associated with regulating plant growth (Teale et al., 2006). The mechanisms of auxinic herbicides are not well understood (Kelley and Riechers, 2007; Mithila et al., 2011). Previous studies of these herbicides' modes of action found that they trigger buildup of abscisic acid and ethylene, which enables accumulation of hydrogen peroxide, and subsequently, reactive oxygen species (ROS) (Romero-Puertas et al., 2004). Consequently, ROS are closely associated with the activity of WBG in plants (Gleason et al., 2011). A growing body of evidence suggests that several methods could be used detect herbicide stresses in plants, including hyperspectral imaging (Henry et al., 2017; Lu et al., 2020; Yan et al., 2021).

Our group previously showed that Raman spectroscopy (RS), a non-destructive, non-invasive analytical method that reveals the chemical structure of analyzed samples, can be used to detect and identify biotic and abiotic stresses in plants (Mandrile et al., 2019; Farber et al., 2020c; Gupta et al., 2020; Farber et al., 2021; Payne and Kurouski, 2021). The efficacy of this method has been demonstrated for the rapid detection of: viral diseases in roses, tomatoes, wheat, and ornamental shrubs; fungal diseases in corn, wheat, and sorghum; and bacterial diseases of orange (Yeturu et al., 2016; Egging et al., 2018; Farber and Kurouski, 2018; Farber et al., 2019a; Mandrile et al., 2019; Sanchez et al., 2019b; Farber et al., 2020a). Additionally, our group has demonstrated RS-based detection of insect larva developing within beans (Sanchez et al., 2019a) and the detection of zebra chip disease in potato (Farber et al., 2021). It was also reported that RS could be used for pre-symptomatic diagnostics of nutritional deficiencies in rise caused by the lack of nitrogen, phosphorus, and potassium (Sanchez et al., 2020).

Expanding upon this, we investigated the accuracy of RS-based confirmatory diagnostics of herbicide stresses in the 'Pink Double Knock Out©' roses that are caused by Roundup and WBG herbicides. Our findings suggest that RS can be used for a fast and accurate diagnostics of herbicide stresses in plants prior to the symptom development. Timely detection of such stresses can be used for timely elimination of the herbicides or adjustment of their concentrations.

Materials and methods

Plants

Twenty-one plants of the rose cultivar 'Pink Double Knock Out©' were used in this experiment. Plants were received as small tissue culture plants and were acclimated indoors for one month. After plants were actively growing, they were placed in a greenhouse to continue growing for 2 months before the experiment. The substrate used in the greenhouse was Jolly Gardener® Pro-Line C/20, which consisted of 80% Canadian Sphagnum peat moss and 20% coarse perlite. We used neither fertilization nor pesticides before or during the experiments. The temperatures in the greenhouse were approximately 29°C during the day (6 am to 6 pm) and 21°C night (6 pm to 6 am). All plants involved in the experiment were screened for rose rosette virus (RRV), as well as for other rose viruses utilizing National Clean Plant Network-Rose screening protocols. All plants were free of the targeted pathogens (Farber et al., 2019a). Seven plants of uniform size and health were selected as replicates for each of the herbicide treatments.

Herbicide treatments

The herbicides used for this experiment were Roundup and WBG; reverse osmosis (RO) water was used as the negative control application. The two herbicides were chosen due to the prevalence of those brands being used on lawns and in landscapes. To simulate herbicide drift damage, all herbicides were diluted to 1/10th of the Ready To Spray (RTS) amounts. This concentration was chosen to model appearance of symptoms that visually resembled RRD since such herbicide-induced stresses are often misdiagnosed as RRD in roses. Herbicide calculations were done based on a six-inch pot size. Roundup and WBG were diluted from the RTS at a rate of 1-part spray to 9 parts water. Application was performed by soaking all

leaves to the point upon which solution droplets appeared on the plant surfaces.

Raman spectroscopy

For each measurement, thirty leaflets were sampled randomly from each group of plants with an average of four leaves per plant. Leaflets were collected from both new, fully emerged leaves and mature leaves to represent the entire plant canopy. We avoided collection of leaflets that had any mechanical damage, signs of wilting, or leaves exhibiting visual signs of herbicide damage to demonstrate robustness of this sensing approach. All leaflets looked healthy without any changes in coloration or chlorosis symptoms. Sampling time intervals were: 1, 7, 14, and 30 day post-treatment. The experiment was repeated twice.

Raman spectra were acquired from leaves using an Agilent Resolve spectrometer equipped with an 830 nm source with a spectral resolution of 15 cm^{-1}. Laser beam size was around 2 mm. For each measurement, the spectrometer was positioned next to the leaf surface. On average 2-3 spectra were collected form one leaf. All scans were taken in the 'surface' mode with a 1-second integration time and 490 mW of power. We found that as the experiment proceeded, the overall intensity of spectra acquired from herbicide-treated plants was much lower than that of the control plants. To address this, we normalized our spectra to the 1440 cm^{-1} peak. All spectral interpretation is based on these normalized spectra. All acquired spectra were baselined by the Agilent Resolve spectrometer. For the data analysis, Raman spectra were extracted using Agilent Resolve software and treated using MATLAB equipped with PLS-Toolbox ((Eigenvector Research Inc.). Reference Raman spectra of round-up and WBG are shown in the Figure S1.

Statistical analysis

After spectra were imported into MATLAB and assigned a class based on their health or herbicide status, Partial Least Squares Discriminant Analysis (PLS-DA) was conducted to differentiate the spectra based on spectral changes associated with their experimenter-assigned classes. Spectra were first split into calibration (66%) and validation (34%) sets using the Kennard-Stone method before building all models, unless otherwise noted (Kennard and Stone, 1969). Spectra were normalized to a total spectral area of 1 then mean centered. All tables reporting PLS-DA results are for the validation of these models. All datasets used for model calibration and validation in this study are summarized in Tables S1–S3.

Results

Description of visual symptoms

At day 7 post-treatment, we observed proliferation of small shoots on Glyphosate-treated plants (Figure 1) (Karlik and Flint). These plants began to exhibit leaf chlorosis at day 30 post-treatment. Roses exposed to WBG produced tiny leaves with abnormal margins at day 7 post-treatment (Figure 2). At day 30, WBG-exposed plants, had leaves of normal size, however, leaf margins remained abnormal. It should be noted that roses exposed to WBG did not change leaf color throughout the experiment.

Herbicide treatment

At day 1 post-treatment, differences in the spectra of control and herbicide-treated plants were observed primarily in the 1610 to 1720 cm^{-1} region, Figure 3 and Table 1. At this timepoint, the control spectra showed greater intensity in 1610 cm^{-1} and 1720 cm^{-1} peaks. The herbicide-treated plant spectra, conversely, had greater intensity than the control at 1669 cm^{-1} peak.

At day 7 post-treatment, the spectra of plants treated with each herbicide began to diverge. In the spectra of Roundup-treated plants, numerous changes throughout the spectrum were observed. Specifically, a new peak appeared at 476 cm^{-1}. Additionally, small variations in intensity relative to the control spectrum were observed from 747 to 1526 cm^{-1} region. In the 1610 to 1720 cm^{-1} region, the spectra collected from Roundup-treated plants showed greater intensity than the spectra collected from both the control and WBG plants, whereas the spectra collected from WBG-treated plants exhibited the greatest intensity at 1669 cm^{-1}. These same patterns were observed at day 14 post-treatment.

At day 30 post-treatment, while the general intensities throughout the spectra showed little relative changes, more alterations were observed in 1610 to 1720 cm^{-1} region. Specifically, all three treatments, Roundup, WBG and control, showed the same intensity of 1720 cm^{-1} peak. Additionally, the intensities of Roundup, WBG and control spectra changed relative to each other at the 1610 cm^{-1} peak. While Roundup-treated spectra continued to show the highest intensity at this band, the WBG and control spectra showed more similar intensities. Finally, the average quality of the Roundup spectra deteriorated to the point where the 476 cm^{-1} peak, previously observed at day 7 and day 14 post-treatment, was no longer distinguishable from the noise.

We then sought to determine whether Raman spectra acquired from control and herbicide-treated plants could be distinguished using multivariate methods. First, for the control and Roundup spectra, we built one PLS-DA model for each timepoint to distinguish these two categories from each other, Table 2. We found that these models enabled accurate identification of both WBG and Roundup-stresses (Chicco, 2017). We also found that PLS-DA enabled 92% accurate identification of control plants. These results demonstrated the RS could be used for the rapid, non-invasive detection of Roundup-associated changes to rose plants.

Next, we repeated this procedure to differentiate control and WBG spectra. It was found that developed PLS-DA models did not perform as well as the models discussed above. We obtained on average76% accuracy in differentiation between control and WBG spectra, Table 3. Furthermore, day 14 model performed very poorly, showing on average 59% accurate classification. However, models at day 7 and day 30 post-treatment performed significantly better showing 82% and 86% accuracy, respectively. These findings showed that RS could be used for the detection of WBG-associated stress in roses in a time-dependent manner.

FIGURE 1
Herbicide induced symptoms on Roundup treated plants observed 21 days post treatment. This included proliferated shoots at nodes and chlorosis on new and old growth. Symptomatic damaged plant parts will not recover. Additionally, some necrosis was observed on new growth up to 30 days after treatment. Plants resumed normal appearing growth patterns about 30 days after the initial application (Shires, 2020).

One may question the need for Raman diagnostics when it is not necessary to identify the specific herbicide that caused the damage. To answer this question, we combined spectra collected form WBG and Roundup treated plants at different stages of the plant vegetation and question the accuracy with which such spectra can be differentiated from the spectra collected from control plants at the same vegetation state, Table 4. We found that overall, such model demonstrates less accurate prediction on the healthy status of plants (control) compared to the herbicide damage. For instance, average accuracy of control plants ranges ~78%, whereas the accuracy of herbicide stress identification is within 85%, on average. We infer that the combined model (Table 4) provides lower accuracies due to different mechanisms of action of WBG and Roundup on plants. Although reported by this model accuracies (78% for control and 85% for the herbicide stress) are likely to be satisfactory for farmers, our data show that if higher accuracy is expected, herbicide-specific models should be used.

FIGURE 2
Typical symptoms caused by Weed-B-Gon include small, strapped leaves with abnormal leaf margins. Symptoms are visible, on the new growth, starting seven days post application. Plants appeared to resume normal new growth 30 days after treatment (Shires, 2020).

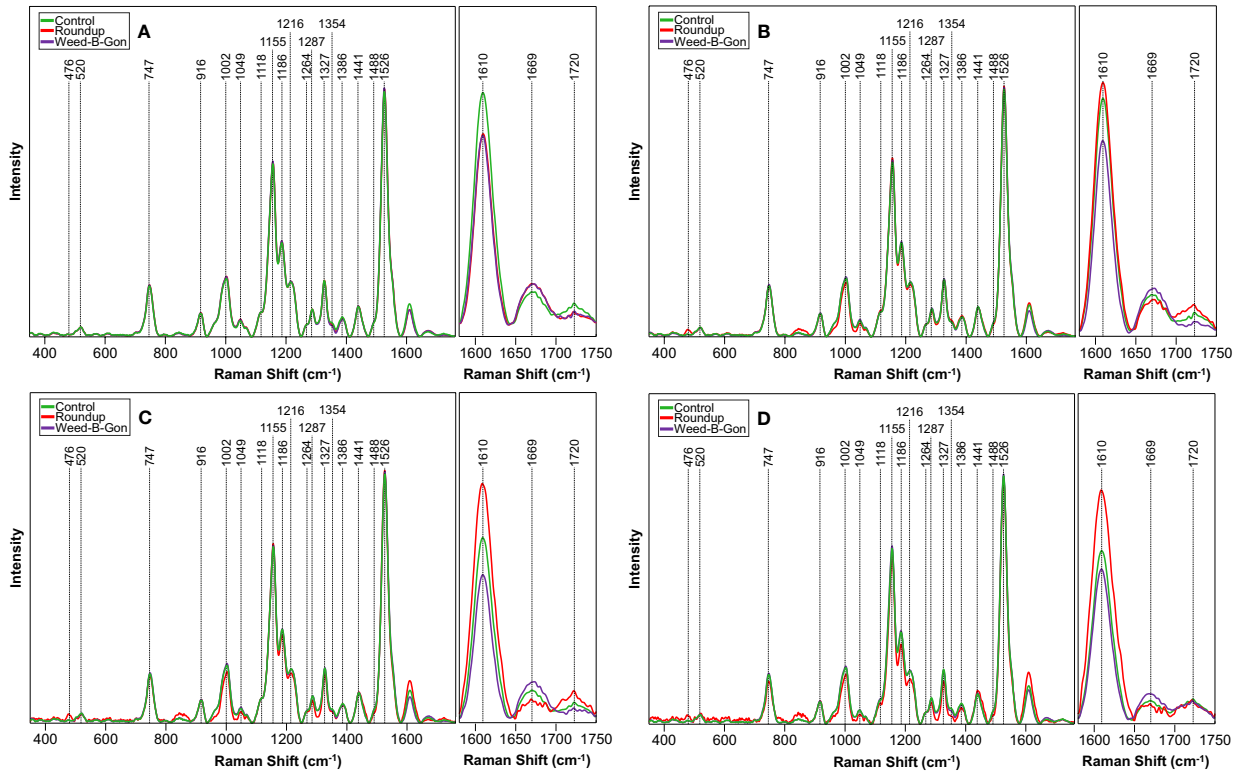


FIGURE 3
Raman spectra acquired from rose plants treated with RO water (control), Roundup, or Weed-B-Gon at **(A)** one day; **(B)** seven days; **(C)** 14 days; **(D)** 30 days after application. Inset: The spectral region 1580 cm^{-1} to 1750 cm^{-1} zoomed in for clarity. All spectra are normalized to the 1440 cm^{-1} peak.

Discussion

In plants, glyphosate concentrates in the meristems where it disrupts the shikimic pathway the prohibition of EPSP synthase (Sherwani et al., 2015). This decelerates plant growth and lowers protein expression. Roundup application also facilitates plant infection by soil-borne pathogens.

We found that spectra acquired from glyphosate-treated plants exhibited an increase in the intensity of ~1610 cm^{-1} peak at day 7, 14 and 30 post-treatment. We previously demonstrated that an increase in intensity of this peak was associated with plant infection by bacterial pathogens (Farber et al., 2019a; Sanchez et al., 2019b). Upon such infection, plants enhance production of *p*-coumaric acid that inhibits bacterial growth (Dou et al., 2021). This molecule is also used for lignin biosynthesis. Thus, producing p-coumaric acid species enhance lignification to limit bacterial propagation thought the plant. The observed in our current study increase in the intensity of ~1610 cm^{-1} peak suggest that glyphosate facilitated plant infection by pathogens present in soli (Hammerschmidt, 2018). This resulted in an increase in the synthesis of *p*-coumaric acid or similar aromatic compounds. Further experiments including mass spectrometry would be required to determine the exact biological origin of observed spectroscopic changes.

We also found that after day 1 post-treatment, the intensity of the 1669 cm^{-1} peak in the spectra acquired form glyphosate-treated plants was lower compared to the intensity of this peak in the spectra acquired from control or WBG-treated plants. This peak originates from amide I, a vibration of the backbone of proteins (Kurouski et al., 2015). A decrease in intensity of the 1669 cm^{-1} peak suggests that the total concentration of proteins is reduced in glyphosate-treated roses. These results are in a good agreement with the discussed above glyphosate-induced suppression of protein expression in plants.

Next, we found an increase in the intensity of 1720 cm^{-1} peak in the spectra collected from glyphosate-treated plants compared to the spectra acquired from control plants. This peak can be assigned to compounds with carboxyl groups, such as salicylic acid (Farber et al., 2019b; Farber et al., 2020b). This important signaling molecule is a product of the shikimate pathway (Gao et al., 2015). Thus, out results point on the accumulation of salicylic acid in plant leaves.

Unlike glyphosate, 2,4-D causes uncontrolled cellular division in plants that are exposed to this herbicide (Teale et al., 2006). This uncontrolled division is caused by cell wall plasticity, biosynthesis of proteins, and production of ethylene. MCPP is similar to 2,4-D, however, it targets auxin pathways causing elongated stems (Kelley and Riechers, 2007; Mithila et al., 2011).

We found that in the spectra collected WBG-treated plants, the intensity of the 1610 cm^{-1} peak was consistently lower compared to the intensity of this peak in the spectra acquired from control plants. These findings suggested that WBG lowered the concentration of *p*-coumaric acid or similar aromatic compounds in roses. Since *p*-coumaric acid is used in lignin biosynthesis (Amarowicz and Pegg, 2019), one can expect that uncontrolled cellular division in plants

TABLE 1 Vibrational peak assignments for the Raman spectra of Rose leaves.

Peak (cm^{-1})	Vibrational mode	Assignment
476	Glycosidic ring	Carbohydrates (Kizil et al., 2002; Almeida et al., 2010)
520	ν(C-O-C) glycosidic	cellulose (Edwards et al., 1997)
740-747	γ(C–O-H) of COOH	pectin (Synytsya et al., 2003)
905-918	ν(C-O-C) in plane, symmetric	cellulose, lignin (Edwards et al., 1997)
1000	in-plane CH_3 rocking of polyene	carotenoids (Schulz et al., 2005)
1048	ν(C-O)+ν(C-C)+δ(C-O-H)	cellulose, lignin (Edwards et al., 1997)
1118	Sym ν(C-O-C), C-O-H bending	cellulose (Edwards et al., 1997)
1157	C-C Stretching; ν(C-O-C), ν(C-C) in glycosidic linkages, asymmetric ring breathing	carotenoids (Schulz et al., 2005), carbohydrates (Wiercigroch et al., 2017)
1186	ν(C-O-H) Next to aromatic ring+σ(CH)	lignin (Mary et al., 2012; Agarwal, 2014)
1216	δ(C-C-H)	aliphatics (Yu et al., 2007), xylan (Agarwal, 2014)
1264	Guaiacyl ring breathing, C-O stretching (aromatic)	lignin (Cao et al., 2006)
1287	δ(C-C-H)	aliphatics (Yu et al., 2007)
1327	δCH_2 Bending	aliphatics, cellulose, lignin (Edwards et al., 1997)
1354	$\delta(CH_2)+\delta(CH_3)$	aliphatics (Yu et al., 2007)
1386	δCH_2 Bending	aliphatics (Yu et al., 2007)
1441	$\delta(CH_2)+\delta(CH_3)$	aliphatics (Yu et al., 2007)
1488	$\delta(CH_2)+\delta(CH_3)$	aliphatics (Yu et al., 2007)
1526	-C=C- (in plane)	carotenoids (Adar, 2017; Devitt et al., 2018)
1610	ν(C-C) Aromatic ring+σ(CH)	lignin (Agarwal, 2006; Kang et al., 2016)
1669	C=O Stretching, amide I	proteins (β-sheet) (Devitt et al., 2018)
1720	C=O Stretching	Esters, aldehydes, carboxylic acids and ketones (Colthup et al., 1990)

should drastically lower the concentration of such molecular analytes in plant tissues.

We also observed an increase in the intensity of 1669 cm^{-1} peak in the spectra acquired from WBG-treated plants compared to the intensity of this peak in the spectra acquired from control plants. As previously described, one effect for auxinic herbicides is dysregulation of auxin-regulated genes from said regulation, leading to an increase of gene products. Examples of these auxin-regulated genes include *Aux/IAA* family genes, GH3 proteins, and small auxin up RNAs (SAURs), which themselves are thought to code other short-lived small proteins (Kelley and Riechers, 2007). Increased intensity at this peak in the WBG spectra suggests that plants express more proteins, potentially due to WBG-associated gene dysregulation.

Finally, we found that at day 7, 14 and 30 post-treatments, the intensity of the 1720 cm^{-1} peak was weaker in WBG spectra compared to control spectra. These findings suggest that carboxyl-containing compounds could be metabolized by plants

Application of this Raman spectroscopy-based sensing approach could be limited due to the high capital cost of the

TABLE 2 Summary of model validation for the control vs. Roundup differentiation.

Days post-treatment	Validation Sample Size		True positive rate of control prediction	True positive rate of Roundup prediction
	Control	Roundup		
Day 1	35	38	85%	92%
Day 14	53	20	100%	100%
Day 14	24	53	70%	100%
Day 30	44	40	100%	97%

TABLE 3 Summary of model validation for the control vs. WBG differentiation.

Days post-treatment	Validation Sample Size		True positive rate of control prediction	True positive rate of WBG prediction
	Control	WBG		
Day 1	46	30	84%	70%
Day 14	40	33	72%	93%
Day 14	48	26	45%	73%
Day 30	38	38	84%	89%

TABLE 4 Summary of model validation for control vs herbicides models.

Days post-treatment	Validation Sample Size		True positive rate of control prediction	True positive rate of herbicide prediction
	Control	Herbicides		
Day 1	109	49	73%	71%
Day 14	113	43	78%	86%
Day 14	75	86	76%	86%
Day 30	58	103	82%	89%

spectrometers ($30,000 to $70,000). However, operational costs are very low. Therefore, if most cases, such testing can be implemented as the service provided to a farmer.

Finally, it is important to determine variability of the observed vibrational peaks in the spectra collected from different cultivars of roses. Such variabilities originate from differences in biochemical profiles of cultivars. Consequently, if biochemical changes among cultivars are greater that the magnitude of changes in plant biochemistry induced by herbicides, individual spectroscopic libraries will be required for each cultivar. At the same time, if the magnitude of changes in plant biochemistry induced by herbicides is greater than differences in biochemical profiles of different cultivars, the discussed above results can be used for all rose cultivars. Additional experiments are needed to disentangle these two possibilities. This work is currently in progress in our laboratory.

Conclusions

Our results show that RS can be used to detect plant exposure to herbicides with high accuracy. We also found that RS could be used to differentiate between WBG and Roundup. We infer that this sensitivity arises from drastically different mechanisms of action of these two herbicides. These findings demonstrate that RS can be a powerful tool for detecting herbicide misuse on ornamental plant.

Data availability statement

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

Author contributions

CF, MS, and JU: investigation, data curation, and methodology. KO and DK: methodology, funding acquisition, and supervision. All authors contributed to the article and approved the submitted version.

Acknowledgments

We acknowledge Star Roses for provided financial support. DK acknowledges financial support of the Institute on Human Health and Agriculture.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1121012/full#supplementary-material

References

Adar, F. (2017). Carotenoids - their resonance raman spectra and how they can be helpful in characterizing a number of biological systems. *Spectroscopy* 32, 12–20.

Agarwal, U. P. (2006). Raman imaging to investigate ultrastructure and composition of plant cell walls: distribution of lignin and cellulose in black spruce wood (Picea mariana). *Planta* 224, 1141–1153. doi: 10.1007/s00425-006-0295-z

Agarwal, U. P. (2014). 1064 nm FT-raman spectroscopy for investigations of plant cell walls and other biomass materials. *Front. Plant Sci.* 5, 1–12. doi: 10.3389/fpls.2014.00490

Almeida, M. R., Alves, R. S., Nascimbem, L. B., Stephani, R., Poppi, R. J., and De Oliveira, L. F. (2010). Determination of amylose content in starch using raman spectroscopy and multivariate calibration analysis. *Anal. Bioanal. Chem.* 397, 2693–2701. doi: 10.1007/s00216-010-3566-2

Amarowicz, R., and Pegg, R. B. (2019). "Chapter one - natural antioxidants of plant origin," in *Advances in food and nutrition research*. Eds. I. C. F. R. Ferreira and L. Barros (New York, USA: Academic Press), 1–81.

Benbrook, C. M. (2016). Trends in glyphosate herbicide use in the united states and globally. *Environ. Sci. Europe* 28, 3–3. doi: 10.1186/s12302-016-0070-0

Cao, Y., Shen, D., Lu, Y., and Huang, J. (2006). A raman-scattering study on the net orientation of biomacromolecules in the outer epidermal walls of mature wheat stems (Triticum aestivum). *Ann. Bot.* 97, 1091–1094. doi: 10.1093/aob/mcl059

Chicco, D. (2017). Ten quick tips for machine learning in computational biology. *BioData Min.* 10, 35. doi: 10.1186/s13040-017-0155-3

Colthup, N. B., Daly, L. H., and Wiberley, S. E. (1990). *Introduction to infrared and raman spectroscopy* (New York, USA: Academic Press).

Devitt, G., Howard, K., Mudher, A., and Mahajan, S. (2018). Raman spectroscopy: an emerging tool in neurodegenerative disease research and diagnosis. *ACS Chem. Neurosci.* 9, 404–420. doi: 10.1021/acschemneuro.7b00413

Dou, T., Sanchez, L., Irigoyen, S., Goff, N., Niraula, P., Mandadi, K., et al. (2021). Biochemical origin of raman-based diagnostics of huanglongbing in grapefruit trees. *Front. Plant Sci.* 12, 680991. doi: 10.3389/fpls.2021.680991

Edwards, H. G., Farwell, D. W., and Webster, D. (1997). FT raman microscopy of untreated natural plant fibres. *Spectrochim. Acta A* 53, 2383–2392. doi: 10.1016/S1386-1425(97)00178-9

Egging, V., Nguyen, J., and Kurouski, D. (2018). Detection and identification of fungal infections in intact wheat and sorghum hrain using a hand-held raman spectrometer. *Anal. Chem.* 90, 8616–8621. doi: 10.1021/acs.analchem.8b01863

Farber, C., Bryan, R., Paetzold, L., Rush, C., and Kurouski, D. (2020a). Non-invasive characterization of single-, double- and triple-viral diseases of wheat with a hand-held raman spectrometer. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.01300

Farber, C., and Kurouski, D. (2018). Detection and identification of plant pathogens on maize kernels with a hand-held raman spectrometer. *Anal. Chem.* 90, 3009–3012. doi: 10.1021/acs.analchem.8b00222

Farber, C., Sanchez, L., and Kurouski, D. (2020b). Confirmatory non-invasive and non-destructive identification of poison ivy using a hand-held raman spectrometer. *RCS Adv.* 10, 21530–21534. doi: 10.1039/D0RA03697H

Farber, C., Sanchez, L., Pant, S., Scheuring, D. C., Vales, M. I., Mandadi, K., et al. (2021). Potential of spatially offset raman spectroscopy for detection of zebra chip and potato virus y diseases of potatoes (Solanum tuberosum). *ACS Agric. Sci. Technol.* 1, 211–221. doi: 10.1021/acsagscitech.1c00024

Farber, C., Sanchez, L., Rizevsky, S., Ermolenkov, A., Mccutchen, B., Cason, J., et al. (2020c). Raman spectroscopy enables non-invasive identification of peanut genotypes and value-added traits. *Sci. Rep.* 10, 7730. doi: 10.1038/s41598-020-64730-w

Farber, C., Shires, M., Ong, K., Byrne, D., and Kurouski, D. (2019a). Raman spectroscopy as an early detection tool for rose rosette infection. *Planta* 250, 1247–1254. doi: 10.1007/s00425-019-03216-0

Farber, C., Wang, R., Chemelewski, R., Mullet, J., and Kurouski, D. (2019b). Nanoscale structural organization of plant epicuticular wax probed by atomic force microscope infrared spectroscopy. *Anal. Chem.* 91, 2472–2479. doi: 10.1021/acs.analchem.8b05294

Gao, Q.-M., Zhu, S., Kachroo, P., and Kachroo, A. (2015). Signal regulators of systemic acquired resistance. *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00228

Gleason, C., Foley, R. C., and Singh, K. B. (2011). Mutant analysis in arabidopsis provides insight into the molecular mode of action of the auxinic herbicide dicamba. *PloS One* 6, e17245. doi: 10.1371/journal.pone.0017245

Gupta, S., Huang, C. H., Singh, G. P., Park, B. S., Chua, N.-H., and Ram, R. J. (2020). Portable raman leaf-clip sensor for rapid detection of plant stress. *Sci. Rep.* 10, 20206. doi: 10.1038/s41598-020-76485-5

Hammerschmidt, R. (2018). How glyphosate affects plant disease development: it is more than enhanced susceptibility. *Pest Manage. Sci.* 74, 1054–1063. doi: 10.1002/ps.4521

Henry, W. B., Shaw, D. R., Reddy, K. R., Bruce, L. M., and Tamhankar, H. D. (2004). Remote sensing to detect herbicide drift on crops. *Weed Technol.* 18, 358–368. doi: 10.1614/WT-03-098

Henry, W. B., Shaw, D. R., Reddy, K. R., Bruce, L. M., and Tamhankar, H. D. (2017). Remote sensing to detect herbicide drift on crops. *Weed Technol.* 18, 358–368. doi: 10.1614/WT-03-098

Kang, L., Wang, K., Li, X., and Zou, B. (2016). High pressure structural investigation of benzoic acid: raman spectroscopy and x-ray diffraction. *J. Phys. Chem. C.* 120, 14758–14766. doi: 10.1021/acs.jpcc.6b05001

Karlik, J. F., and Flint, M. L. *Diseases and abiotic disorders of outdoor roses* (American Phytopathological Society). Available at: https://www.apsnet.org/edcenter/apsnetfeatures/Pages/Roses.aspx (Accessed January 21 2021).

Kelley, K. B., and Riechers, D. E. (2007). Recent developments in auxin biology and new opportunities for auxinic herbicide research. *Pesticide Biochem. Physiol.* 89, 1–11. doi: 10.1016/j.pestbp.2007.04.002

Kennard, R. W., and Stone, L. A. (1969). Computer aided design of experiments. *Technometrics* 11, 137–148. doi: 10.1080/00401706.1969.10490666

Kizil, R., Irudayaraj, J., and Seetharaman, K. (2002). Characterization of irradiated starches by using FT-raman and FTIR spectroscopy. *J. Agric. Food Chem.* 50, 3912–3918. doi: 10.1021/jf011652p

Kurouski, D., Van Duyne, R. P., and Lednev, I. K. (2015). Exploring the structure and formation mechanism of amyloid fibrils by raman spectroscopy: a review. *Analyst* 140, 4967–4980. doi: 10.1039/C5AN00342C

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

Mandrile, L., Rotunno, S., Miozzi, L., Vaira, A. M., Giovannozzi, A. M., Rossi, A. M., et al. (2019). Nondestructive raman spectroscopy as a tool for early detection and discrimination of the infection of tomato plants by two economically important viruses. *Anal. Chem.* 91, 9025–9031. doi: 10.1021/acs.analchem.9b01323

Mary, Y. S., Panicker, C. Y., and Varghese, H. T. (2012). Vibrational spectroscopic investigations of 4-nitropyrocatechol. *Orient. J. Chem.* 28, 937–941. doi: 10.13005/ojc/280239

Mehdizadeh, M., Mushtaq, W., Siddiqui, S. A., Ayadi, S., Kaur, P., Yeboah, S., et al. (2021). Herbicide residues in agroecosystems: fate, detection, and effect on non-target plants. *Rev. Agricult. Sci.* 9, 157–167. doi: 10.7831/ras.9.0_157

Mithila, J., Hall, J. C., William, G. J., Kevin, B. K., and Dean, E. R. (2011). Evolution of resistance to auxinic herbicides: historical perspectives, mechanisms of resistance, and implications for broadleaf weed management in agronomic crops. *Weed Sci.* 59, 445–457. doi: 10.1614/WS-D-11-00062.1

Payne, W. Z., and Kurouski, D. (2021). Raman-based diagnostics of biotic and abiotic stresses in plants. a review. *Front. Plant Sci.* 11, 616672. doi: 10.3389/fpls.2020.616672

Romero-Puertas, M. C., Mccarthy, I., Gómez, M., Sandalio, L. M., Corpas, F. J., Del Río, L. A., et al. (2004). Reactive oxygen species-mediated enzymatic systems involved in the oxidative action of 2,4-dichlorophenoxyacetic acid*. *Plant Cell Environ.* 27, 1135–1148. doi: 10.1111/j.1365-3040.2004.01219.x

Sanchez, L., Ermolenkov, A., Biswas, S., Septiningshih, E. M., and Kurouski, D. (2020). Raman spectroscopy enables non-invasive and confirmatory diagnostics of salinity stresses, nitrogen, phosphorus, and potassium deficiencies in rice. *Front. Plant Sci.* 11, 573321. doi: 10.3389/fpls.2020.573321

Sanchez, L., Farber, C., Lei, J., Zhu-Salzman, K., and Kurouski, D. (2019a). Noninvasive and nondestructive detection of cowpea bruchid within cowpea seeds with a hand-held raman spectrometer. *Anal. Chem.* 91, 1733–1737. doi: 10.1021/acs.analchem.8b05555

Sanchez, L., Pant, S., Xing, Z., Mandadi, K., and Kurouski, D. (2019b). Rapid and noninvasive diagnostics of huanglongbing and nutrient deficits on citrus trees with a handheld raman spectrometer. *Anal. Bioanal. Chem.* 411, 3125–3133. doi: 10.1007/s00216-019-01776-4

Schulz, H., Baranska, M., and Baranski, R. (2005). Potential of NIR-FT-Raman spectroscopy in natural carotenoid analysis. *Biopolymers* 77, 212–221. doi: 10.1002/bip.20215

Sherwani, S. I., Arif, I. A., and Khan, H. A. (2015). "Modes of action of different classes of herbicides, herbicides, physiology of action, and safety," in *Herbicides: physiology of action and safety*. Eds. A. Price, J. Kelton and L. Sarunaite (London, UK: IntechOpen Limited).

Shires, M. K. (2020). *Study of resistance to rose rosette disease utilizing field research, molecular methods, and transmission methods.*

Synytsya, A., Čopíková, J., Matějka, P., and Machovič, V. (2003). Fourier Transform raman and infrared spectroscopy of pectins. *Carbohydr. Polym.* 54, 97–106. doi: 10.1016/S0144-8617(03)00158-9

Tataridas, A., Jabran, K., Kanatas, P., Oliveira, R. S., Freitas, H., and Travlos, I. (2022). Early detection, herbicide resistance screening, and integrated management of invasive plant species: a review. *Pest Manag Sci.* 78, 3957–3972. doi: 10.1002/ps.6963

Teale, W. D., Paponov, I. A., and Palme, K. (2006). Auxin in action: signalling, transport and the control of plant growth and development. *Nat. Rev. Mol. Cell Biol.* 7, 847–859. doi: 10.1038/nrm2020

Wiercigroch, E., Szafraniec, E., Czamara, K., Pacia, M. Z., Majzner, K., Kochan, K., et al. (2017). Raman and infrared spectroscopy of carbohydrates: a review. *Spectrochim. Acta A* 185, 317–335. doi: 10.1016/j.saa.2017.05.045

Yan, Y., Ren, J., Tschannerl, J., Zhao, H., Harrison, B., and Jack, F. (2021). Nondestructive phenolic compounds measurement and origin discrimination of peated barley malt using near-infrared hyperspectral imagery and machine learning. *IEEE Trans. Instrum. Measur.* 70, 5010715. doi: 10.1109/TIM.2021.3082274

Yeturu, S., Vargas Jentzsch, P., Ciobotă, V., Guerrero, R., Garrido, P., and Ramos, L. A. (2016). Handheld raman spectroscopy for the early detection of plant diseases: abutilon mosaic virus infecting abutilon sp. *Anal. Methods* 8, 3450–3457. doi: 10.1039/C6AY00381H

Yu, M. M., Schulze, H. G., Jetter, R., Blades, M. W., and Turner, R. F. (2007). Raman microspectroscopic analysis of triterpenoids found in plant cuticles. *Appl. Spectrosc.* 61, 32–37. doi: 10.1366/000370207779701352

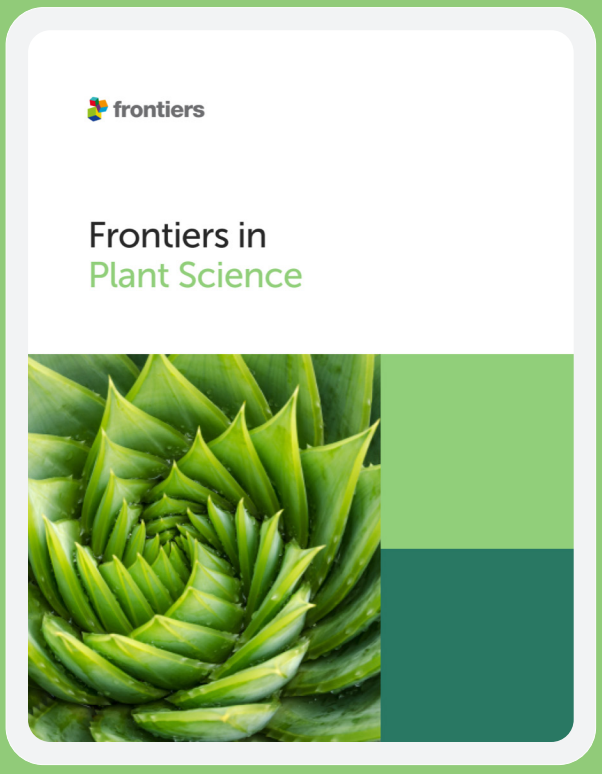
Frontiers in
Plant Science

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