REVIEW OF RESEARCH ON RECOGNITION AND MONITORING OF PLANT GROWTH PHENOTYPE BASED ON DEEP LEARNING
/
基于深度学习的植物生长表型识别监测研究现状

Zheying ZONG[1,2,3], Biao FENG[1], Shuai WANG[1], Chunhui ZHANG[*,1], Changfeng LI[1], Yongchao XU[1]
[1] College of Electromechanical Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China
[2] Inner Mongolia Engineering Research Center of Intelligent equipment for the entire process of forage and feed production, Hohhot 010018, China
[3] Key Laboratory of biopesticide Creation and Resource Utilization of Universities, Inner Mongoilia, China, Hohhot 010018, China
Corresponding author: Chunhui ZHANG
Tel: 0471-5315864; E-mail: zhangchunhui@imau.edu.cn
DOI: https://doi.org/10.35633/inmateh-77-07

Keywords: *Phenotypic analysis; Deep learning; Growth monitoring; Identification; Plant stress*

ABSTRACT
Accurate measurement of plant phenotypic data can provide a comprehensive understanding of plant physiology and help to study the relationship between plant genes and the environment. The application of visible light and other multi-source and multi-dimensional imaging sensing technology can provide a rich data source for plant phenotype identification and monitoring. With the continuous development and application of computer technology in the field of plant phenotype analysis, deep learning techniques have achieved significant progress in plant phenotype recognition and monitoring. On the basis of reviewing the relevant research results at home and abroad at this stage, this paper firstly describes the common ways of plant phenotype image acquisition. Then, it discusses in detail the current status of the application of deep learning technology in the fields of classification, detection and segmentation of plant phenotypes, crop development and yield prediction, as well as plant drought and pest stress, etc. Finally, it discusses the challenges and future development goals of the deep learning method in the monitoring and recognition of plant phenotypes. This paper aims to provide theoretical support and technical reference for the development and application of deep learning technology in the field of agricultural plant phenotyping.

摘要
精确测量植物表型数据可以全面了解植物生理状况，有助于深入研究植物基因与环境之间的关系。应用可见光等多源多维度成像感知技术可为植物表型识别与监测提供丰富的数据源，随着计算机技术在植物表型分析领域的不断发展应用，深度学习技术在植物表型识别与监测方面已取得了显著进展。在梳理现阶段国内外相关研究成果的基础上，本文首先阐述了常见的植物表型图像采集方式；之后详细探讨了深度学习技术在植物表型的分类、检测与分割，作物生长发育及产量预测，以及植物干旱与病虫害胁迫等领域的应用现状；最后讨论了深度学习方法在植株表型监测与识别中的挑战与未来发展目标。本文旨在为深度学习技术在农业植物表型领域的发展与应用提供理论支持和技术参考。

INTRODUCTION
Phenotype refers to observable morphological traits of organisms across different growth environments and stages *(Zhou et al., 2018)*. Plant phenotyping research focuses on acquiring high-quality trait data to quantify the influence of genotype and environment on key traits, such as quality, yield, and stress tolerance *(Tester & Langridge, 2010, 2010; Ribaut et al., 2010)*. Accurate identification and monitoring of crop morphological traits during growth enable precise water and fertilizer management, enhancing intelligent agricultural production. Current plant phenotyping research emphasizes developing specialized algorithms to expand observable and measurable plant traits while enhancing analytical accuracy and efficiency *(Zhao, 2019)*. Table 1 illustrates classifications of common plant phenotypes and their specific characteristics. Optical sensors now enable extraction of plant growth parameters from captured images, and image technology applications in crop growth monitoring, pest and disease detection, and yield estimation are expanding. For instance, features like leaf area, plant height, and stem width can be derived directly from 2D images, while indices such as color metrics and normalized difference vegetation index (NDVI) facilitate quantitative plant growth analysis *(Sancho-Adamson et al., 2019)*.

Table 1

Common plant phenotypic classifications and their specific characteristics

Phenotypic classification	Subphenotypic classification	Characteristic description
Morphological	Root system	Root type, length, density, diameter, and number of branches
	Stem	Stem thickness and growth habit (upright or prostrate)
	Leaf blade	Leaf shape, color, size, thickness, and arrangement
	Flower	Flower pattern, color, size, and arrangement
	Fruit	Fruit shape, size, color, number and seed distribution
Physiological	Photosynthesis	Photosynthetic rate and chlorophyll content
	Transpiration	Stomatal density, opening and closing behavior, and water-use efficiency
	Stress tolerance	Drought, salt, flood, cold, and high-temperature resistance
	Nutrient absorption	Efficiency of elemental absorption
Developmental biology	Seed germination	Germination rate, time, and potential
	Seedling stage	Height, growth rate, and color
	Growth cycle	Growth rate, flowering period, maturity period, and lifespan
	Branching and sprouting	Number of tillers and bud distribution

Extensive studies have shown that deep learning excels in solving intricate problems, with image recognition being a prominent success case *(Syuen et al., 2022)*. Researchers have applied deep learning techniques to agriculture. Single-layer neural network methods are limited to specific plant phenotypic traits in specific environments *(Hou et al., 2017; Zhang et al., 2020)*. In contrast, multi-layer neural network structures, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) models, exhibit better adaptability and have achieved impressive results in agricultural applications *(Gowri et al., 2024)*.

This paper systematically reviews traditional methods for capturing plant phenotyping images and summarizes the current applications of deep learning in plant classification, detection, and segmentation, as well as in growth, development, and yield prediction, and in assessing drought, pest, and disease stress. Finally, it discusses the challenges and future goals of deep learning in plant phenotypic monitoring and identification. The study aims to provide theoretical support and technical references for advancing deep learning applications in agricultural plant phenotyping.

TRADITIONAL METHODS FOR PLANT PHENOTYPIC IMAGE ACQUISITION

By processing multi-source plant image data, crop phenotypic parameters can be monitored in real time under different environmental conditions. This paper introduces five commonly used plant phenotyping acquisition techniques: visible light imaging, fluorescence imaging, infrared thermal imaging, 3D imaging, and hyperspectral imaging. Table 2 summarizes the different imaging techniques and their applications in plant phenotyping.

Visible light imaging

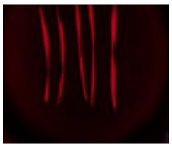
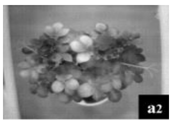
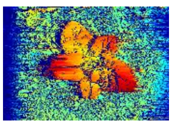
Visible light imaging captures reflected or transmitted light signals in the visible spectrum, converts them into digital data, and outputs the image *(Aboalia et al., 2024)*. *Marchetti et al., (2019)*, designed and validated a non-invasive, reproducible indoor phenotyping technique based on RGB images for screening barley genotypes with potential tolerance to water stress. Results showed significant correlations between canopy height and above-ground fresh weight ($R^2 = 0.90$, $P < 0.001$), leaf length, width, and relative water content (RWC). However, RGB imaging alone provides only morphological data. Comprehensive analysis of barley physiological responses requires integration with fluorescence imaging or destructive sampling.

Gée et al. (2023), calculated the normalized dark green color index (nDGCI) from RGB images of winter wheat captured under uniform lighting to estimate the nitrogen nutrition index (NNI), thereby predicting nitrogen status. Under single-variety conditions, the coefficient of determination R^2 between nDGCI and NNI reached 0.73–0.91, indicating a high correlation between the two indices. *Coswosk et al. (2024),* utilized RGB images to analyze the relationships between vegetation indices and yield, morphological, and physiological traits across different maize genotypes. Their findings revealed that visible vegetation indices such as VARI, NGRDI, and GLI showed strong correlations with grain yield (reaching up to r=0.99). However, RGB sensors can only capture visible light bands and cannot capture near-infrared information, limiting their precise assessment of physiological parameters such as chlorophyll fluorescence and water stress.

Visible light imaging technology is cost-effective and versatile, but it cannot capture depth or near-infrared information. It struggles to extract effective features from sparse canopies and fails to obtain additional phenotypic data. Factors such as lighting conditions, background interference, and crop organ occlusion also significantly impact its analytical accuracy. These limitations hinder its large-scale deployment in practical production.

Table 2

Different Techniques for Plant Phenotyping

Imaging technology	Sensor	Original image	Extracted parameters	Advantages	Image example	Reference
Visible light imaging	Charge-coupled device(CCD)	Grayscale or color image	Canopy size, plant structure, projected area, and other parameters	Low cost, provides detailed color information	 *(Coswosk et al. 2024)*	*Marchetti et al., 2019;* *Gée et al. 2023;* *Coswosk et al. 2024*
Fluorescence imaging	Fluorescence imager (FLI)	Fluorescence emission spectrum in the red and far-red regions	Chlorophyll content monitoring and photosynthetic status	Enables monitoring of photosynthetic efficiency, physiological status, and early detection of plant	 *(G. Liu et al., 2024)*	*Si et al., 2023;* *Liu G. et al., 2024;* *Banah et al., 2024*
Thermal infrared imaging	Infrared camera	Continuous or discrete spectrum in the infrared region	Leaf temperature, and pest/disease conditions	Enables monitoring of plant moisture content, stress level, and temperature distribution	 *(Sashuang et al. 2023)*	*Sashuang et al. 2023;* *Ma et al. 2024*
Three-dimensional imaging	Stereo or depth camera	Depth image	Plant height, branching, and canopy structure	Fast and cost-effective 3D data acquisition	 *(Ge et al. 2020)*	*Wang R. et al. 2022; Tunca et al. 2024; Chen et al., 2024; Wang Y. et al., 2022; Kumar et al., 2022; Wang Y.et al., 2024; Zheng et al., 2022; Ge et al. 2020; X. Liu et al., 2024; Y. Zhou et al., 2024*
Hyperspectral imaging	Hyperspectral imager (HSI)	Continuous/discrete spectrum	Leaf/canopy moisture, vegetation health index, and canopy density	Enables identification of nutritional status, plant type, and disease presence	 *(Eva et al. 2023)*	*Eva et al. 2023;* *Almoujahed et al. 2025;* *Rahman et al. 2025*

Fluorescence imaging

Fluorescence imaging captures the emitted light from fluorescent substances that are excited by specific wavelengths, generating images based on their luminescent properties. *Si et al., (2023)*, conducted drought stress experiments on Longjing tea seedlings in a dark chamber. By analyzing chlorophyll fluorescence excited by 248.6 nm deep ultraviolet laser, they found that the slope fluorescence index (SFI) was highly correlated with the number of stress days (R^2 = 0.94). *Liu G. et al., (2024)*, employed multispectral fluorescence imaging to assess chlorophyll content in wheat. By processing multiple fluorescence images and calculating the ultraviolet-to-visible fluorescence ratio parameter, they found ten fluorescence parameters—including Fr740—to be highly correlated with chlorophyll content. However, fluorescence signals are susceptible to interference from ambient light, necessitating measurements under low-light conditions. *Banah et al. (2024)*, used a deep ultraviolet fluorescence spectrometer to collect emission spectra from healthy maize leaves and those infected with Southern leaf blight (SLB). Detection at the excitation wavelength revealed that SLB-infected leaves exhibited significantly higher fluorescence intensity at 325 nm compared to healthy leaves, indicating that enhanced fluorescence at this wavelength directly correlates with fungal infection phenotypes. However, the large diameter of the fluorescence spots made it difficult to distinguish fungal spore concentration gradients and their spatial distribution.

Although fluorescence imaging offers advantages such as molecular-level sensitivity and real-time analysis of physiological processes, its application is constrained by drawbacks including significant interference from light exposure and background noise, relatively large fluorescence spot diameters, and the requirement for species- or genotype-specific calibration.

Infrared thermography

Infrared thermography captures and analyzes infrared radiation emitted by objects to reveal spatial temperature distributions, converting the radiation intensity into pseudo-color images *(Zhang et al., 2022)*. *Sashuang et al. (2023)*, utilized thermal imaging fused with visible light to extract multimodal features from potato canopy fusion images. They constructed a partial least squares regression model to predict potato photosynthesis and fluorescence dynamics, achieving an R^2 = 0.85 for net photosynthetic rate and R^2 = 0.66 for stomatal conductance in Zhongshu No. 5 potatoes. However, thermal infrared cameras (640×480 resolution in this study) generally have lower resolution than RGB cameras, making it difficult to capture fine features or early stress responses. *Ma et al. (2024)*, utilized infrared thermal imaging to obtain winter wheat canopy temperature data, thereby calculating the crown water stress index (CWSI). Findings revealed that during the jointing stage, CWSI exhibited extremely significant negative correlations ($p<0.01$) with soil moisture content and leaf relative water content, and significant negative correlations ($p<0.05$) with stomatal conductance, transpiration rate, and photosynthetic rate. During the flowering and grain filling stages, CWSI exhibited extremely significant negative correlations ($p<0.01$) with all physiological and soil parameters. However, to mitigate environmental fluctuations, infrared thermography measurements typically require specific time windows, limiting measurement flexibility and real-time monitoring capabilities.

Infrared thermal imaging can directly reflect plant physiological states and capture changes in plant physiological functions. However, it is highly sensitive to environmental conditions and has relatively low resolution, making it difficult to detect micro-scale temperature variations at the leaf level. Therefore, it must be combined with other data to interpret physiological mechanisms and avoid misinterpretation.

Current status of three-dimensional stereoscopic plant phenotypic analysis and detection

Three-dimensional (3D) measurement technology allows fast, accurate, and non-invasive measurements, offering fundamental data for quantitative studies on plant growth patterns *(Jiang et al., 2025)*. To enhance measurement accuracy and capture detailed crop spatial morphology, researchers have integrated 3D data into non-destructive crop inspection methods *(Lei et al., 2024)*. Based on measurement methods, 3D phenotypic inspection techniques primarily include structure-from-motion (SfM), structured light methods, laser scanning, and binocular stereo vision techniques *(Yu et al., 2024)*.

The Structure-from-Motion (SfM) method reconstructs 3D information of target objects by processing sequential images or videos. *Wang R. et al. (2022)*, achieved a segmentation accuracy of 0.961 for stem and leaf parts of seedling-stage Chinese cabbage using an improved hypervoxel segmentation method based on the SfM algorithm. The coefficient of determination R^2 for measured phenotypic parameters such as plant height and leaf length exceeded 0.98, with the MAPE for all parameters except leaf area remaining below 2.5%.

Tunca et al. (2024), calculated sorghum plant height using the difference between the digital surface model (DSM) and digital terrain model (DTM) generated through SfM. The results showed high consistency with ground-truth measurements (R^2 = 0.97, RMSE = 8.77 cm, MAPE = 5.98%), demonstrating that SfM combined with DSM, DTM, and ground control points can accurately estimate sorghum plant height. Although SfM achieves high accuracy under ideal conditions, it relies on feature matching across multi-view images. The resulting point cloud data is voluminous, making reconstruction of complex structures challenging.

The structured light method obtains 3D information of an object by projecting a specific light pattern onto the object and capturing the deformation of the pattern using a sensor. Researchers acquired point cloud data of black-skinned chicken-of-the-woods fruiting bodies using an SR300 depth camera based on structured light technology, and obtained point cloud data of spherical fruit bodies using a dual-camera structured light system. Through point cloud reconstruction techniques and processing algorithms, they successfully extracted target three-dimensional phenotypic parameters. The former achieved an R^2 value exceeding 0.97, while the latter demonstrated a relative error in diameter reconstruction below 3.32% *(Chen et al., 2024; Y. Wang et al., 2022)*. Although the structured light method yielded favorable results, it also faces limitations such as high demands on ambient lighting conditions, algorithmic complexity requiring manual intervention, and potentially reduced reconstruction accuracy for irregular shapes.

3D laser scanning technology reconstructs 3D data of objects with high accuracy by scanning. Researchers employed LiDAR technology, 3D digital scanning methods, and 3D reconstruction techniques to capture 3D point clouds of strawberries at different growth stages, morphological structures of wheat seedlings, and multi-angle fruit point clouds of sweet peppers. This enabled non-destructive measurement of strawberry, wheat seedling stage, and bell pepper fruit. Among these, the R^2 values for individual strawberry point cloud counts, volume, plant height, and canopy area were relatively high (0.98/0.90/0.93/0.96). For wheat seedlings, the extraction accuracy of stem length, leaf length, stem diameter, and stem-leaf angle was high (R^2=0.93/0.98/0.93/0.85). For bell pepper fruits, the mean relative errors for fruit width and height were small, at 1.72% and 1.60%, respectively *(Kumar et al., 2022; Wang Y. et al., 2024; Zheng et al., 2022)*. Although 3D laser scanning technology offers high non-destructive fitting accuracy, its system integration, calibration, and algorithms are extremely complex, resulting in high deployment costs.

The binocular stereo vision method measures depth information of objects using the parallax between two cameras with different viewing angles. *Ge et al. (2020)*, employed binocular stereovision technology to identify cauliflower seedlings, utilizing the differential sum-of-squares algorithm for the stereo matching step and the k-medoids algorithm for the clustering step. Results demonstrated that this method achieved an average correct recognition rate of 98.75% across 240 pairs of cauliflower seedling images. To address the accuracy challenges in image depth estimation models caused by insufficient effective photometric loss metrics, researchers employed a stereo camera to capture RGB and depth images of field corn and Chinese cabbage. Segmentation and depth prediction were then performed. Compared to Monodepth2, the average relative error and average absolute error for field corn depth estimation decreased by 48.2% and 17.1%, respectively, while the average absolute error for Chinese cabbage leaf inclination was less than 5.5°, validating the feasibility of automated non-destructive crop measurement *(X. Liu et al., 2024; Y. Zhou et al., 2024)*. Although the binocular stereoscopic vision method is low-cost and generates highly accurate 3D data, it requires strict calibration and alignment procedures. Matching accuracy is affected by lighting conditions, and computational complexity is relatively high.

Various 3D imaging technologies exhibit significant differences in accuracy, cost, and environmental requirements, creating complementary application scenarios. Compared to structured light methods, which are susceptible to ambient light interference, and stereo vision, which involves complex calibration and is affected by lighting conditions, Structure from Motion (SfM) relies less on lighting. However, it generates massive point cloud datasets, posing challenges for reconstructing complex plant structures. In terms of accuracy, 3D laser scanning typically achieves the highest precision but involves complex system integration and high deployment costs. SfM and stereo vision offer a better balance between cost and accuracy. Therefore, selecting the appropriate 3D imaging technology requires weighing accuracy requirements, budget constraints, and environmental conditions.

Hyperspectral imaging technology

Hyperspectral imaging captures both spatial and spectral information of a target, providing continuous spectral band data for each pixel. *Eva et al. (2023)*, investigated the effects of nematode feeding and drought stress on the spectral characteristics of maize leaves to evaluate the effectiveness of hyperspectral imaging in distinguishing between biotic and abiotic stresses in maize. The results showed that the accuracy rates for pest detection and drought stress detection reached 98% and 95.9%, respectively. However, the hyperspectral images in this study contained hundreds of bands, posing significant challenges for data processing. *Almoujahed et al. (2025)*, employed hyperspectral cameras to detect Fusarium head blight (FHB) in winter wheat, successfully distinguishing healthy from infected ears before visual symptoms appeared, achieving an overall classification accuracy of 84.7%. However, hyperspectral data in the experiment proved susceptible to environmental interference, requiring multiple calibrations. Among the 224 bands in the full spectrum, only a few were correlated with FHB, making the optimization process cumbersome. *Rahman et al. (2025),* developed a high-throughput plant phenotyping system that successfully predicted leaf water content under drought conditions by analyzing spectral reflectance from acquired hyperspectral images, achieving an $R^2 = 0.81$ on the test set. However, hyperspectral data comprise numerous bands (224 bands used in this study), with adjacent bands exhibiting high correlation, thereby increasing the complexity of data processing and model training.

Despite demonstrating significant potential in plant phenotyping research, hyperspectral imaging technology faces multiple constraints in its application, including environmental interference, data noise, dimensional redundancy, and hardware costs. Future efforts should integrate multispectral fusion, field adaptive calibration, or low-cost sensor optimization to enhance its practicality.

CURRENT STATUS OF PLANT PHENOTYPIC ANALYSIS AND MONITORING BASED ON DEEP LEARNING

Deep neural networks consist of multiple layers of trainable structures, through which data passes to become more abstract and is eventually represented as a high-dimensional feature vector. Developed from early simple perceptual machines, deep neural networks can process complex information nonlinearly and automatically learn hierarchical features of the data, eliminating the need for manual feature extraction. Current research directions in deep learning for plant phenotyping and monitoring include: identification of crop morphological traits, prediction of crop growth, development, and yield, regression-based yield forecasting, and studies on plant stress responses *(Bini et al., 2021)*.

Recognition of plant phenotypic traits using deep learning

In recent years, plant analysis research has gradually evolved from basic classification and regression to object detection and segmentation. In the field of crop phenotype recognition, the application of deep learning techniques is primarily categorized into classification, detection, and segmentation. This paper will systematically explore the application of deep learning in plant phenotypic feature recognition, focusing on these three categories. Figure 1 presents the classification, detection, and segmentation diagram of corn plants *(Murphy et al., 2024)*, and Table 3 provides the plant morphological feature recognition based on deep learning.

Fig. 1 - Example of classification, detection and segmentation of corn plants
(*Murphy et al., 2024*)

（1）Classification

Plant image classification is generally divided into two main categories: cross-species coarse-grained semantic classification, which involves recognizing different plants by acquiring salient features from plant images, and same-species fine-grained classification, which requires the model to focus on subtle features in images to distinguish similar plants. In plant image classification, convolutional neural networks (CNNs) demonstrate excellent performance. Their network structure includes feature extraction and mapping modules, achieving neuron weight sharing through convolution, activation, pooling, and fully connected layers, directly extracting plant features from images to improve processing speed and accuracy. Figure 2 presents the schematic diagram of the convolutional neural network.

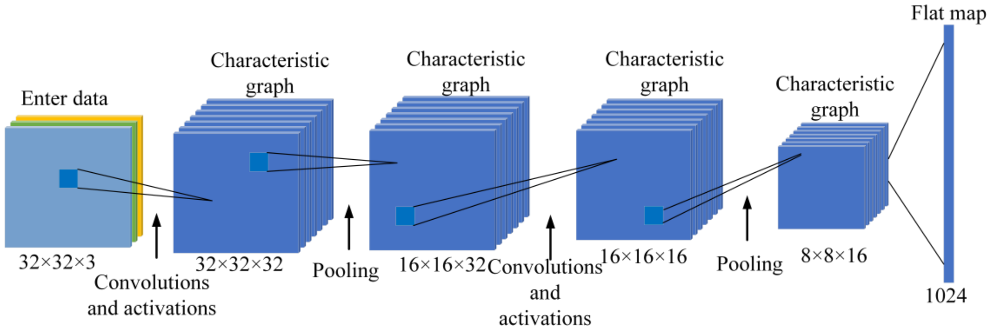


Fig. 2 - Convolutional neural network schematic*(Cao et al., 2024)*

Currently, ResNet and MobileNet are the two most popular neural networks for plant image classification tasks, representing two distinct optimization approaches *(Yang et al., 2025)*. ResNet's core advantage lies in its effective mitigation of gradient vanishing in deep networks through residual modules, enhancing feature propagation capabilities to better capture complex, abstract features. MobileNet, on the other hand, prioritizes extreme efficiency. By employing separable convolutions, it significantly reduces computational complexity and parameter count, making it well-suited for deployment on resource-constrained mobile devices while maintaining precise capture of fine-grained features.

Ali et al. (2022), developed a plant leaf classification model based on the MobileNetV2 architecture using transfer learning. This model eliminates the need for manual feature extraction, employs a shallower architecture with fewer asynchronous parameters, and achieves classification accuracy exceeding 90%. However, it exhibits high system complexity, requiring the design of intricate termination conditions and knowledge transfer logic, along with parameter tuning. *Vásconez et al. (2024)*, conducted a classification study on tomato wilt disease, comparing and analyzing 14 deep learning models. They evaluated performance using multiple metrics, with MobileNet-v2 and Xception models ultimately demonstrating the best results, both achieving an accuracy rate of 97.7%. However, the study's conclusions are based on a specific dataset and disease, and their generalizability to other plant diseases remains to be verified. Due to the unreliability of classification results generated from limited labeled crop samples, *Wang H. et al. (2024),* developed a customized network architecture, VPSNet. This network draws inspiration from ResNet's residual connection concept, introducing shortcut connections within each block to mitigate the vanishing gradient problem in deep networks. It enables efficient extraction of multi-level semantic features from limited labeled samples. Experiments demonstrate that VPSNet achieves outstanding performance across three limited-labeled-sample domains (SA II, SA III, SA V). However, this method was designed for satellite imagery and cannot be directly applied to RGB plant images.

When pursuing extreme precision with ample computational resources, ResNet-based models should be prioritized. For mobile deployment or real-time processing, MobileNet and its variants (such as MobileNetV2) are the preferred choice. When data is scarce, customized network architectures and domain adaptation algorithms (like VPSNet) can be designed based on data characteristics (e.g., time series) to fundamentally enhance the model's ability to learn from limited samples.

（2）Detection

The detection task involves both classification and localization of targets, with the main models for plant image detection currently including YOLO and Faster R-CNN. Deep learning frameworks for target detection typically require large and diverse datasets for training. However, practical applications often encounter issues such as insufficient data volume and imbalanced sample sizes across categories, which can hinder the performance of trained models.

Seyam et al. (2024), conducted a study on leaf disease detection using nine pre-trained models, including DenseNet201, alongside a custom CNN (LDDTA). All models were trained and evaluated on the PlantVillage dataset. Results indicate that LDDTA achieves comparable performance to pre-trained models while requiring the shortest training time (1891.22 seconds). With only 184,890 parameters and a model size of 2.36 MB, it significantly undercuts other models, achieving an optimal balance between efficiency and resource consumption. *Rai and Pahuja (2024)*, designed a deep convolutional neural network (ETL-NET) based on ensemble transfer learning to detect diseases in cotton leaves and plants. By employing a bagging strategy to average and fuse the probability outputs from five optimal models including InceptionV3, the approach enhanced robustness. In real field conditions, ETL-NET achieved a near-perfect classification accuracy of 99.48%. However, it should be noted that due to integrating five pre-trained models, ETL-NET's training time is significantly longer than that of a single model. *Mhala et al. (2025)*, developed an efficient and robust potato disease diagnosis system by integrating transfer learning, customized data augmentation, and regularization techniques. Performance comparisons of NASNetMobile, DenseNet201, and ResNet152v2 models were conducted on detecting three types of diseased potato leaf images captured in real field environments. Results showed DenseNet201 achieved the highest accuracy of 77.14% on the original dataset, but its large parameter count necessitates future exploration of distillation and quantization techniques.

Table 3

Deep Learning-Based Recognition of Plant Morphological Features

Task	Database	Data preprocessing	Data enhance-ment	Model	Effect	Comparison algorithm	Refe-rence
Tomato Wilt Classification	3,737 RGB images	Scale-dependent image	Not applicable	MobileNet-v2, Xception	Accuracy: 97.7%	12 Mainstream CNN models used	*(Vásconez et al. 2024)*
Plant leaf stage identification	MalayaKew dataset	Rotated by 7 degrees	Rotation, shifting, mirroring, and shuffling	S-LeafNET, W-LeafNET,P-LeafNET	The accuracy rate exceeds 99%	ResNet26, Alex-Net, and ResNet50	*(H. Wang et al. 2024)*
Blade disease detection	PlantVillage dataset	Image standardization, and normalization of image values	Rotation, flipping, cropping, and adjustment	Custom convolutional neural network	Outperforms pretrained models	DenseNet201, among others	*(Seyam et al. 2024)*
Classification of diseased cotton leaves	Kaggle's datasets with 2,293 and 1,711 images	Normalized image size	Random permutation of image data	Bagging integrated	Accuracy rate exceeding 98.5%	InceptionV3, among others	*(Rai & Pahuja , 2024)*
Detection of potato leaf blight	3,076 disease-infected leaf images	Dataset allocation and pixel value normalization	Rotation and data enhancement techniques	DenseNet201	Achieves the highest accuracy	NASNetMobile and ResNet152v2	*(Mhala et al. 2025)*
Weed detection	3,857 annotated images	Image standardization, segmentation, and preprocessing	Not applicable	Optimized DeepLabV3+ model	Average accuracy rate exceeding 99.5%	CCNet,GCNet ,ISANet, and DeepLabV3	*(T. Liu et al., 2024)*

When models need to be deployed for edge computing and mobile applications, the LDDTA model is recommended. When computational cost is not a concern and peak performance is required, the ETL-NET model is the preferred choice. For detecting complex backgrounds and unevenly distributed samples of different diseases, the DenseNet201 model is suitable.

（3）Segmentation

The effectiveness of leaf segmentation is often influenced by dataset quality as well as environmental factors such as complex backgrounds and lighting conditions *(Wang & Cao, 2021)*.

In recent years, image segmentation research has incorporated various deep learning methods, with segmentation tasks categorized into semantic segmentation and instance segmentation. Semantic segmentation classifies images at the pixel level, accurately identifying different regions of crops and backgrounds. It is widely used for recognizing structures such as crop morphology, leaves, and roots, with U-Net and DeepLabV3+ being common semantic segmentation architectures. Figure 3 illustrates the U-Net model *(Cao et al., 2024)*. Instance segmentation, in contrast, not only distinguishes between different categories but also identifies individual instances within the same category, with Mask R-CNN being a typical method.

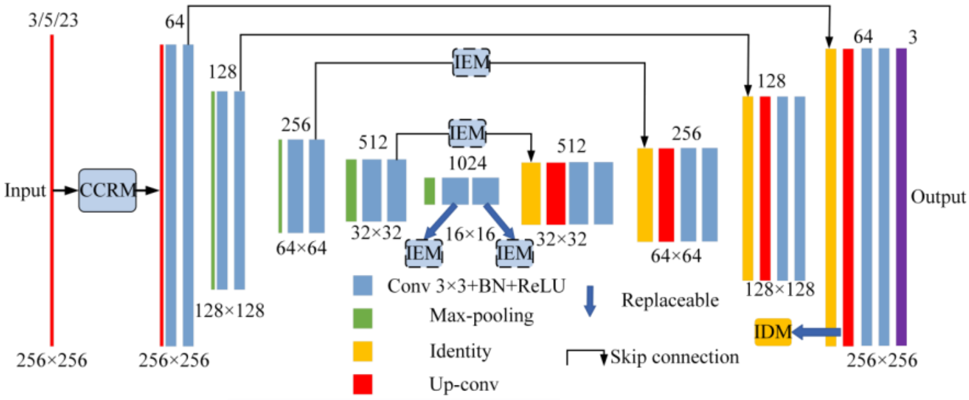


Fig. 3 - U-Net model diagram *(Cao et al., 2024)*

In plant image semantic segmentation, *Li Q. et al. (2021)*, proposed an ensemble U-Net segmentation model designed for small-sample datasets. By integrating residual blocks and gated convolutions to construct edge structures, this model effectively extracts boundary semantic information from target images. It achieved a segmentation accuracy exceeding 92.14%, significantly improving fruit segmentation precision and model generalization. However, the integrated U-Net model's incorporation of residual blocks and other structures substantially increases its parameter count and computational complexity. *Liu T. et al., (2024)*, proposed a semantic segmentation-based weed detection method aimed at simplifying the detection process through crop pixel segmentation and vegetation identification outside crop masks. The indirect segmentation strategy avoids the complexity of directly detecting different weed species. By optimizing the ResNet50-based semantic segmentation model DeepLabV3+ using knowledge distillation techniques, it achieves faster and lighter inference. Results show the model size reduced from 349.1 MB to 265.2 MB, while inference speed increased from 10.2 FPS to over 34 FPS, achieving an average accuracy exceeding 99.5%. However, this approach assumes all non-corn green vegetation is weeds, potentially failing when corn and weeds share similar colors.

In plant image instance segmentation, *Zeng et al. (2023)*, developed MT-SegNet, a multi-task segmentation network for 3D point clouds. By integrating a multi-value conditional random field model, it successfully achieved instance segmentation of Colocasia esculenta leaf point clouds, achieving an average segmentation accuracy of 88.10% and an average recall of 78.44%. Results demonstrate superior performance across all metrics compared to multiple state-of-the-art models such as PointNet. Although the model demonstrates strong segmentation performance, some misclassified points persist at the junctions between stems, leaves, and overlapping foliage. To address the limitations of remote sensing in individual crop analysis, *Yuan et al. (2024)*, combined multispectral data with an enhanced YOLOv8-Seg model for segmenting complex-morphology individual Chinese cabbage plants. This model supports independent segmentation masks and detection at different scales, employs a Path Aggregation Feature Pyramid Network (PAFPN) for multiscale feature integration, and optimizes sample matching. Results demonstrate that the method achieves a detection mAP@0.5:0.95 of 92.6% and segmentation mAP@0.5:0.95 of 86.3%, proving its ability to maintain high segmentation accuracy even at lower spatial resolutions. To address the high cost of manual annotation for leaf segmentation, *Zhou L. et al. (2024)*, proposed an automated annotation method based on zero-shot learning and transfer learning: first, a visual model locates leaves and generates initial masks; then, a lightweight YOLOv8-Segment model is trained. Requiring only 6.5MB of parameters, this model achieves an AP_{50} segmentation accuracy of 0.888, outperforming traditional methods relying on 150 manually annotated plants (AP_{50}=0.841) while completely eliminating manual annotation. However, its performance heavily depends on the accuracy of initial detection boxes, with early localization failures leading to cascading errors.

In balancing accuracy and complexity, *Li Q. et al. (2021)* and *Yuan et al. (2024)* enhanced accuracy through complex structures but sacrificed efficiency. *Liu T. et al., (2024)* and *Zhou L. et al. (2024)* achieved lightweighting via strategy design or distillation techniques but relied on specific prerequisites. Regarding application scenarios, *Zeng et al. (2023)* focused on 3D point clouds, *Yuan et al. (2024)* targeted multispectral remote sensing, while *Liu T. et al., (2024)* and *Zhou L. et al. (2024)* prioritized cost and efficiency in practical deployments. Most methods demonstrate strong performance on specific datasets or under controlled conditions, yet stability issues persist in complex natural environments.

DEEP LEARNING-BASED PREDICTION OF CROP GROWTH, DEVELOPMENT AND YIELD

Predicting crop growth involves modeling environmental and historical data to forecast future growth conditions and yields. Traditional methods rely on statistics and machine learning, but require manual feature selection, resulting in complex models that struggle to handle nonlinear relationships. In recent years, researchers have developed time-series prediction models to extract temporal features, predict the impact of environmental factors on crop growth, and provide novel approaches for phenotyping simulation and prediction(*Nguyen et al., 2025*).

Li L. et al. (2021), constructed a tomato transpiration prediction model based on a Long Short-Term Memory (LSTM) network, using air temperature and humidity, light intensity, and canopy relative leaf area index as input variables. Results showed that on the test set, the ST-LSTM model achieved a coefficient of determination R^2 of 0.9925, with a mean absolute error (MAE) of 4.53 g and root mean square error (RMSE) of 11.02 g, significantly outperforming comparison models (NARX, Elman, and RNN). However, this model requires extensive high-quality time-series data for training, and data gaps or noise may compromise its performance. *Wang C. et al. (2022)*, developed a plant growth and development prediction model based on spatio-temporal long short-term memory (ST-LSTM) networks, incorporating environmental factors. By leveraging the spatio-temporal dependency of plant growth and integrating spatio-temporal deep features, the model successfully predicted future plant growth and development image sequences. Experimental results showed prediction R^2 values of 0.9619, 0.9158, and 0.9087 for canopy leaf area, canopy width, and leaf count, respectively, validating the effectiveness of ST-LSTM in plant growth and development forecasting. However, the model exhibits certain limitations, such as image blurring with extended prediction periods and reduced accuracy for complex plant structures. 0.9087, and 0.9158, respectively, validating the effectiveness of ST-LSTM in plant growth prediction. However, the model exhibits limitations, such as image blurring with extended prediction time, insufficient accuracy for complex phenotypic features, and restricted processing capability to background-removed RGB images. *Yang et al. (2024)*, proposed a hybrid model, TMEAD-BiLSTM, combining mutation point detection with deep learning to predict alfalfa leaf area index (LAI). This model integrates environmental factors (temperature, soil moisture, etc.), meteorological data, and growing days. By employing a BiLSTM network with an encoder-attention-decoder architecture to capture long-term dependencies, it achieves prediction accuracy with $R^2 > 0.99$, significantly outperforming traditional models. However, the model exhibits high computational complexity and relies on large-scale, high-quality datasets for training.

Addressing the limitation of existing studies that focus solely on predicting dynamic changes in single phenotypic traits without comprehensively illustrating the entire plant growth process, *Duan et al. (2024)*, proposed a multi-variety rice growth visualization prediction technique based on an improved Pix2Pix-HD model. Employing a data-driven strategy, the method achieved an average correlation coefficient of 0.762 between 15 morphological and textural phenotypic traits extracted from predicted images and their actual values, enabling high-resolution growth visualization simulation. Notably, this model only predicts for a single growth stage and does not cover the entire rice growth cycle.

Li L. et al. (2021) and *Yang et al. (2024)* focused their research on high-precision numerical predictions (e.g., transpiration, LAI), making their models highly suitable for decision support in precision agriculture. *Wang C. et al. (2022)* and *Duan et al. (2024)* dedicated their efforts to generating visual images, providing richer and more comprehensive phenotypic information while enhancing interpretability. All these studies indicate that high-performance plant growth prediction models require large-scale, high-quality data. As models become more complex (e.g., incorporating attention mechanisms or GANs), predictive capability increases alongside computational costs and data quality requirements. Furthermore, accuracy degradation in long-term predictions—such as blurred images or incomplete cycle coverage—remains a critical challenge that demands urgent solutions.

In recent years, academic research on crop yield prediction has primarily employed neural network models. These models monitor growth environments by collecting feature data, identify key factors, and digitally input them into the model. Once established, the model's performance is evaluated by comparing predicted values with actual values. Existing research shows a shift in both data and model structure—from single data sources to multi-source data fusion, and from single deep learning architectures to multi-model coupling *(Li et al., 2024)*.

In current research, remote sensing technology is widely employed for yield prediction. Researchers extract key yield-indicating features by analyzing remote sensing imagery and feed these features into neural networks for predictive analysis. *Tian et al. (2024)*, proposed the AMCN model (AMCN1 architecture), which concurrently integrates CNN and LSTM based on remote sensing and meteorological data to estimate wheat yield. This model extracts spatial features and temporal dependencies from input data, avoiding potential information loss associated with sequential connections. The model achieved a coefficient of determination R^2 = 0.68 and a root mean square error RMSE = 22.97 kg/ha on the test set, demonstrating high prediction accuracy. However, its performance significantly deteriorates under extreme conditions, and its high computational complexity limits large-scale application. *Toledo et al. (2024)*, proposed a multimodal framework integrating heterogeneous multimodal data to predict maize yield. This architecture assigns weights based on pattern characteristics and temporal variations, thereby revealing the dynamic processes of plant growth interacting with the environment. In experiments, the models achieved coefficient of determination R^2 values ranging from 0.82 to 0.93 for yield predictions, demonstrating exceptionally high predictive accuracy. However, model performance remains constrained by data quality, and its complex structure and high computational demands limit deployment. *Jian et al. (2024)*, developed the GCBA hybrid deep learning model by deeply integrating multi-source heterogeneous data and enhancing the GOA and attention mechanisms. This model demonstrated outstanding performance in handling complex time series and diverse remote sensing datasets. In estimating U.S. county-level soybean yields for 2019–2020, this model significantly outperformed five comparison models including SVR, RFR, and CNN, achieving optimal performance with R^2 = 0.7057 and RMSE = 4.4612 bushels/acre, demonstrating high accuracy and stability. However, limitations include suboptimal performance under extreme climatic conditions, structural complexity, and high dependency on high-quality data.

From the dual-source "remote sensing + meteorological" data of *Tian et al. (2024)*, to the deep fusion of "multi-source heterogeneous remote sensing data + photosynthetic parameters" in *Jian et al. (2024)*, and further to the "heterogeneous multimodal data" adopted by *Toledo et al. (2024)*, a clear shift from single-source data to multi-source, multimodal data fusion is evident. Similarly, from the parallel CNN-LSTM architecture (AMCN) of *Tian et al. (2024)*, to the serial coupling of CNN-BiGRU-Attention enhanced by GOA optimization (GCBA) in *Jian et al. (2024)*, and finally to the dynamic weight allocation framework designed for multimodal data in *Toledo et al. (2024)*, this progression illustrates a transition from single-model approaches to multi-model coupling and optimization.

However, it is worth noting that the aforementioned crop yield prediction models generally face challenges such as insufficient generalization capabilities for extreme weather events, high model complexity, and substantial computational costs.

Deep Learning-Based Research on Plant Stress and Related Factors

Crops are frequently affected by biotic and abiotic stresses, such as pests, diseases, climate change, and microbial activity *(Singh et al., 2018)*, leading to reduced yield and quality. Drought and pests are two major constraints on global food production *(Cen et al., 2020)*. By continuously monitoring environmental parameters and combining them with physiological and biochemical indicators of crops, researchers can develop sensitive and accurate stress response models. Table 4 presents the detection of plant pest and disease stress using deep learning techniques.

（1）Drought stress

Water deficit reduces photosynthetic and transpiration rates, inhibits chlorophyll fluorescence, and ultimately affects crop growth and development *(Yin et al., 2024)*. Monitoring crop growth conditions helps breed superior varieties with strong water stress resistance, thereby enhancing crop yields in arid regions. A review of water stress research literature indicates that studies over the past 20 years have primarily focused on canopy temperature, transpiration rate, and chlorophyll content *(Yin et al., 2024)*.

Zhang et al. (2024), developed a multi-task classification model based on infrared and RGB images using 1DCNN to evaluate drought resistance in poplar seedlings. The study demonstrated that when using four PCA principal components as input, the model achieved classification accuracies of 81.8% for variety drought resistance and 62.3% for individual stress level classification, outperforming traditional machine learning methods. However, the accuracy for stress level classification remained significantly lower than that for variety classification, indicating that distinguishing varying degrees of stress is more challenging. It should be noted that this study was validated only during the seedling stage of poplar trees, and its applicability to stress monitoring throughout the entire growth cycle remains unknown. *Yao et al. (2024),* proposed a drought stress monitoring model for key growth stages of winter wheat based on DenseNet121. The study collected drought stress images across three critical growth stages. By optimizing training strategies, adjusting learning rates, and incorporating attention mechanisms, the model achieved an average recognition accuracy of 94.67%, validating its effectiveness. However, since drought stress is a continuous physiological process, the phenotypic boundaries between adjacent stress levels are blurred, leading to misclassification of these adjacent levels by the model. *Wang L. et al. (2024),* compared the performance of LSTM, ResNet18-LSTM, and ResNet18-CBAM-LSTM models in dual-task classification of poplar seedling varieties and drought severity using multi-source temporal data. Among these, the ResNet18-CBAM-LSTM model demonstrated optimal performance, achieving a drought severity classification accuracy of 90.94%. This enabled continuous monitoring of the dynamic responses of multiple poplar seedling varieties under drought stress. However, feature extraction relies on manually designed image processing workflows and is sensitive to image quality.

Table 4

Application of Deep Learning Technology for Plant Pest and Disease Stress Detection

Purpose	Data set	Data preprocessing	Data enhancement	model	Effect	Comparison Algorithm	Reference
Field aphid detection and Identification	Self-built aphid image collection	Mark aphids and their densely populated areas	Image rotation, flipping	Two-stage CFN model	The average accuracy rate is 76.8%.	Faster R-CNN, DSSD, R-FCN, FPN	*Li et al. 2019*
Accurate classification of pest Images	D0, SMALL, IP102 datasets	Input image size normalization	Image rotation, random scaling, mirror flipping	GAEnsemble	Accuracy: 98.81%	VGG-16, VGG-19, and others	*Ayan et al. 2020*
Detection of cassava disease	Cassava Image Dataset, Cassava Plant Disease Consolidated (2019-2020) Dataset	Combined case	Angle rotation, saturation adjustment, and contrast enhancement	CDDNet	All models achieved over 97% accuracy	VGG19, VGG16, and others	*Dosset et al. 2024*
Prediction of fruit tree disease extent	2,010 images of lobar disease	Harmonize image sizes and tag assignments	Random noise, blurring operations, etc.	DINOV2-FCS	Predictive accuracy 95.68%	FCN, Deeplabv3+, and others	*Bai et al. 2024*
Quantifying the extent of tar spots	254 RGB pictures	Background removal, homogenization, gaussian filtering, etc.	Not applicable	SCDA v2	Accuracy: 73.7%	SCDA v1	*Lee et al. 2024*

The aforementioned studies each have distinct focuses: *Zhang et al. (2024)* emphasize the effectiveness of multi-task learning and integrating biological prior knowledge; *Yao et al. (2024)* concentrate on enhancing recognition accuracy during critical periods through advanced CNN architectures and training techniques; *Wang L. et al. (2024)* focus on achieving continuous dynamic monitoring using temporal models.

Future research should continue to advance in the areas of temporal modeling, multimodal data fusion, and addressing category ambiguity to develop more robust and precise stress monitoring systems.

（2）Pest and disease stresses

Globally, food production losses due to diseases and pests account for approximately 14% and 10% of total production, respectively *(Huang et al., 2018)*. To enable large-scale pest detection, methods are gradually shifting towards artificial intelligence-based approaches. To enable large-scale pest detection, methods are gradually shifting towards artificial intelligence-based approaches. Pest recognition and detection methods using deep learning techniques can be categorized into three types: feature-optimized detection, attention-enhanced feature detection, and network tuning detection. Figure 5 illustrates a schematic diagram of various plant pests.

To address the issue of low detection accuracy for aphids in fields, *Li et al. (2019)*, designed a two-stage aphid detector employing a “coarse-to-fine network” architecture. This model first utilizes a coarse convolutional neural network (CCNN) to search for aphid clusters, then employs a fine convolutional neural network (FCNN) to precisely identify aphid regions, ultimately achieving an average detection accuracy of 76.8%. However, due to the minuscule size of aphid targets—occupying only 1.5% of the image—feature loss after multi-layer pooling downsampling fundamentally limits further improvements in model performance. *Ayan et al. (2020)*, proposed a deep learning model ensemble approach based on genetic algorithms (GA) to address the challenge of precise classification of agricultural pests. This method utilized GA to automatically optimize the weight distribution of three CNN models (Inception-V3, Xception, MobileNet) within the ensemble (w1=0.509, w2=0.463, w3=0.921), thereby maximizing ensemble performance. Validation accuracy reached up to 98.81% across three public datasets, demonstrating the strategy's effectiveness. However, this approach also faces challenges including high computational complexity, poor interpretability, and limited performance on ultra-large-scale datasets. *Shifeng et al. (2021)*, addressed the challenges of low detection performance due to high similarity among small pest features by proposing the CRA-Net model, which integrates a Channel Recalibration Feature Pyramid Network (CRFPN) with an Adaptive Anchor Module (AA). This approach improved the average detection accuracy to 67.9%. However, for categories with extremely small relative sizes (occupying only about 0.03% of the image), detection accuracy remains low (AP=24.6%), highlighting the technical bottleneck in detecting extremely small objects.

Li et al. (2019) and *Shifeng et al. (2021),* focused on detection tasks, aiming to both identify pest categories and locate their positions. The former provided a dedicated detection solution for specific small targets, while the latter introduced an advanced module to enhance the representation capabilities of general small-object features. The research by *Ayan et al. (2020)* centered on classification tasks—determining the pest category of an entire image—and proposed model optimization strategies that achieved exceptionally high accuracy. Future pest detection research must continue to pursue breakthroughs in underlying network architecture (e.g., preventing loss of small features), computational efficiency, and robustness toward targets at extreme scales.

Convolutional neural networks are widely used for plant disease recognition, extracting deep image features. The network learns features layer by layer, with the bottom layers recognizing basic features and the top layers generating semantically rich advanced features. As shown in Figure 4, the convolutional neural network for disease recognition extracts feature vectors from the input image using convolutional and pooling layers, with disease type predicted by the fully connected layer classifier.

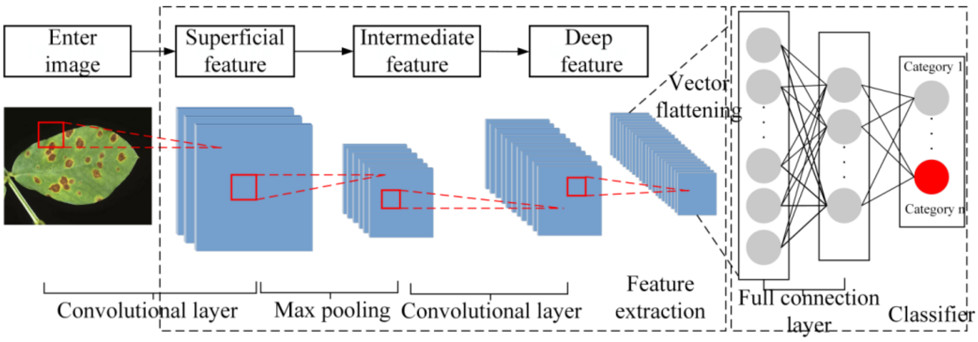


Fig. 4 - Schematic diagram of a convolutional neural network for disease recognition
(*Yu et al., 2024*)

Plant disease recognition is classified into two categories: type recognition and severity assessment. *Too et al. (2019)*, compared the performance of five convolutional networks, including VGG-16, in plant disease recognition. Experiments demonstrated that the DenseNet-121 architecture achieved the highest accuracy of 99.75% in testing. However, the model's generalization capability in real field environments remains questionable, and its analysis of class imbalance and rare class samples is insufficient. *Dosset et al. (2024),* proposed an efficient, lightweight framework named CDDNet for detecting and classifying cassava leaf diseases. This framework achieved over 97% accuracy across various datasets in real-time environments. However, the model exhibits limited environmental adaptability and struggles to process images in complex settings. Its processing speed on devices reaches only 9.76 FPS, falling short of meeting the demands for high-frame-rate real-time processing. *Too et al. (2019)* prioritize accuracy, exploring the performance ceiling of deep models under ideal conditions. *Dosset et al. (2024)* emphasize application-first approaches, advancing technology toward practical implementation. Through lightweight design and attention mechanisms, they achieve a favorable balance between accuracy and efficiency.

Identifying disease severity requires precise differentiation of categories, enabling both accurate disease type identification and severity determination. Such models require high-quality datasets. However, leaf images from different diseases often resemble each other, and the severity boundaries within the same disease are unclear, complicating the classification task. Figure 5 shows images of several plant disease types.

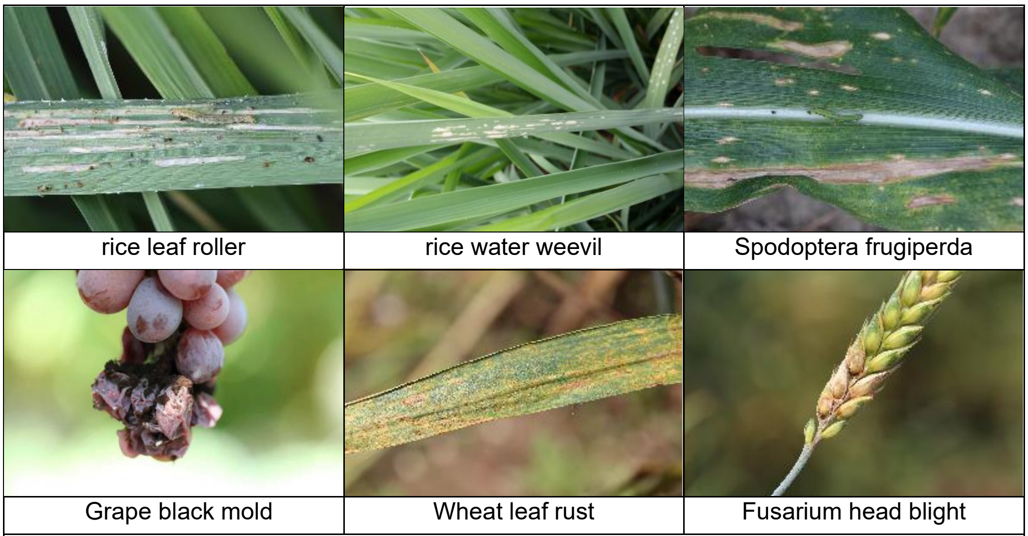


Fig. 5 - IDADP-based schematic diagrams of various plant pests and diseases
(*Hefei Institutes of Physical Science et al., 2006*)

In the task of identifying disease severity, different research teams have adopted approaches with distinct focuses. *Bai et al. (2024)*, focused on enhancing the extraction of subtle features in diseased leaves by improving network architecture. They employed the DINOV2-fruit leaf segmentation model, utilizing DINOv2-B as the backbone feature extraction network, and introduced the Class-Patch Feature Fusion Module (C-PFFM), Explicit Feature Fusion Architecture (EFFA), and Adaptive Kernel Attribute Space Pyramid Pooling (AKASPP). These innovations effectively improved the segmentation performance for various fruit tree leaf diseases. In contrast, *Lee et al. (2024)*, proposed Streamline Contour Detection Algorithm Version 2 (SCDA v2) to quantify the severity of tar spots. This algorithm eliminates the need for empirical optimization of decision input parameters, achieving an overall accuracy of 73.7% in tar spot laminar flow detection. It aims to deliver acceptable accuracy without relying on extensive annotated data, making it more suitable for scenarios with limited annotation resources. While the former may achieve higher precision with ample data, the latter offers greater advantages in practicality and scalability.

Currently, plant pest and disease recognition systems primarily rely on supervised learning, though labeled data remains costly. Future research should prioritize unsupervised learning to reduce data dependency. Additionally, resources should be integrated to create a data-sharing platform, build large-scale, diverse datasets, and develop lightweight recognition models to meet the demands of mobile applications with limited resources and high real-time requirements.

CHALLENGES

Although deep learning methods have achieved significant progress in agricultural fields such as crop morphological feature analysis, crop growth and development prediction, yield forecasting, and drought and pest/disease stress assessment, the following issues remain to be addressed:

(1) Data Requirements and Quality Bottlenecks: The generalization capability of plant phenotyping models relies on large-scale, high-quality, and diverse annotated datasets. Currently, large-scale, high-quality, and diverse public datasets covering different growth stages, seasons, and geographic regions remain scarce. Moreover, capturing plant images is often hindered by organ overlap and occlusion. While 3D point cloud technology partially mitigates this issue, its massive data volume, high computational and storage costs, and susceptibility to data registration errors limit its application in real-time systems.

(2) Model Optimization and Hyperparameter Selection: Deep learning models face significant challenges in practical deployment on resource-constrained mobile platforms, including high memory consumption, slow inference speeds, and large parameter counts. Furthermore, hyperparameters such as learning rate, filter size, and stride exhibit strong interdependencies, making the tuning process heavily reliant on experience and low in automation. This severely impedes model reproducibility and broader application.

(3) Spatio-Temporal Modeling and Dynamic Forecasting: Plant development exhibits irreversible temporal and phasic sequences. Current plant growth monitoring primarily relies on static or short-cycle image analysis, failing to adequately capture spatiotemporal dynamics across growth cycles. To achieve accurate plant growth prediction, challenges in long-term growth forecasting and multi-stage phenotypic correlation analysis must be addressed to support precise modeling and systematic analysis of plant growth processes.

DISCUSSION AND OUTLOOK

To address these challenges, future research should focus on synergistic innovation across three levels—data, models, and systems—to advance plant phenotyping toward standardization, intelligence, and practical application:

(1) Building Multimodal Standardized Datasets and Data Generation Methods: Integrate domestic and international resources to establish an open-source phenotypic dataset covering major crops, multiple growth stages, and diverse environmental conditions, while formulating unified annotation standards. Combine generative adversarial networks (GANs) and diffusion models to synthesize high-quality samples, thereby enhancing model robustness under imbalanced and occluded conditions.

(2) Model Lightweighting and Adaptive Optimization: To address platform computing and real-time processing demands under limited resources, research should focus on lightweight techniques such as neural network pruning, quantization, and knowledge distillation to build embedded models that balance accuracy and efficiency. For hyperparameter optimization, automated tuning strategies like meta-learning should be adopted to enhance model training efficiency and reproducibility.

(3) Develop cross-stage spatio-temporal forecasting and universal models: To capture plant phenotyping data throughout the entire growth cycle, a spatiotemporal hybrid network integrating temporal imaging with environmental factors should be constructed to enable early diagnosis and prediction of growth trends and stress responses. Research should explore cross-crop, cross-task universal pre-trained models, leveraging transfer learning to enhance model adaptability across diverse scenarios and reduce redundant modeling efforts.

(4) Promote the integrated development of learning and systems: In complex field environments, a single model struggles to address all challenges. Therefore, we should develop multi-model fusion networks based on ensemble learning to fully leverage the complementary advantages of multiple models and build robust, interpretable analytical systems. Moving forward, plant phenotyping analysis systems should gradually evolve toward platformization and cloud-edge collaboration, providing integrated solutions for precision agriculture and smart breeding.

ACKNOWLEDGEMENT

We greatly appreciate the careful and precise reviews by the anonymous reviewers and editors. This work was supported by the Inner Mongolia Natural Science Foundation (Grant No. 2024MS06016), the Basic Research Operating Expenses Project for Higher Education Institutions (Grant No. BR241401), and the Basic Research Operating Expenses Project for Higher Education Institutions (Grant No. BR230124).

REFERENCES

[1] Aboalia, H., Hussein, S., & Mahmoud, A. (2024). Enhancing Power Lines Detection Using Deep Learning and Feature-Level Fusion of Infrared and Visible Light Images. *Arabian Journal for Science and Engineering*(prepublish), 1-13.

[2] Ali, B., Karim, F., & Ali, M. (2022). SWP-LeafNET: A novel multistage approach for plant leaf identification based on deep CNN. *Expert Systems With Applications, 202*.

[3] Almoujahed, M. B., Apolo, O. E. A., Whetton, R. L., Kazlauskas, M., Kriaučiūnienė, Z., Šarauskis, E., & Mouazen, A. M. (2025). Field-based hyperspectral imaging for detection and spatial mapping of fusarium head blight in wheat. *European Journal of Agronomy, 164*, 127485-127485.

[4] Ayan, E., Erbay, H., & Var, F. (2020). Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks. *Computers and Electronics in Agriculture, 179*, 105809-.

[5] Bai, C., Zhang, L., Gao, L., Peng, L., Li, P., & Yang, L. (2024). DINOV2-FCS: a model for fruit leaf disease classification and severity prediction. *Frontiers in Plant Science, 15*, 1475282-1475282.

[6] Banah, H., Kurti, P. J. B., Houdinet, G., Hawkes, C. V., & Kudenov, M. (2024). The quantification of southern corn leaf blight disease using deep UV fluorescence spectroscopy and autoencoder anomaly detection techniques. *PloS one, 19*(5), e0301779-e0301779.

[7] Bini, D., Pamela, D., Shajin, P., Elena, P. D., & Jude, H. D. (2021). Recognition of Bloom/Yield in Crop Images Using Deep Learning Models for Smart Agriculture: A Review. *Agronomy, 11*(4), 646-646.

[8] Cao, Y., Zhang, H., Guo, F., Feng, S., Yang, L., & Wei, S. (2024). Research Progress of Crop Disease Monitoring Based on UAV Remote Sensing. *Journal of Shenyang Agricultural University, 55*(05), 616-628.

[9] Cen, H., Zhu, Y., Sun, D., Zhai, L., Wan, L., Ma, Z., Liu, Z., & He, Y. (2020). Current status and future perspective of the application of deep learning in plant phenotype research. *Transactions of the Chinese Society of Agricultural Engineering, 36*(09), 1-16.

[10] Chen, X., Peng, Y., Yu, S., & Hu, D. (2024). Measuring the appearance indicators of spherical fruits using binocular structured light 3D reconstruction. *Transactions of the Chinese Society of Agricultural Engineering, 40*(20), 187-194.

[11] Coswosk, G. G., Gonçalves, V. M. L., Lima, V. J. d., Souza, G. A. R. d., Junior, A. T. d. A., Pereira, M. G., Oliveira, E. C. d., Leite, J. T., Kamphorst, S. H., Oliveira, U. A. d., Crevelari, J. A., Santos, K. D. d., Marques, F. C. R., & Campostrini, E. (2024). Utilizing Visible Band Vegetation Indices from Unmanned Aerial Vehicle Images for Maize Phenotyping. *Remote Sensing, 16*(16), 3015-3015.

[12] Dosset, A., Dang, L. M., Alharbi, F., Habib, S., Alam, N., Park, H. Y., & Moon, H. (2024). Cassava disease detection using a lightweight modified soft attention network. *Pest management science*.

[13] Duan, L., Wang, X., Wang, Z., Geng, Z., Lu, Y., & Yang, W. (2024). Growth visualized prediction method of multi-variety rice based on improved Pix2Pix-HD network. *Acta Agronomica Sinica, 50*(12), 3083-3095.

[14] Eva, P., Andrej, V., Primož, Ž., Matej, K., Nik, S., Saša, Š., Dominik, V., David, L., Janez, L., Uroš, Ž., & Jaka, R. (2023). Early Detection of Wireworm (Coleoptera: Elateridae) Infestation and Drought Stress in Maize Using Hyperspectral Imaging. *Agronomy, 13*(1), 178-178.

[15] Ge, L., Luzhen, G., & Yuzhi, T. (2020). Cauliflower Seedling Recognition Using Stereo Vision and K-Medoids Algorithm under Weedy Conditions. *Journal of Physics: Conference Series, 1621*(1), 012056

[16] Gée, C., Denimal, E., Yparraguirre, M. d., Dujourdy, L., & Voisin, A. S. (2023). Assessment of Nitrogen Nutrition Index of Winter Wheat Canopy from Visible Images for a Dynamic Monitoring of N Requirements. *Remote Sensing, 15*(10).

[17] Gowri, L., Manjula, K. R., Pradeepa, S., & Amirtharajan, R. (2024). Predicting agricultural and meteorological droughts using Holt Winter Conventional 2D-Long Short-Term Memory (HW-Conv2DLSTM). *Environmental monitoring and assessment, 196*(10), 875.

[18] Hefei Institutes of Physical Science, Chinese Academy of Sciences, Institute of Subtropical Agriculture, Chinese Academy of Sciences, & Aerospace Information Research Institute, Chinese Academy of Sciences (2006). Image Dataset for Agricultural Diseases and Pests Research (IDADP) [Data set]. Retrieved from http://www.agriculture-dataset.org/idadp

[19] Hou, Y., Zhou, H., & Wang, Z. (2017). Overview of speech recognition based on deep learning. *Application Research of Computers, 34*(8), 2241-2246.

[20] Huang, W., Liu, L., Dong, Y., Shi, Y., Ye, H., & Zhang, J. (2018). Research progress on crop pest and disease monitoring based on remote sensing technology. *Agricultural Engineering Technology, 38*(09), 39-45.

[21] Jian, L., Hongkun, F., Xuhui, T., Zhao, L., Jujian, H., Wenlong, Z., Hui, C., Yue, S., Xiangyu, N., & Jian, L. (2024). GOA-optimized deep learning for soybean yield estimation using multi-source remote sensing data. *Scientific reports, 14*(1), 7097-7097.

[22] Jiang, L., Li, C., & Fu, L. (2025). Apple tree architectural trait phenotyping with organ-level instance segmentation from point cloud. *Computers and Electronics in Agriculture, 229*, 109708-109708.

[23] Kumar, S. K., Nikos, T., Cornelia, W., & Manuela, Z. (2022). Estimation of Vegetative Growth in Strawberry Plants Using Mobile LiDAR Laser Scanner. *Horticulturae, 8*(2), 90-90.

[24] Lee, D. Y., Na, D. Y., Canul, C. G., Beitia, F. J., Goodwin, S., Cruz, A., Delp, E. J., Acosta, A. G., Lee, J. S., Falconi, C. E., & Cruz, C. D. (2024). Optimizing Corn Tar Spot Measurement: A Deep Learning Approach Using Red-Green-Blue Imaging and Stromata Contour Detection Algorithm for Leaf-Level Disease Severity Analysis. *Plant disease*.

[25] Lei, Z., Wei, J., Boyan, S., Mingtang, C., Saiju, L., & Yimin, D. (2024). Three-dimensional Reconstruction and Phenotype Parameters Acquisition of Seeding Vegetables Based on Neural Radiance Fields. *Transactions of the Chinese Society for Agricultural Machinery, 55*(04), 184-192+230.

[26] Li, L., Li, W., Ma, D., Yang, C., & Meng, F. (2021). Prediction Model of Transpiration of Greenhouse Tomato Based on LSTM. *Transactions of the Chinese Society for Agricultural Machinery, 52*(10), 369-376.

[27] Li, M., Zhou, Q., Han, X., & Lv, P. (2024). Prediction of reference crop evapotranspiration based on improved convolutional neural network (CNN) and long short-term memory network (LSTM) models in Northeast China. *Journal of Hydrology, 645*(PA), 132223-132223.

[28] Li, Q., Jia, W., Sun, M., Hou, S., & Zheng, Y. (2021). A novel green apple segmentation algorithm based on ensemble U-Net under complex orchard environment. *Computers and Electronics in Agriculture, 180*, 105900-.

[29] Li, R., Wang, R., Xie, C., Liu, L., Zhang, J., Wang, F., & Liu, W. (2019). A coarse-to-fine network for aphid recognition and detection in the field. *Biosystems Engineering, 187*(C), 39-52.

[30] Liu, G., Wang, N., An, L., Liu, Y., Sun, H., Li, M., Tang, W., Zhao, R., & Qiao, L. (2024). Crop chlorophyll detection based on multiexcitation fluorescence imaging analysis. *Biosystems Engineering, 246*, 41-53.

[31] Liu, T., Jin, X., Han, K., He, F., Wang, J., Chen, X., Kong, X., & Yu, J. (2024). Semantic segmentation for weed detection in corn. *Pest management science*.

[32] Liu, X., Lu, W., Wang, D., Zheng, J., Peng, Y., & An, K. (2024). Point cloud based phenotypic parameter extraction of Chinese white cabbage. *Journal of Shanghai Normal University(Natural Sciences), 53*(04), 478-488.

[33] Ma, S., Liu, S., Gao, Z., Wang, X., Ma, S., & Wang, S. (2024). Water Deficit Diagnosis of Winter Wheat Based on Thermal Infrared Imaging. *Plants (Basel, Switzerland), 13*(3).

[34] Marchetti, C. F., Ugena, L., Humplík, J. F., Polák, M., Ćavar Zeljković, S., Podlešáková, K., Fürst, T., De Diego, N., & Spíchal, L. (2019). A Novel Image-Based Screening Method to Study Water-Deficit Response and Recovery of Barley Populations Using Canopy Dynamics Phenotyping and Simple Metabolite Profiling. *Frontiers in Plant Science, 10, Article 1252*.

[35] Mhala, P., Bilandani, A., & Sharma, S. (2025). Enhancing crop productivity with fined-tuned deep convolution neural network for Potato leaf disease detection. *Expert Systems With Applications, 267*, 126066-126066.

[36] Murphy, K. M., Ludwig, E., Gutierrez, J., & Gehan, M. A. (2024). Deep Learning in Image-Based Plant Phenotyping. *Annual review of plant biology*.

[37] Nguyen, D., Voil, P. d., Potgieter, A., Dang, Y. P., Orton, T. G., Nguyen, D. T., Nguyen, T. T., & Chapman, S. C. (2025). Multimodal sequential cross-modal transformer for predicting plant available water capacity (PAWC) from time series of weather and crop biological data. *Agricultural Water Management, 307*, 109124-109124.

[38] Rahman, M. H., Busby, S., Ru, S., Hanif, S., Saez, A. S., Zheng, J., & Rehman, T. U. (2025). Transformer-Based hyperspectral image analysis for phenotyping drought tolerance in blueberries. *Computers and Electronics in Agriculture, 228*, 109684-109684.

[39] Rai, C. K., & Pahuja, R. (2024). An ensemble transfer learning-based deep convolution neural network for the detection and classification of diseased cotton leaves and plants. *Multimedia Tools and Applications, 83*(36), 83991-84024.

[40] Ribaut, J. M., Vicente, M. C., & Delannay, X. (2010, Apr). Molecular breeding in developing countries: challenges and perspectives [Review]. *Current Opinion in Plant Biology, 13*(2), 213-218.

[41] Sancho-Adamson, M., Trillas, M. I., Bort, J., Fernandez-Gallego, J. A., & Romanyà, J. (2019). Use of RGB Vegetation Indexes in Assessing Early Effects of ;Verticillium; Wilt of Olive in Asymptomatic Plants in High and Low Fertility Scenarios. *Remote Sensing, 11*(6), 607.

[42] Sashuang, S., Lijun, Z., Ning, L., Yiyin, H., Zhao, W., Si, C., Jiangang, L., Haiyan, C., Yong, H., & Zhenjiang, Z. (2023). Monitoring drought induced photosynthetic and fluorescent variations of potatoes by visible and thermal imaging analysis. *Computers and Electronics in Agriculture, 215*.

[43] Seyam, A. T., Barua, J. C., & Abhijit, P. (2024). Comparing pre-trained models for efficient leaf disease detection: a study on custom CNN. *Journal of Electrical Systems and Information Technology, 11*(1).

[44] Shifeng, D., Rujing, W., Kang, L., Lin, J., Rui, L., Jianming, D., Yue, T., & Fenmei, W. (2021). CRA-Net: A channel recalibration feature pyramid network for detecting small pests. *Computers and Electronics in Agriculture, 191*.

[45] Si, B., Jing, M., Jiang, N., He, Z., Han, H., & Chen, M. (2023). Analysis of Drought Stress in Tea Tree Seedlings Based on Slope Fluorescence Index. *Optical Technique, 49*(06), 685-691.

[46] Singh, A. K., Ganapathysubramanian, B., Sarkar, S., & Singh, A. (2018). Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives. *Trends in Plant Science, 23*(10), 883-898.

[47] Syuen, T. Y., Hong, L. T., & Sam, Y. K. (2022). Deep Learning for Image-Based Plant Growth Monitoring: A Review. *International Journal of Engineering and Technology Innovation, 12*(3), 225-246.

[48] Tester, M., & Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science*, *327*(5967), 818-822.

[49] Tian, H., Wang, P., Tansey, K., Wang, J., Quan, W., & Liu, J. (2024). Attention mechanism-based deep learning approach for wheat yield estimation and uncertainty analysis from remotely sensed variables. *Agricultural and Forest Meteorology, 356*, 110183-110183.

[50] Toledo, C. A., Crawford, M. M., & Tuinstra, M. R. (2024). Integrating multi-modal remote sensing, deep learning, and attention mechanisms for yield prediction in plant breeding experiments. *Frontiers in Plant Science, 15*, 1408047-1408047.

[51] Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture, 161*, 272-279.

[52] Tunca, E., Köksal, E. S., Taner, S. Ç., & Akay, H. (2024). Crop height estimation of sorghum from high resolution multispectral images using the structure from motion (SfM) algorithm. *International Journal of Environmental Science and Technology, 21*(2), 1981-1992.

[53] Vásconez, J. P., Vásconez, I. N., Moya, V., Díaz, M. J. C., Valenzuela, M., Besoain, X., Seeger, M., & Cheein, F. A. (2024). Deep learning-based classification of visual symptoms of bacterial wilt disease caused by Ralstonia solanacearum in tomato plants. *Computers and Electronics in Agriculture, 227*(P2), 109617-109617.

[54] Wang, C., Pan, W., Li, X., & Liu, P. (2022). Plant Growth and Development Prediction Model Based on ST-LSTM. *Transactions of the Chinese Society for Agricultural Machinery, 53*(06), 250-258.

[55] Wang, H., & Cao, L. (2021). Automatic Segmentation and Phenotypic Analysis of Plant Leaves Based on Fully Convolutional Networks. *Journal of Chinese Agricultural Mechanization, 42*(08), 161-168.

[56] Wang, H., Yao, Y., Ye, Z., Chang, W., Liu, J., Zhao, Y., Li, S., Liu, Z., & Zhang, X. (2024). Solution for crop classification in regions with limited labeled samples: deep learning and transfer learning. *GIScience & Remote Sensing, 61*(1).

[57] Wang, L., Zhang, H., Bian, L., Zhou, L., Wang, S., & Ge, Y. (2024). Poplar seedling varieties and drought stress classification based on multi-source, time-series data and deep learning. *Industrial Crops & Products, 218*, 118905-.

[58] Wang, R., Liu, D., Wang, X., & Yang, H. (2022). Segmentation and measurement of key phenotype for Chinese cabbage sprout using multi-view geometry. *Transactions of the Chinese Society of Agricultural Engineering, 38*(16), 243-251.

[59] Wang, Y., Xiao, H., Yin, H., Luo, S., Le, Y., & Wan, J. (2022). Measurement of morphology of oudemansiella raphanipies based on RGBD camera. *Transactions of the Chinese Society of Agricultural Engineering, 38*(20), 140-148.

[60] Wang, Y., Yi, Z., Tan, W., Guo, J., Zhou, X., & Zhao, J. (2024). Multi Perspective Point Cloud Reconstruction Method for Sweet Pepper Fruit under Occlusion Conditions. *Transactions of the Chinese Society for Agricultural Machinery, 55*(05), 218-225.

[61] Zhou, J., Tardieu, F., Pridmore, T., Doonan, J., Reynolds, D., Hall, N., Griffiths, S., Cheng, T., Zhu, Y., Wang, X., Jiang, D., & Ding, Y. (2018). Plant phenomics: history, present status and challenges. *Journal of Nanjing Agricultural University*, *41*(4), 580-588.

[62] Yang, H. C., Zhou, J. P., Zheng, C., Wu, Z., Li, Y., & Li, L. G. (2025). PhenologyNet: A fine-grained approach for crop-phenology classification fusing convolutional neural network and phenotypic similarity. *Computers and Electronics in Agriculture, 229*, 109728-109728.

[63] Yang, S., Ge, Y., Wang, J., Liu, R., & Fu, L. (2024). Predicting alfalfa leaf area index by non-linear models and deep learning models. *Frontiers in Plant Science, 15*, 1458337-1458337.

[64] Yao, J., Wu, Y., Liu, J., & Wang, H. (2024). Research on Drought Stress Monitoring of Winter Wheat during Critical Growth Stages Based on Improved DenseNet-121. *Applied Sciences, 14*(16), 7078-7078.

[65] Yin, X., Wang, H., Wang, C., Wang, S., Ren, X., & Zong, Z. (2024). Effect of Watering Gradient on Chlorophyll Fluorescence Parameters of Prunus domestica Trees. *Agronomy, 14*(10), 2373-2373.

[66] Yin, X., Wang, S., Wang, C., Wang, H., Zong, Z., & Ban, Z. (2024). Research Progress on Economic Forest Water Stress Based on Bibliometrics and Knowledge Graph. Phyton, 93(5), 843-858.

[67] Yu, S., Hu, D., & Liu, W. (2024). Vision-Based Optical 3D Reconstruction and the Application in Crop Information Perception. *Laser & Optoelectronics Progress, 61*(04), 49-57.

[68] Yuan, X., Yu, H., Geng, T., Ma, R., & Li, P. (2024). Enhancing sustainable Chinese cabbage production: a comparative analysis of multispectral image instance segmentation techniques. *Frontiers in Sustainable Food Systems, 8*, 1433701-1433701.

[69] Zeng, A., Luo, L., Pan, D., Xian, Z., Jiang, X., Xian, Y., & Liu, L. (2023). Multi-task segmentation network for the plant on 3D point cloud. *Transactions of the Chinese Society of Agricultural Engineering, 39*(12), 132-140.

[70] Zhang, H., Zhou, Z., Bian, L., Zhou, L., Zou, Y., & Tian, Y. (2024). Assessment of plant drought stress level based on 1DCNN fusion of multi-source phenotypic data. *Transactions of the Chinese Society for Agricultural Machinery*, 1-15.

[71] Zhang, R., Bilige, S., & Chaolu, T. (2020). Fractal Solitons, Arbitrary Function Solutions, Exact Periodic Wave and Breathers for a Nonlinear Partial Differential Equation by Using Bilinear Neural Network Method. *Journal of Systems Science and Complexity*(prepublish), 1-18.

[72] Zhang, S., Yang, H., & Du, T. (2022). Evaluating the water status of greenhouse tomatoes using thermal infrared imaging. *Transactions of the Chinese Society of Agricultural Engineering, 38*(18), 229-236.

[73] Zhao, C. (2019). Big Data of Plant Phenomics and Its Research Progress. Journal of Agricultural Big Data, 1(2), 5-18.

[74] Zheng, C., Wen, W., Lu, X., Guo, X., & Zhao, C. (2022). Phenotypic Traits Extraction of Wheat Plants Using 3D Digitization. *Smart Agriculture, 4*(02), 150-162.

[75] Zhou, L., Zhang, H., & Bian, L. (2024). Poplar Leaf Phenotyping Methods Using Zero-shot Deep Learning. *Journal of Forestry Engineering, 9*(06), 152-160.

[76] Zhou, Y., Liu, Z., Deng, H., Miao, T., & Wang, C. (2024). Depth estimation for corn plant images based on hybrid group dilated convolution. *Journal of South China Agricultural University, 45*(02), 280-292.