ITC 2/55

Information Technology and Control

Vol. 55 / No. 2/ 2026

pp. 489-510

DOI 10.5755/j01.itc.55.2.43803

A Multi-Modal Deep Learning Framework for Graded Diagnosis of Lettuce Water-Nitrogen Stress and Pest Infestation

Received 2025/12/09 | Accepted after revision 2026/04/10

HOW TO CITE: Dong, M., Liu, B., Fu, D., Zhao, R. (2026) A Multi-Modal Deep Learning Framework for Graded Diagnosis of Lettuce Water-Nitrogen Stress and Pest Infestation. *Information Technology and Control*, 55(2), 489-510. https://doi.org/10.5755/j01.itc.55.2.43803

A Multi-Modal Deep Learning Framework for Graded Diagnosis of Lettuce Water-Nitrogen Stress and Pest Infestation

Mo Dong, Boqiang Liu, Dan Fu, Ruohan Zhao*

Mudanjiang Medical University, Mudanjiang 157000, Heilongjiang, China

Corresponding author: zhaoruohan@mdjmu.edu.cn

Lettuce is extremely sensitive to the demand for water and nitrogen throughout its growth phase. These elements are crucial for plant photosynthesis and resistance against diseases and pests. When lettuce experiences water or nitrogen stress, it leads to a decline in quality, with the leaves turning yellow, severely affecting its quality, yield, and market value. Therefore, it becomes particularly important to swiftly and accurately identify the water-nitrogen stress status of lettuce and to carry out precise graded diagnosis. This paper explores methods for recognition and graded diagnosis of water-nitrogen stress status in lettuce. We propose a comprehensive deep learning framework that integrates multi-modal data and multi-functional analysis capabilities. Specifically, we construct three specialized deep learning pipelines utilizing phenotypic-, image-, and detection-based methods, with attention mechanisms integrated. The DNN network model established using the Feature Phenotype Dataset (FPD) can efficiently and rapidly accomplish the recognition and graded diagnosis of lettuce water-nitrogen stress status, with an accuracy of 0.8302, providing a feasible basis for monitoring and diagnosing water-nitrogen stress status using non-image data. The ResNet50Evo-SE model, which uses image data combined with the Channel-wise Attention Mechanism (SE), achieves an accuracy of 0.9897 for the recognition and graded diagnosis of water-nitrogen stress status. The YoloV8-CBAM target detection model, which uses image data combined with the Convolutional Block Attention Module (CBAM), can accurately detect predefined visual proxies (e.g., leaf color and physiological symptoms) of water-nitrogen stress; these detected proxies are then combined according to diagnostic rules to infer the stress status, achieving an overall precision of 0.9614, recall of 0.9782, and mAP (mean average precision) of 0.9832 and 0.7665, respectively, on the held-out test set. The YoloV8-CBAM model can also effectively identify and detect lettuce leafminer pest damage, with an early-stage pest detection precision of 0.9443, and recall of 0.9531. These methods provide reliable technical support for optimizing planting management and advancing smart agriculture, contributing to the efficient and high-quality development of precision agriculture.

KEYWORDS: Water-nitrogen Stress, Pest Prevention, Lettuce, Deep Learning, Phenotype

1. Introduction

Biotic and abiotic stresses are major factors that affect crop growth and development, including diseases, pests, light exposure, temperature, climate change, soil structure, microbial activity, and other elements that may trigger biotic or abiotic stress states in crops [13, 31]. Thus, identifying and grading the stress conditions of crops is instructive for improving the comprehensive agricultural productivity as well as the yield and quality of crops [4, 19]. By implementing refined management and intervening with stress factors at the appropriate time, farmers can optimize crop management methods, improve resource utilization efficiency, achieve scientific management of crops, and also effectively reduce the negative impact of agriculture on the environment [1].

Lettuce is highly sensitive to water and nitrogen during its growth [29]. These elements profoundly influence plant photosynthesis and resistance to disease and pests. When lettuce is subjected to water or nitrogen stress, it results in a decrease in quality, with leaf yellowing, which heavily impacts its quality, yield, and market value. Therefore, rapidly and accurately identifying the water-nitrogen stress status of the lettuce and performing accurate graded diagnosis becomes especially important.

Deep learning is a potent instrument capable of modeling highly complex problems and mimicking human neural networks. With the expanding field of applications for deep learning, it is increasingly being employed in the agricultural sector, particularly in the identification of crop stress [3, 30]. Researchers can extract rich information such as crop morphology, color, and texture from images, utilizing computer vision technology to achieve real-time monitoring and scientific assessment of crop growth [14].

This research will implement a series of deep learning models to identify, grade, and analyze the water-nitrogen stress conditions and pest situations in lettuce crops. The outcomes of this study will provide a productive, accurate, and objective monitoring and management approach for agricultural production, which is of practical significance for ensuring the quality and yield of crops, enhancing economic benefits, and also supplying critical technical support for optimizing planting management and advancing smart agriculture, thus contributing to the efficient and quality development of precision agriculture.

2. Related Work

2.1. Advances in Research on Recognition and Diagnosis of Crop Stress

Traditional methods for identifying and diagnosing crop stress primarily rely on collecting and analyzing sensor data, frequently constrained by the capabilities of the collection equipment, and have limited capacity to model the stress response relations for different environmental factors [22, 24]. In recent years, the application of deep learning technology in crop stress has been on the rise. Researchers collect crop images and sensor data, then apply deep learning models for feature extraction and classification, enabling rapid stress detection and prediction.

Recent research has focused on using deep learning models, particularly Convolutional Neural Networks (CNN) and other machine learning algorithms combined with CNN, such as Self-Organizing Maps (SOM), Learning Vector Quantization (LVQ), and Support Vector Machines (SVM), to enhance accuracy and efficiency [8, 18]. Deep learning not only improves crop classification accuracy [2] but also enables superior identification of biotic/abiotic stresses [7]. These models facilitate large-scale crop health monitoring, thereby providing efficient decision-support tools for agricultural management. Furthermore, significant improvements have been made in the automatic recognition and classification of crop health status through enhancement of image datasets, optimization of algorithm structure, and the adoption of transfer learning techniques [12]. Additionally, some studies are dedicated to developing decision-support systems for crop stress [17]. These systems, by integrating crop growth models, environmental data, etc., provide farmers with real-time stress monitoring and decision-making advice, aiding them in better managing the crop production process.

2.1.1. Data-feature-based Networks

Identification and classification based on data features commonly involve classic machine learning methods such as Support Vector Machines (SVM), Random Forests (RF), K-Nearest Neighbors (KNN), and others. In the realm of deep learning, The Deep Neural Network (DNN) is a commonly used network structure for classification based on data features [9]. A DNN consists of a multi-layer neural network that extracts features and performs classification through multiple fully connected layers. DNN models are simple to build, consume fewer computational resources, train quickly, and are more amenable to parameter tuning. Compared to other traditional machine learning models, DNNs can learn deeper representations of data and have a stronger capacity to fit complex nonlinear relationships.

2.1.2. Image-feature-based Networks

Image-feature-based networks can be mainly categorized into two types:

1. **Convolutional Neural Networks (CNN)**

   CNN is a classical network structure in deep learning for processing images. CNNs employ multiple layers of nonlinear transformations through convolutional layers, pooling layers, and fully connected layers, capable of extracting rich and complex features from raw pixel data. Particularly for image classification problems, CNNs leverage end-to-end training to automatically learn cascaded features most relevant to object recognition.

2. **Vision Transformer (ViT)**

   ViT is a relatively new model based on Transformer architecture [5] that divides input images into a series of small patches, processes these patches independently, and utilizes the Transformer's self-attention mechanism to handle the relationships between the patches, thereby capturing global relationships within the image.

   Both CNN and ViT models have their strengths, and both have demonstrated their significant performance in numerous visual tasks. Furthermore, recent advancements in network designs, such as lightweight geometric component transformers and ambiguity-aware cooperative learning, have further enhanced the robustness of these architectures in extracting features from indistinct targets within complex environments [26].

2.1.3. Object Detection Networks

In the field of deep learning, object detection is a crucial computer vision task. Object detection networks integrate image classification and object localization techniques to accurately detect and locate multiple objects in complex scenes. Popular object detection network models include:

1. The R-CNN (Region-based Convolutional Neural Network) series. These models use a Region Proposal Network (RPN) to generate candidate object regions and combine this with convolutional neural networks to classify objects and regress their positions [21].
2. The YOLO [20] (You Only Look Once) series, which divide the image into a grid and use convolutional neural networks to simultaneously predict the classes of objects and bounding boxes within each grid cell.
3. SSD [16], which performs object detection by applying convolutional kernels of various sizes on feature maps at different levels.
4. RetinaNet [15], which addresses the class imbalance issue in object detection by using an FPN to extract features of different scales and by designing efficient classification and regression heads.

These object detection networks have their own characteristics and performance in different application scenarios and have made significant advancements in terms of accuracy, speed, and efficiency in object detection.

2.2. Problems in Current Research

1. There are few applications that rely on non-image information for the recognition and diagnostic grading of stress states, and their accuracy cannot be guaranteed.
2. Although existing datasets can provide a good level of accuracy for diagnosing the stress states of crops in specific environments, models often fail to maintain their original performance when exposed to new environmental variables.
3. Current stress diagnosis models are mostly focused on the recognition of single stress states, and for complex agricultural scenarios, such as the coexistence of multiple stresses, the models' judgment criteria and categories are not sufficient to comprehensively cover the actual requirements.

2.3. Motivations and Contributions

Building on the current research status and issues, this paper targets lettuce as the research subject and adopts common nutritional stress during the lettuce growth process as experimental conditions. By collecting RGB images throughout the complete growth cycle of lettuce, this study develops several deep learning-based computational models to address key issues during lettuce growth. It involves extracting commonly used phenotypic indicators to achieve recognition, grading diagnosis of water-nitrogen stress conditions, and early prevention of pest infestations in lettuce. Ultimately, the study accomplishes intelligent monitoring and management throughout the entire lifespan of lettuce, providing technical support for its high-quality and efficient production.

3. Experimental Environment and Methods

The experimental setup was consistent with our prior study [27], but the detailed parameters are listed below for completeness.

3.1. Greenhouse Conditions

The experiment was conducted in a greenhouse facility located at 44°50′N, 125°18′E. The cultivation period lasted from March to June 2023. Throughout the growing season, environmental parameters were strictly controlled to ensure consistency. The temperature was maintained at 25°C during the day and 15°C at night, relative humidity was kept between 40% and 60%, and the photoperiod was set to 12 hours of light per day.

3.2. Water-Nitrogen Stress Design

To investigate the effects of varying water and nitrogen levels, a five-level gradient system was established. The nitrogen source was urea (46% nitrogen content). The plants were divided into five groups (A–E), representing different stress intensities. Irrigation was controlled using specific intervals ranging from 0 to 4 days to induce varying water stress levels. The five nitrogen gradients and five moisture gradients established in this study are presented in Tables 1-2, respectively.

Table 2
Irrigation schemes under different moisture gradient levels.

Moisture Gradient	Irrigation Interval (days)
WA	0
WB	1
WC	2
WD	3
WE	4

3.3. Image Acquisition and Processing

3.3.1. Image Acquisition

The RGB image dataset used in this study was constructed in-house. The imaging system and protocol remained unchanged from [27], featuring:

- 12MP RGB camera (3024×4032 pixels JPG)
- Black-background shooting environment
- Daily 8-angle capture protocol (2 top-down + 6 side-views)

3.3.2. Image Processing

1 **Image Augmentation**

Image augmentation is a common practice in constructing image datasets, used to improve model generalization and robustness. In this study, we adopted a targeted augmentation strategy rather than augmenting the entire dataset. Given that the

Table 1
Fertilization schemes under different nitrogen gradient levels.

Nitrogen Gradient	Nitrogen Application Rate (g/m^2)	Urea (g/pot)	Base Fertilizer (g/pot)	1st Additional Fertilizer (g/pot)	2nd Additional Fertilizer (g/pot)
NA	60	1.860	0.930	0.465	0.465
NB	45	1.400	0.700	0.350	0.350
NC	30	0.933	0.466	0.233	0.233
ND	15	0.467	0.234	0.117	0.117
NE	0	0	0	0	0

original dataset of approximately 37,000 multi-angle RGB images had relatively low background noise and a sufficient sample size, a full augmentation was unnecessary. Instead, we randomly selected one-fifth of the images for augmentation (generating ~7,400 enhanced images). This specific ratio was chosen to balance the need for feature diversity with the preservation of the original data distribution, preventing the introduction of potential artifacts that could bias the model's understanding of real-world phenotypic features.

2. **Image Cropping**

   Cropping the images serves to adjust them to specific sizes to accommodate the requirements of various deep learning neural network architectures. In this study, two different image sizes were adjusted, as shown in Table 3.

Table 3

Image sizes used for training deep learning models.

Image size	Models
288×288	CNN/Vision Transformer
640×640	YoloV8

4. Graded Diagnostic Methods for Lettuce Stress Identification

4.1. Identification of Phenotypic Indicators

Building upon our validated framework [27], we extracted 45 phenotypic traits covering:

- **Geometry:** 12 parameters including canopy projection and convex hull features
- **Color:** 28 indices (RGB/HSV statistics + 10 vegetation indices)
- **Texture:** 5 descriptors with improved fractal dimension calculation

Current experimental images are shown in Figure 1.

4.2. Model Construction Based on Phenotypic Features

4.2.1. Dataset Construction

To avoid train–test leakage, the dataset was strictly split at the plant level. All images, daily records, and multi-angle views of the same plant were assigned to only one subset. A total of 120 lettuce plants were randomly divided into three groups: 72 plants for training (2,520 samples), 12 plants for validation (420 samples), and 36 plants for testing (1,260 samples).

Figure 1

Lettuce images from multiple angles.

Feature selection was performed exclusively on the training set to prevent data leakage. A correlation analysis was conducted on the 45 phenotypic indicators, retaining features with a correlation coefficient $r > 0.9$ and removing highly redundant variables. Cluster analysis and principal component analysis (PCA) were then applied only on the training set to identify the most representative and stable traits.

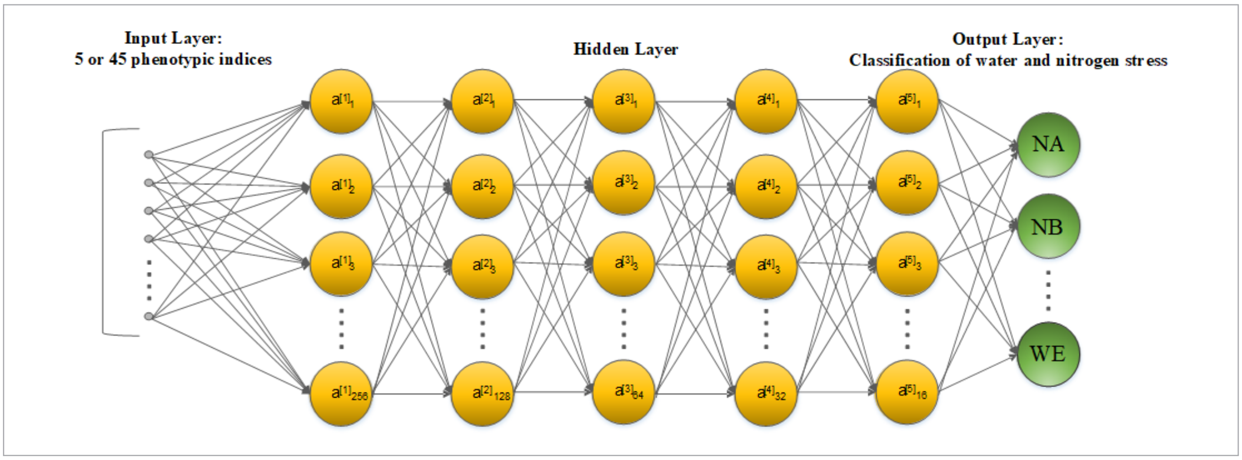
After screening, five key phenotypic indicators were retained: plant height, area of the circumscribed convex polygon, density of leaf gap in the circumscribed rectangle, canopy perimeter, and GLI. These indicators were selected using solely training data, ensuring no information leakage during feature selection.

Based on these selected features, the Feature Phenotype Dataset (FPD) was constructed using the five key indicators. The Complete Phenotype Dataset (CPD) retained all 45 phenotypic indicators for reference and comparison.

4.2.2. Network Structure

In this study, a Deep Neural Network (DNN) was constructed to utilize phenotypic indicators for feature extraction and to classify water-nitrogen stress in lettuce. The lettuce phenotypic indicator set contains 45 feature variables with complex internal correlation structures; hence, selecting the DNN model can achieve better classification performance. To handle the nonlinear correlations within these complex indicators, non-linear activation functions (ReLU) are employed in the hidden layers, enabling the model to learn intricate patterns. For feature selection, instead of relying on external filtering methods that might discard information, we utilized the DNN's inherent

Figure 2
Architecture of DNN.



capability to perform implicit feature weighting. This allows the model to effectively distinguish significant features from noise without compromising the integrity of the phenotypic dataset.

Although the DNN model has a simple network structure, constructing a network model suitable for specific tasks requires repeated experiments. This is to ensure a balance between extracting deep features and preventing overfitting when building the DNN model. After numerous experiments, a six-layer deep DNN model was finally constructed, with the network structure illustrated in Figure 2.

4.2.3. Model Training and Performance Evaluation

The model training employed the Adam optimizer with a cross-entropy loss function to calculate loss, a learning rate set to 0.001, and a batch size of 128 for a total of 10,000 training epochs. For both CPD and FPD datasets, the aforementioned training method was used, and the validation set loss curve is shown in Figure 3.

Analysis from Figure 3 reveals that after 10,000 training epochs, the model trained with FPD had a lower loss value on the validation set compared to the model trained with CPD.

To precisely evaluate the performance of the DNN models constructed with FPD and CPD datasets for classifying water-nitrogen stress in lettuce, five evaluation metrics: Accuracy, Precision, Recall, F1 score, and mean AUC were used for comprehensive assessment. The performance of the models trained on the two datasets on the test set is presented in Table 4.

Analysis of Table 4 indicates that the DNN model trained with FPD outperforms the one trained with CPD across all metrics for the recognition and classification of water-nitrogen stress. Investigating the

Figure 3
Loss curves of the models trained on CPD and FPD.

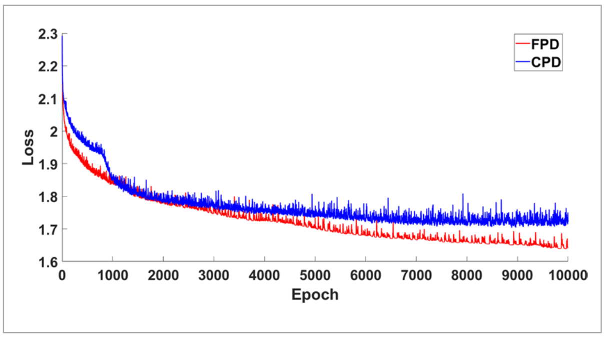
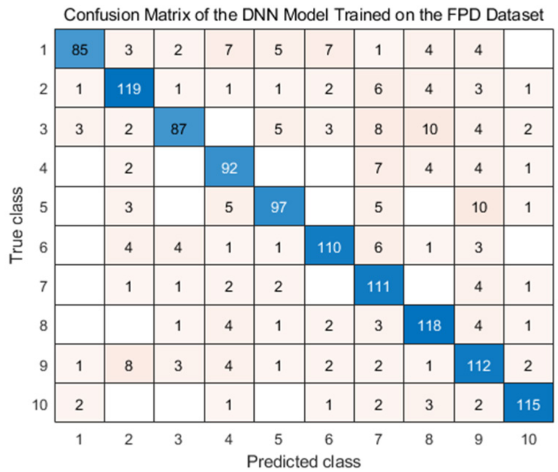


Table 4
Performance comparison of the models trained on CPD and FPD.

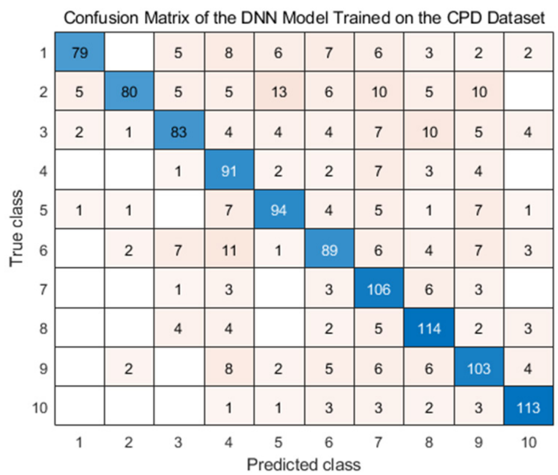
Datasets	Accuracy	Precision	Recall	F1 score	Mean AUC
CPD	0.7556	0.7731	0.7556	0.7546	0.8652
FPD	0.8302	0.8370	0.8302	0.8300	0.9049

Figure 4

Confusion matrix of the DNN model trained on FPD and CPD.



(a) FPD



(b) CPD

Figure 5

Images of lettuce from the ND and WB groups.

(a) ND (b) WB

reasons, firstly, the FPD dataset is simpler. With only five phenotypic indicators, the model requires much less information to create effective decision boundaries, making it easier for the model to understand and learn the data patterns, thereby reducing the risk of overfitting. Secondly, the 45 phenotypic indicators come with more noise and data diversity, which can confound the model's learning, leading to excessive adaptation to the additional phenotypic features during training. Moreover, from the perspective of training efficiency, the size and complexity of the training set are significant factors. A training set with five phenotypic indicators requires processing less information, which can enhance training efficiency, speed up convergence, and ultimately result in lower loss values and higher accuracy.

The accuracy of the model was further analyzed using a confusion matrix, as shown in Figure 4. In the confusion matrix, the numeric indices 0-9 on the axes correspond to the following experimental groups: 'NA', 'NB', 'NC', 'ND', 'NE', 'WA', 'WB', 'WC', 'WD', 'WE', respectively. The results indicate that for both FPD and CPD trained DNN models, the recognition performance for the water-deficient group (WB) and the low nitrogen group (ND) is relatively lower compared to other groups. To investigate the underlying causes, we conducted a detailed visual analysis of the raw images. As illustrated in Figure 5, which depicts randomly selected lettuce from the ND and WB groups at 25 days of growth, there are minimal observable differences in leaf morphology and color compared to the control group. This indicates that the feature space of "mild stress" significantly overlaps with that of the normal state. In other words, the lack of distinct visual markers makes it inherently difficult to differentiate these classes. The confusion in the model's prediction correlates with this high visual similarity, demonstrating that the misclassification is largely due to the objective subtlety of the phenotypic changes under mild stress conditions, rather than a defect in the model's decision mechanism.

In summary, using phenotypic indicators enables the classification and identification of water-nitrogen stress in lettuce, and effective characteristic phenotypic indicators can improve the precision of recognition and grading. The classification and diagnosis of water-nitrogen stress based on phenotypic indicators offer advantages such as a simple model structure, high recognition efficiency, and swift performance, providing a feasible basis for stress monitoring and diagnosis via phenotypic indicators. However, a recognition accuracy of 0.8302 leaves room for further improvement; relying solely on a few phenotypic indicators is insufficient for a more precise and comprehensive diagnosis of water-nitrogen stress. Therefore, an image analysis method can be employed, leveraging images of lettuce throughout the growth period, where deep learning models can directly learn the correlation between images and water-nitrogen stress states, thereby facilitating a more precise recognition and classification.

4.3. Model Construction Based on Image Features

Employing DNN for lettuce water-nitrogen recognition and classification yields an accuracy above 0.83. Although using phenotypic indicators for water-nitrogen stress classification and diagnosis is more straightforward and resource-efficient when considering model training difficulty and efficiency, to further improve the accuracy of recognition and classification, we aim to construct a deep learning-based diagnostic model for lettuce water-nitrogen stress from the perspective of extracting image features. The model training dataset consists of images featuring lettuce under all water-nitrogen stress conditions throughout a 35-day growth period, captured from various angles. Following manual selection and the plant-level split described in Section 4.2.1, 24,000 images (72 plants) were used for training, 6,000 (12 plants) for validation, and 6,000 (36 plants) for testing.

4.3.1. Construction of ViT-Based Model

The ViT network model constructed in this study is illustrated in Figure 6.

The ViT model was trained using the Adam optimizer with a learning rate of 0.0001. A cross-entropy loss function was employed to compute the loss. To ensure the reproducibility of the experimental results and eliminate the impact of randomness, fixed random seeds were applied to all experiments, including data loading, weight initialization, and data augmentation. Experiments revealed that the ViT model requires at least 200 training epochs to achieve satisfactory recognition and classification results. This necessity stems from ViT's reliance on Transformer architecture, which is fully grounded in attention mechanisms without the local receptive fields typical of CNN models. This necessitates additional training to establish global dependencies. While ViT can learn global features directly, its lack of hierarchical feature extraction necessitates multiple learning iterations to capture and integrate local features of the input data effectively. The loss curve for the ViT model on the validation set is depicted in Figure 7. Ultimately, after 250 training epochs, the loss stabilized.

Figure 6

Architecture of ViT.

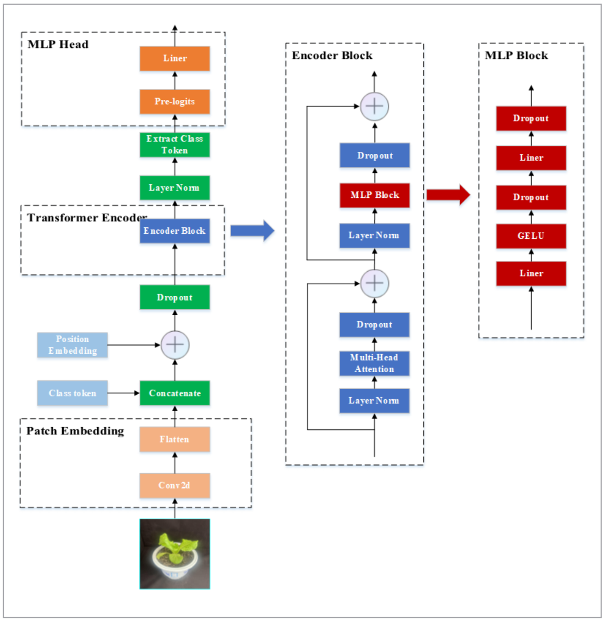
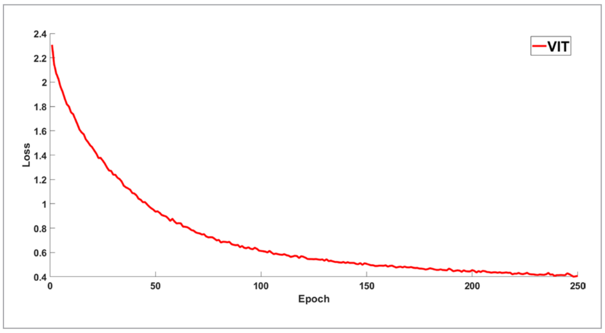


Figure 7

Loss curve of the ViT model on validation set.



4.3.2. Construction of CNN-Based Model

There are numerous image recognition and classification network models based on CNNs. In this study, we selected the Residual Network (ResNet), a convolutional neural network architecture that demonstrates exemplary performance in computer vision tasks. The principal advantage of ResNet lies in its ability to facilitate very deep network structures that provide more nuanced feature extraction capabilities. Using the classic ResNet50 as the backbone, we adopted the ResNeXt bottleneck module method [25] in each residual unit. This module, premised on grouped convolution, enhances the network's expressive power and diversifies feature extraction by increasing the number of groups. Specifically, we configured the grouped convolution with a cardinality of 32, meaning that the input feature maps were divided into 32 groups for independent convolution operations. This approach allows the model to capture a broader and more nuanced set of visual patterns associated with the subtle and varied symptoms of water-nitrogen stress in lettuce leaves, which often exhibit high intra-class similarity.

The model's network structure is shown in Figure 8. Compared to traditional ResNet, this approach introduces greater diversity into the feature extraction process, enhancing the model's performance and representational ability. Appropriate adjustments and optimizations were made to other aspects of the network structure, such as the size of the convolutional kernels, stride, and padding, to further enhance the model's performance. The constructed network model has been designated ResNet50Evo.

To precisely compare the performance between ViT and ResNet50Evo models, the ResNet50Evo model was trained using the same parameters and dataset as the ViT model. Experiments indicated that, in comparison to ViT, the ResNet50Evo model required only 50 training epochs to attain loss convergence on the validation set, as depicted in Figure 9.

Figure 8

Comparison of ResNet and ResNeXt [26].

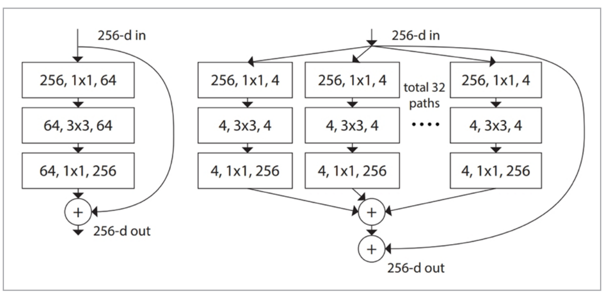
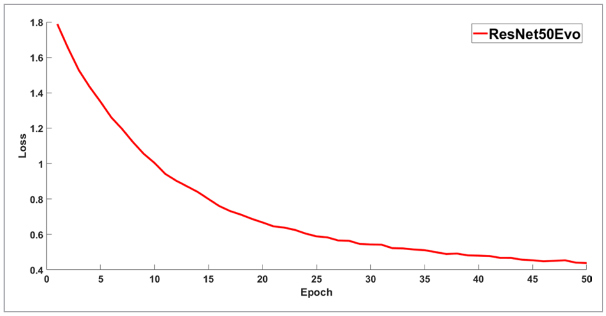


Figure 9

Loss curve of the ResNet50Evo model on validation set



4.3.3. Performance Comparison Between ViT and CNN Models

To thoroughly evaluate the efficacy of the ViT and ResNet50Evo models in the recognition and classification of water-nitrogen stress in lettuce, the performance of the models was compared on a test dataset using the following five metrics: Accuracy, Precision, Recall, F1 score, and mean AUC. Table 5 presents the performance comparison among ViT, ResNet50Evo, and the standard ResNet50. Additionally, a comparison of the model training processes and computational costs for ViT, ResNet50, and ResNet50Evo is presented in Table 6.

Analysis of Tables 5-6 indicates that the ViT model is slightly superior to the ResNet50Evo model in categorizing water-nitrogen stress in lettuce. Even though each training epoch for the ResNet50Evo model requires over three times as long as that for the ViT model, the total number of training epochs for ViT far exceeds that of ResNet50Evo. Consequently, training ViT incurred approximately 3 hours more time than ResNet50Evo for this experiment.

Recently, the performance comparison between ViT and CNN models has been a topic of discussion among many researchers. Some studies have shown that ViT surpasses CNN in many image recognition tasks. However, when considering dataset quality, computing resources, and training efficiency in their entirety, CNNs still maintain specific advantages. Therefore, this study will continue to explore optimizations of the ResNet50Evo model to further enhance its performance.

The design of the ResNet50Evo model represents our first-step architectural innovation tailored for agricultural image analysis. Compared to the standard ResNet50, we specifically employed the ResNeXt bottleneck with a cardinality (number of

Table 5

Performance comparison of the ViT /ResNet50/ResNet50Evo models.

Models	Accuracy	Precision	Recall	F1 score	Mean AUC
ViT	0.9848	0.9851	0.9848	0.9849	0.9874
ResNet50	0.9720	0.9700	0.9714	0.9714	0.9814
ResNet50Evo	0.9775	0.9776	0.9775	0.9775	0.9837

Table 6

Training comparison of the ViT /ResNet50/ResNet50Evo models.

Models	Total Training Epochs	Total Training Time (s)	Average Time per Epoch (s)	Parameters (M)	Inference Time (ms/img)
ViT	250	26880	107.5	86.4	14.5
ResNet50	50	16035	320.7	25.8	6.2
ResNet50Evo	50	16120	322.4	26.4	6.4

groups) of 32. This grouped convolution strategy diversifies the feature extraction pathways within the network without significantly increasing computational complexity. Theoretically, this enhanced feature diversity is crucial for our task, as it enables the model to capture a broader and more nuanced set of visual patterns associated with the subtle and varied symptoms of water-nitrogen stress in lettuce leaves, which often exhibit high intra-class similarity.

4.3.4. Optimization with the Introduction of the SE Attention Mechanism

Compared to the CNN model, the ViT model has the significant advantage of employing a pure attention mechanism from the Transformer, abandoning convolutional operations inherent to the CNN model entirely. This allows the ViT model to establish global dependency relations, unrestricted by the limitations of local convolutional operations. Attention mechanisms excel at learning global contextual information, which is beneficial for processing long-range dependencies in images. Theoretically, the model's performance could be improved by incorporating certain attention mechanism modules into the ResNet50Evo model. In light of this perspective, the current study made the following optimization adjustments: the Squeeze-and-Excitation (SE) module [10] was added to the architectural foundation of the ResNet50Evo model. The principal function of the SE module is to introduce dynamic channel weight adjustments into the ResNet50Evo model, thereby enhancing the model's sensitivity to inter-channel relationships. The SE module allocates divergent weights to feature maps across different channels, strengthening the representation of key information required and accentuating areas of interest within the image.

By embedding the SE module within the forward propagation function of the ResNet50Evo model, it has been integrated with other layers of the model. The enhanced model, now incorporating the SE module, is defined as ResNet50Evo-SE. The architecture of the ResNet network post-introduction of the SE module is illustrated in Figure 10.

The same parameters used for the ViT and ResNet50Evo models have been selected to train the ResNet50Evo-SE model. The training comprised a total of 50 epochs. The ResNet50Evo-SE model was evaluated using the test set, as indicated in Table 7.

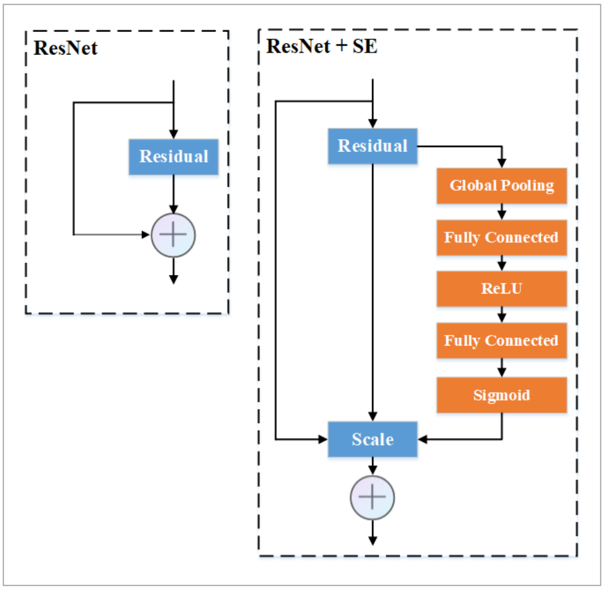
Table 7

Performance comparison of all models on test set.

Models	Accuracy	Precision	Recall	F1 score	Mean AUC
ViT	0.9848	0.9851	0.9848	0.9849	0.9874
ResNet50Evo	0.9775	0.9776	0.9775	0.9775	0.9837
ResNet50Evo-SE	0.9897	0.9898	0.9896	0.9897	0.9898

Figure 10

Schematic diagram of the integration of the ResNet and SE module.

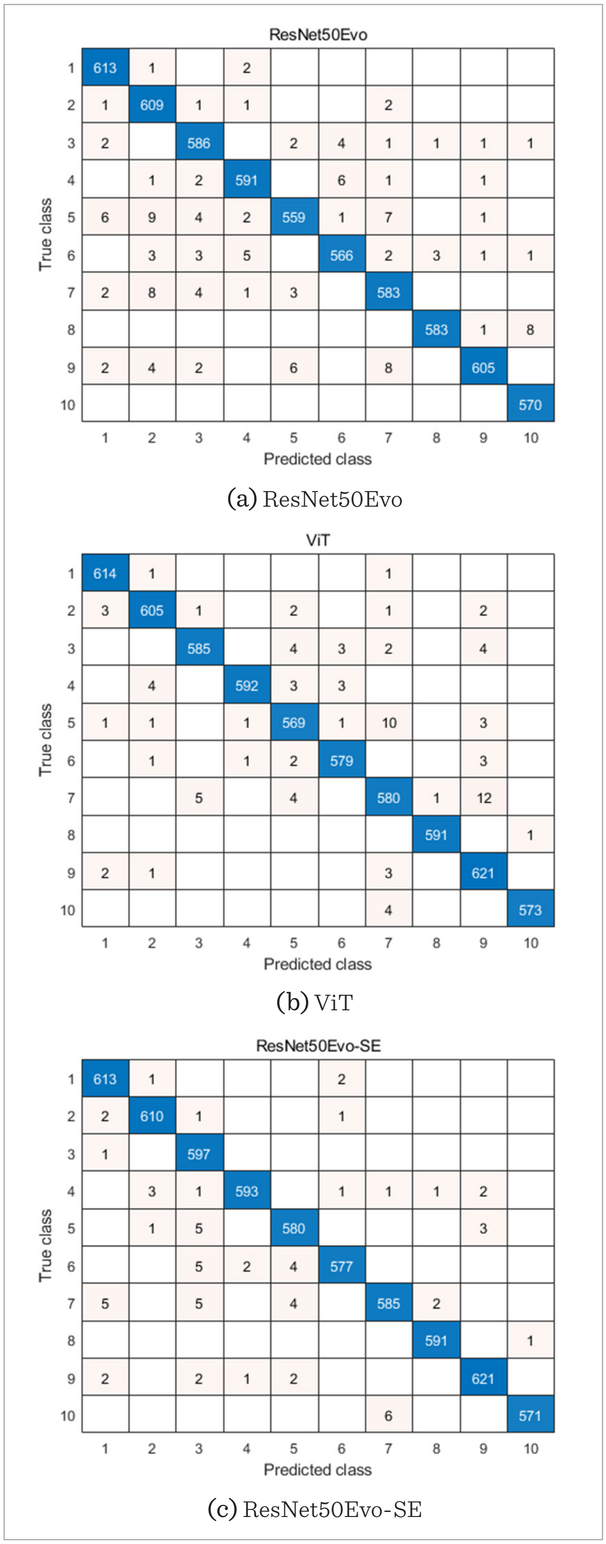


Analysis from the table reveals that the ResNet50Evo-SE model, augmented with an attention mechanism, improved on all evaluation metrics compared to the original ResNet50Evo and ViT models. Specifically, accuracy increased by 1.25% compared to the original ResNet50Evo model, and by 0.49% compared to the ViT model. This demonstrates that the application of attention mechanisms has indeed contributed to the enhancement of CNN models. Although the improvement over the ViT model is not substantial, the ResNet50Evo-SE model is superior in terms of training time and computational resource consumption. This further elaborates that on different image recognition tasks, both ViT and CNN models maintain their respective advantages, and in certain tasks, such as the classification and diagnostic work of lettuce water-nitrogen stress recognition in this study, the CNN model proves more effective.

The computation of the confusion matrices for the aforementioned three models in the test set continues, as depicted in Figure 11. In the confusion matrix, the numeric indices 0-9 on the axes correspond to the following experimental groups: ‘NA’, ‘NB’, ‘NC’, ‘ND’, ‘NE’, ‘WA’, ‘WB’, ‘WC’, ‘WD’, ‘WE’, respectively. To provide a more detailed performance analysis, we have also computed the precision, recall, and F1-score per class for each model, as shown in Figure 12.

Figure 11

Confusion matrices of the three model.



(a) ResNet50Evo

(b) ViT

(c) ResNet50Evo-SE

The figure presents a comprehensive comparison of the per-class performance metrics across the three models: VIT (gray bars), ResNet50Evo (blue bars), and our proposed ResNet50Evo-SE (red bars). These results further highlight the effectiveness of our proposed ResNet50Evo-SE model in addressing the challenges posed by the high similarity of lettuce images under different stress conditions. The improved precision, recall, and F1-score per class suggest that our model can better distinguish between subtle variations in the visual features of lettuce plants, thereby enhancing the accuracy of classification.

Furthermore, analyses of the chart reveal that regardless of using the CNN or ViT models, misclassifications are predominantly concentrated in the groups WB/WC/ND, which correspond to lettuce with slight nitrogen and water deficiencies, respectively. This result is analogous to the classification of water-nitrogen stress states using a DNN model based on phenotypic indicators. The lettuces in these images exhibit a high degree of similarity concerning their color, shape, and other characteristics, which further underscores the existing limitations of deep neural network models based on image features. Despite these challenges, our proposed ResNet50Evo-SE model demonstrates superior performance, indicating its potential for improving classification accuracy in similar scenarios.

4.4. Model Construction Based on Object Detection

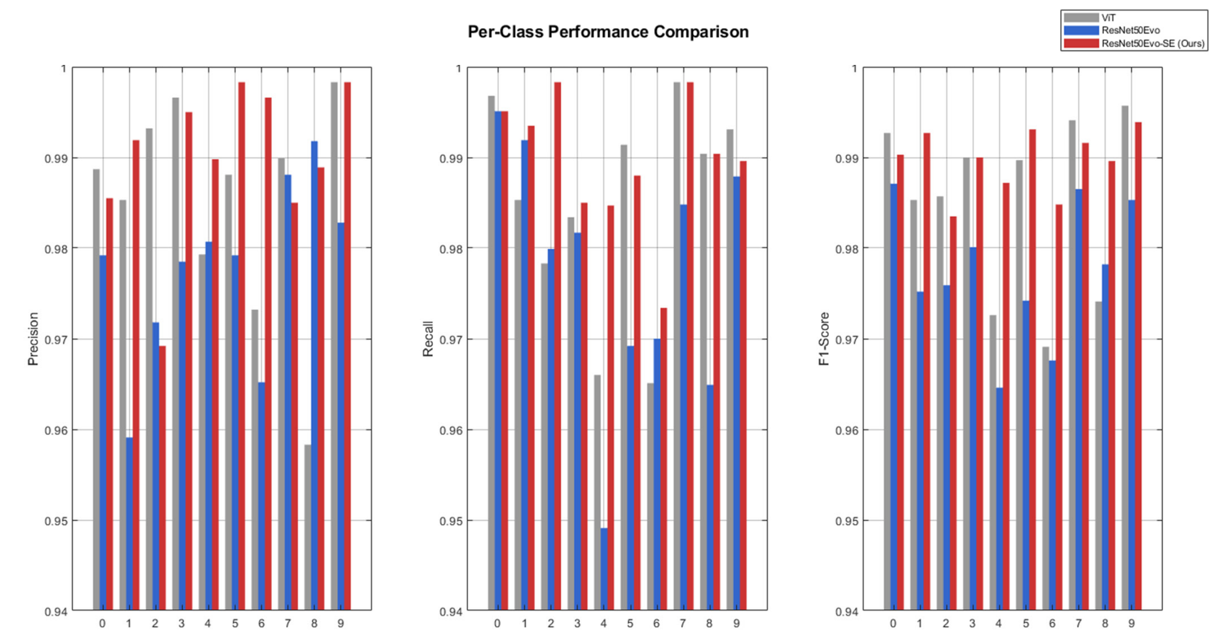
In addition to image-level classification, we explored an object detection approach to identify fine-grained visual symptoms. Unlike CNN and ViT models, which output a single class label for the whole image, object detection localizes specific symptomatic regions (e.g., leaf color, wilting, scorch) via bounding boxes. This provides a different level of output detail rather than a direct performance comparison. Accordingly, we adopted YoloV8 as the base detector and augmented it with CBAM to detect predefined visual proxies, which are then combined into diagnostic rules (Table 8) to infer water-nitrogen stress levels.

4.4.1. Dataset Construction

Using the plant-level split from Section 4.2.1, 14,968 images from training plants and 4,800 images from validation plants were selected based on clear stress

Figure 12

Per-Class Performance Metrics Comparison Among VIT, ResNet50Evo, and ResNet50Evo-SE Model.



(a) Precision per Class (b) Recall per Class (c) F1-Score per Class responses. The remaining images from test plants served as the test set. As identified in the previous section, the model encounters certain difficulties in classifying and identifying the water-nitrogen stress of lettuce when the traits, such as color and morphology, are very similar. Therefore, when establishing the object detection dataset, the fundamental principle was to distinguish different grades of water-nitrogen stress with the most evident features. This study redefined the labels for the object detection model and recombined different object detection features to serve as the final criteria for the classification and diagnosis of lettuce's water-nitrogen stress states. For details on labels and group relationships, refer to Tables 8, with labels defined specifically for lettuce leaves.

The features of lettuce leaves were manually annotated according to the bounding box labels defined in Table 8. The goal of this research is to perform classification and identification of lettuce's water-nitrogen stress by combining different detection box labels. The actual images of lettuce corresponding to the labels in Table 8 are shown in Figure 13. Images 15(a)-(d) are comparative images of lettuce leaf colors, 15(e)-(g) are images comparing lettuce leaf tips, and 15(h) and 15(i) compare images of burnt lettuce leaf edges. In addition to the label information above, instances of pest infestations in lettuce were discovered during the experiments; discussions regarding the annotation of pest damage images and pest prevention strategies will be detailed in Section 4.5.

4.4.2. Integrating CBAM into the YoloV8 Model

In the process of constructing deep learning network models, this study has repeatedly utilized attention mechanisms. Empirical evidence has shown that attention mechanisms are effective for both image segmentation and image classification models. Therefore, for object detection models, this study also attempts to add an attention mechanism to the YoloV8 model framework. YoloV8 is distinct from previous autoencoder models as it is a pre-packaged model that makes it direct and easy to use. This research tries to incorporate the Convolutional Block Attention Module (CBAM) [23], which is a CBAM module combining channel and spatial attentions. CBAM is an effective hybrid attention mechanism that assigns weight on the feature map level importance by learning on the basis of global average pooling. It contains both a channel attention module and a spatial attention module, which, when sequenced in the convolutional network, can significantly enhance the model's learning capabilities. The architecture of the CBAM is depicted in Figure 14. The YoloV8 model that incorporates this attention mechanism is defined as Yolo-CBAM.

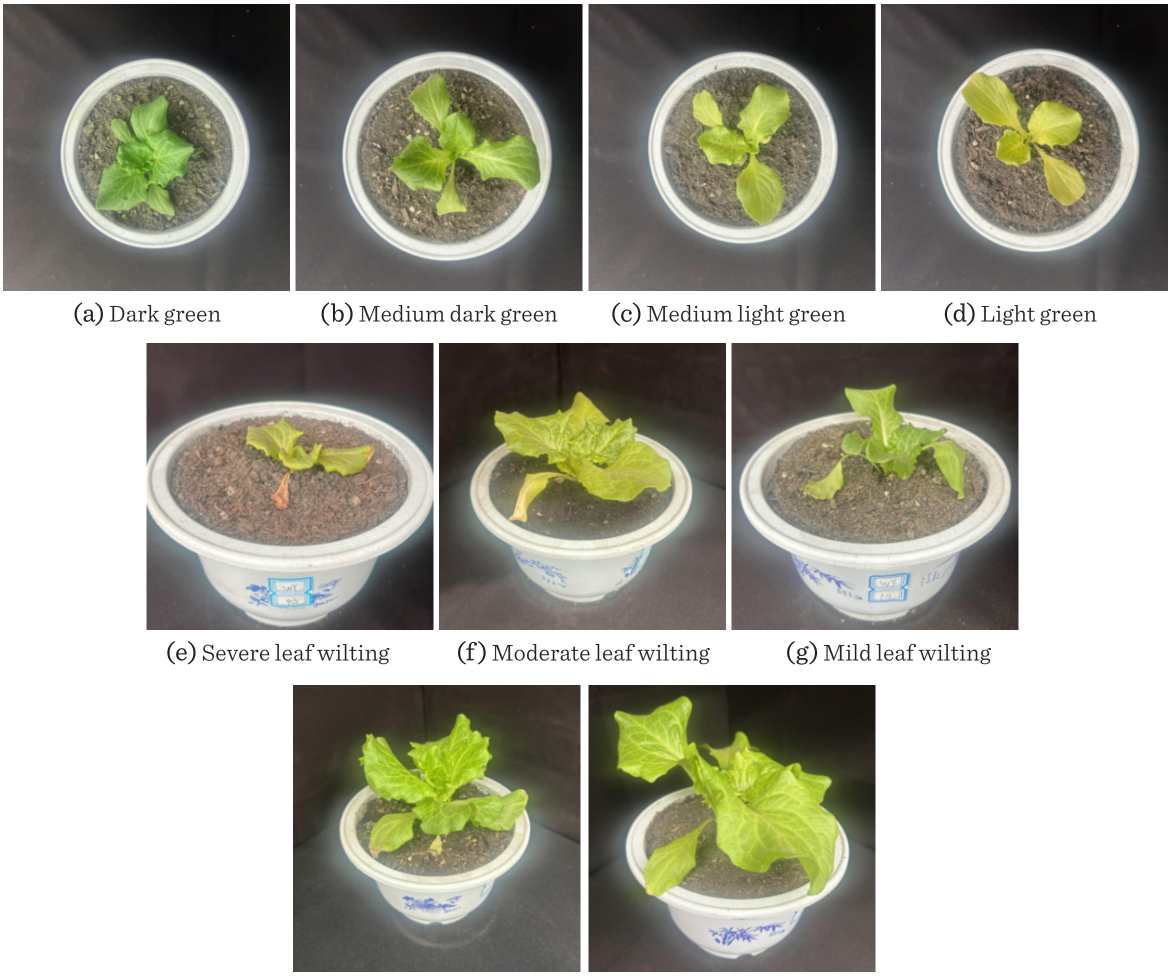
Table 8
Labels for the object detection model.

Groups	Label 1(Leaf Color)	Label 2(Physiological Symptoms)
NA	Dark green	Severe leaf scorch
NB	Dark green	Mild leaf scorch
NC	Medium dark green	Healthy / No visible symptoms
ND	Medium light green	Healthy / No visible symptoms
NE	Light green	Healthy / No visible symptoms
WA	Medium dark green	Healthy / No visible symptoms
WB	Medium dark green	Mild leaf wilting
WC	Medium light green	Mild leaf wilting
WD	Medium light green/Light green	Moderate leaf wilting
WE	Light green	Severe leaf wilting

Note: The groups (NA–NE, WA–WE) represent the treatment combinations defined in Tables 1-2. For example, NA stands for the treatment under Nitrogen Gradient A.

Figure 13

Actual lettuce images with different labels.



(a) Dark green (b) Medium dark green (c) Medium light green (d) Light green

(e) Severe leaf wilting (f) Moderate leaf wilting (g) Mild leaf wilting

(h) Severe leaf scorch (i) Mild leaf scorch

Figure 14

Architecture of CBAM module.

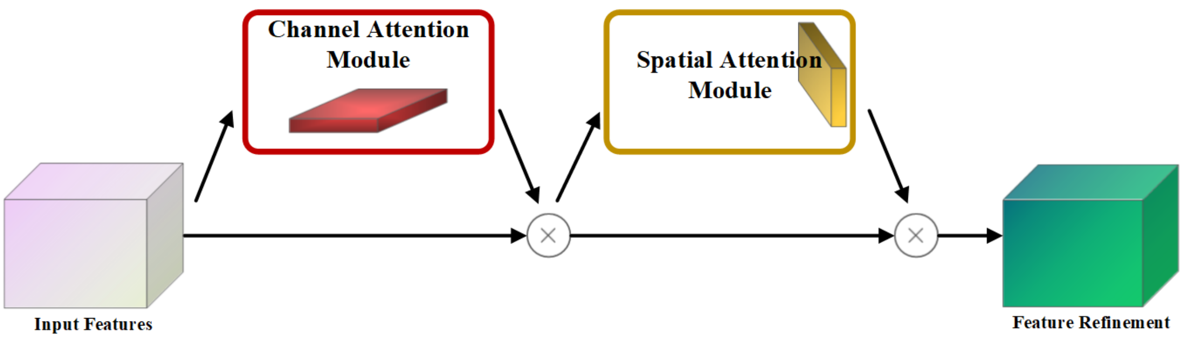
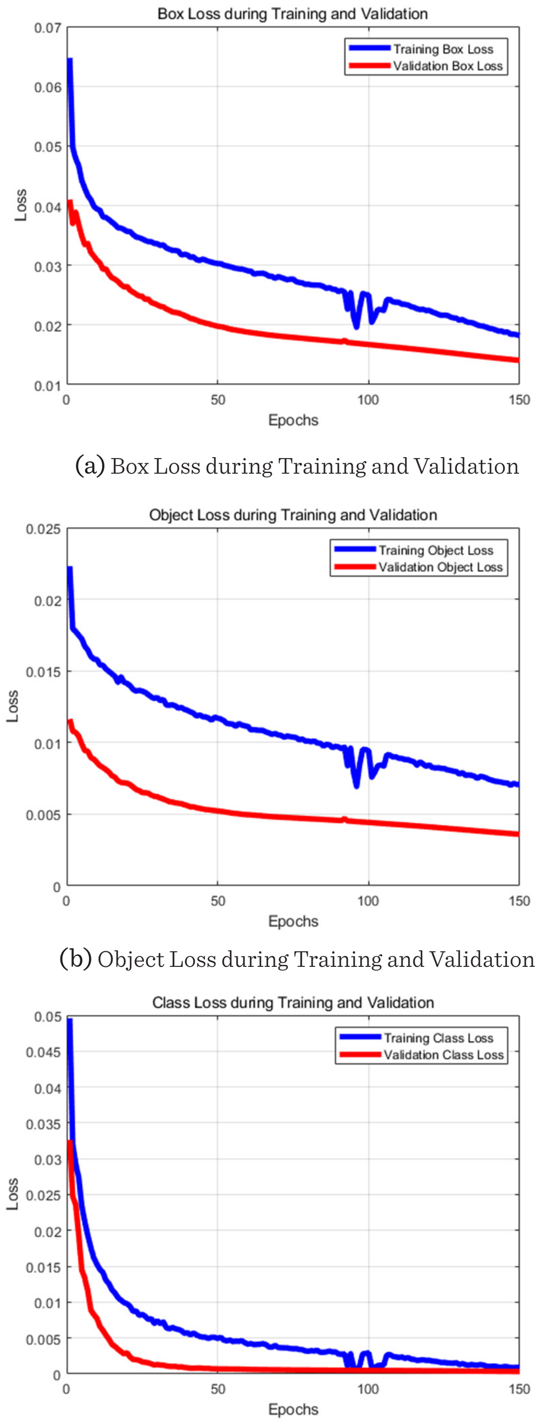


Figure 15

Comparison of three loss curves on the training and validation sets.



(a) Box Loss during Training and Validation

(b) Object Loss during Training and Validation

(c) Class Loss during Training and Validation

The following training parameters are summarized based on the characteristics of this experimental task after multiple trials:

1 The model's training task is set to detection and is fine-tuned using the pre-trained model yolov8s.pt.

2 During the training process, 150 training epochs are set. The initial learning rate is 0.01, which remains the same for the final learning rate. The momentum is set to 0.937, weight decay is set at 0.0005, warm-up epochs are designated as 3.0, warm-up momentum is set to 0.8, and warm-up bias learning rate is established at 0.1.

The loss curves for the training and validation datasets during the training process are shown in Figure 15. After 150 rounds of training, all three losses have dropped to a relatively ideal state on both the training and validation sets. Additionally, the experiments found that after 150 epochs, the improvement in the model's indicators was minimal. Therefore, considering computational resources and other factors, training was ceased after 150 epochs, and further evaluation of the model's performance was conducted.

4.4.3. Performance Evaluation of Yolo-CBAM

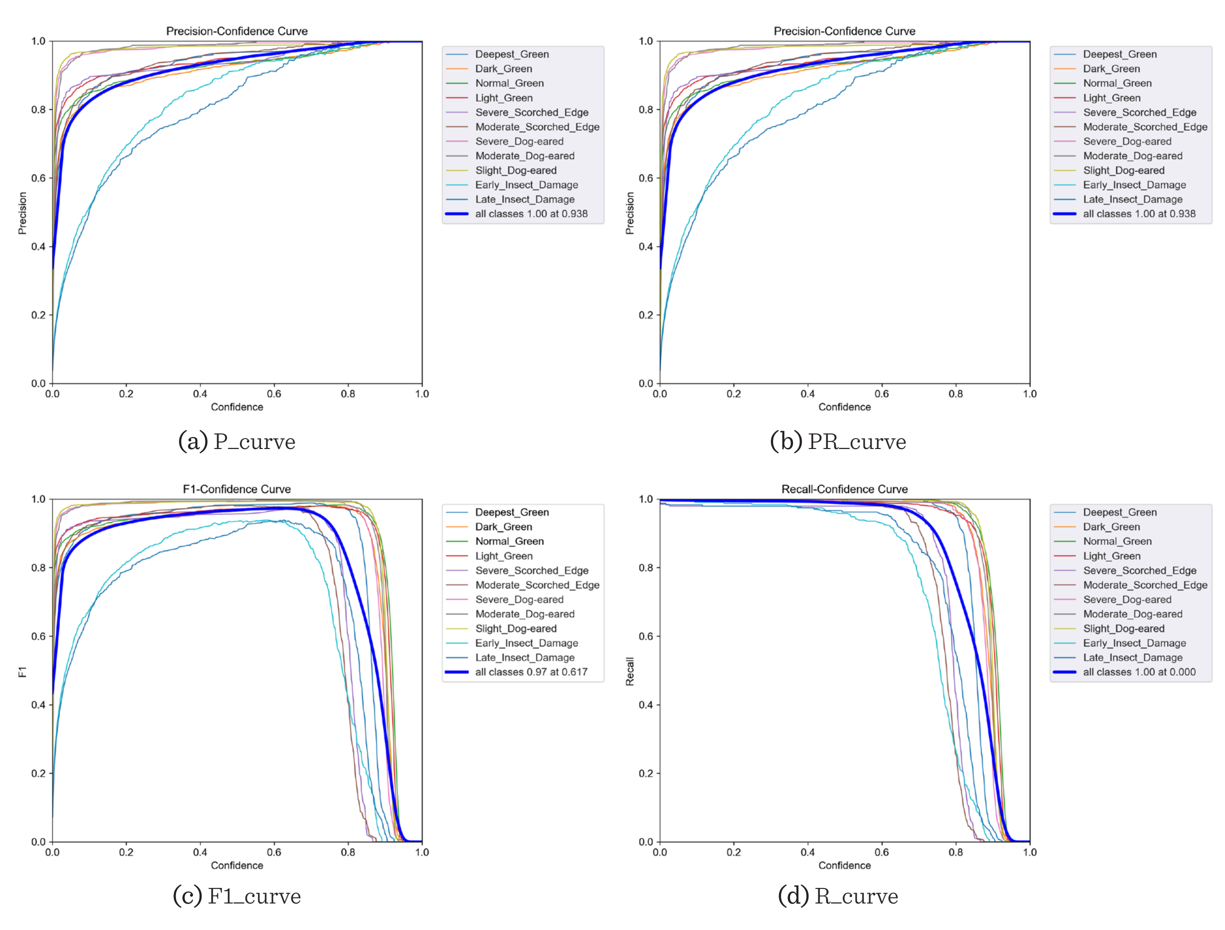
Following the training of the YOLOv8-CBAM model, performance was comprehensively evaluated using Precision, Recall, and mAP. Validation metrics were used for model tuning during training, and final results were reported on a strictly held-out test set. Table 9 compares the original YOLOv8 and the improved YOLOv8-CBAM on both validation and test sets. The four evaluation curves of YOLOv8-CBAM on the validation set are shown in Figure 16.

From the analysis of Table 9 and Figure 16, the performance of the object detection model Yolo-CBAM is as follows:

1 On an overall basis, the Yolo-CBAM model has met performance expectations, achieving excellent results across all indicators. Notably, the overall precision reached 0.9653, recall attained 0.9820, and the mAP also exhibited robust results, achieving scores of 0.9869 and 0.7711, respectively. The high mAP_0.5 indicates that the trained Yolo-CBAM model can effectively detect the predefined visual proxies (color and symptom labels) associated with water-nitrogen stress in lettuce. The gap between mAP_0.5 and

Figure 16

Four evaluation curves of the Yolo-CBAM model on validation set.



mAP_0.5:0.95 is largely attributed to the geometric characteristics of lettuce; specifically, the irregular and serrated edges of leaves pose an inherent challenge for standard axis-aligned bounding boxes to fit precisely under strict Intersection over Union (IoU) thresholds.

2 Analysis of individual labels indicates that the identification and detection of leaf color is highly successful. For the four different colors, both detection precision and recall rates exceed 0.965, providing an effective means of using leaf color to diagnose and grade water-nitrogen stress in let-

Table 9

Evaluation metrics of the Yolo-CBAM and YoloV8 model on validation and test set.

Model	Set	Precision	Recall	mAP_0.5	mAP_0.5:0.95
YoloV8	Validation	0.9401	0.9710	0.9782	0.7130
YoloV8	Test	0.9362	0.9671	0.9745	0.7085
Yolo-CBAM	Validation	0.9653	0.9820	0.9869	0.7711
Yolo-CBAM	Test	0.9614	0.9782	0.9832	0.7665

tuce. Particularly for the issue discussed in the previous section, where lettuce characteristics such as color and morphology are too similar to discern, the object detection method achieved robust discrimination for these subtle variations.

3 The detection of lettuce leaf tip burn and curvature also achieved anticipated results. These two features are relatively easier to identify and detect compared to leaf color, resulting in high accuracy and recall rates, both surpassing 0.975, with mAP_0.5:0.95 also exceeding 0.75. The use of leaf tip burn and curvature features for the identification and grading of lettuce in high nitrogen and severe water stress groups was very effective.

4 The precision rates for the identification and detection of late-stage and early-stage pest damage were slightly lower, at 0.9062 and 0.9003 respectively, below the overall precision, and the mAP did not perform as well as for the other labels. Investigating the cause, it is noted that there are fewer pest damage label samples than for other features. Of the 14,968 images in the training dataset, less than 1,000 were labeled for pest damage, and early-stage pest damage images comprised even fewer than 300. The insufficiency in the number of training images is attributed to the lower detection effectiveness for these two categories compared to others.

In summary, the model development pipeline presented in Sections 4.3-4.4 follow a progressive and theoretically grounded strategy. We began with a structurally modified CNN (ResNet50Evo) to enhance basic feature diversity. We then integrated channel-wise attention (SE) to enable adaptive feature recalibration within this improved backbone. Finally, recognizing the spatial localization challenge, we adopted an object detection framework (YOLOv8) to directly focus on symptomatic regions. This multi-stage optimization reflects a deliberate effort to tailor deep learning architectures to the specific characteristics of agricultural stress imagery, rather than applying generic off-the-shelf models.

4.5. Early Pest Prevention Based on Object Detection

This study was originally designed to diagnose and grade water-nitrogen stress in lettuce. During the experiments, early-stage pest damage (leaf miners) was also observed and recorded. Recognizing the scarcity of datasets for this specific pest, we expanded the framework to include object detection capabilities. Although the two tasks share the same input data, they are processed as parallel branches to avoid feature interference, with the current study focusing on validating these capabilities under isolated stress conditions. Specifically, during the construction of the object detection dataset, pest damage was identified on some lettuce leaves through manual labeling, which was attributed to leafminer flies upon investigation. According to literature [28], temperatures ranging from 20°C to 30°C are suitable for the growth and development of leafminer flies, and their damage is more severe in greenhouses than in open fields. Adult flies are most active between 9-11 AM and 2-4 PM. In this experiment, the greenhouse environment for growing lettuce was consistent with the conditions described in the literature. Female leafminer flies lay eggs within the host leaf tissue, and upon hatching, the larvae feed on the plant tissue under the leaf epidermis. Adults often feed on the sap at the puncture sites, causing the plants to grow slowly, leaves to fall off, and consequently, a reduction in yield. Typically, the eggs hatch within 2-5 days, and the larval stage lasts 4-7 days. Therefore, it takes time for leafminer flies to damage lettuce. If detection and intervention can occur before the eggs hatch or before the larvae start damaging the leaves, the impact of leafminer flies on lettuce can be effectively prevented. Consequently, this study supplemented the dataset used to train the Yolo-CBAM model by adding two labels for early-stage and late-stage pest damage, defined as follows:

Figure 17

Image of leafminer damage in lettuce.

(a) Early-stage (b) Late-stage

1 **Early-stage pest damage:** The eggs laid by female flies within the host leaf tissue are either unhatched or have just hatched. As shown in Figure 17(a), the dotted points represent the early features of pest damage.

2 **Late-stage pest damage:** The lettuce leaf epidermis is chewed upon, showing streaked patterns, which may ultimately lead to wilting. As shown in Figure 17(b), the streaked patterns are characteristic of late-stage pest damage, and as the stripes increase, the leaf gradually wilts.

As indicated by Table 9 and Figure 16, the trained Yolo-CBAM model can effectively detect features of 11 different labels, including the early and late stages of pest damage. However, due to the limited number of dataset samples for early and late pest damage features, the evaluation metrics are lower than those of other label features. Consequently, a thorough investigation of the original dataset images was conducted, and approximately 1000 pest-damaged images were manually re-annotated and combined with previous images. This forms an enlarged dataset comprising about 2000 images of both early and late-stage pest damage. Concurrently, the well-trained Yolo-CBAM model served as the foundation for continued training. Following this extended training, the model's performance is presented in Table 10.

The images of pest-infested lettuce were subjected to classified analysis. Observations from multiple image sets revealed that it generally takes several days for leafminers to progress from early to late stages of pest damage. As shown in Figure 18, the series of images capture the same leaf of lettuce over 12 consecutive days, documenting the progression from early to late stages of pest damage. The images reveal that the leaves were normal during the first two days. Around day 3-4, partial black and white spots appeared, and the leaf surface became rough. Streaking began on the fifth day, gradually expanding over the days, eventually leading to leaf wilting.

Table 10

Evaluation for pest object detection after continued training.

Labels	Precision	Recall	mAP_0.5	mAP_0.5:0.95
Early-stage	0.9443	0.9531	0.9722	0.6908
Late-stage	0.9489	0.9568	0.9793	0.7035

The Yolo-CBAM model accurately identifies the early stages of pest damage. As depicted in Figure 19, the images are the automatically saved feature images resulting from using the Yolo-CBAM model to perform object detection on the identified damaged areas. These feature images correspond to the specific regions of the lettuce affected by pests. If manual intervention [6, 11] is promptly applied during the early stages, effective control of leafminer fly infestation can be achieved.

Figure 18

Development process of leafminer.

(a) 1-6 days

(b) 7-12 days

Figure 19
Feature maps extracted by YOLO-CBAM model.

4.6. Discussion on the Universality of Grading Standards

The grading criteria established in this study were calibrated specifically under controlled greenhouse conditions with a black background, fixed lighting, and a standardized multi-angle imaging protocol. While the correlation between visual symptoms (e.g., leaf wilting, color changes) and physiological stress is consistent across most leafy vegetables, the absolute thresholds for classification may vary among different cultivars or growth stages. For instance, varieties with inherently thicker leaves may exhibit a higher tolerance to water stress before visible wilting occurs. Therefore, the proposed deep learning framework should be considered greenhouse-specific; its direct applicability to open-field environments with variable lighting, cluttered backgrounds, or different imaging setups requires further external validation.

5. Conclusions and Prospects

5.1. Conclusions

This paper addresses critical issues in the lettuce growth process by developing a series of precision computational models based on deep learning. These models enable accurate identification and grading of water-nitrogen stress in lettuce, as well as early prevention of pest damage.

Deep learning models for the identification and grading of water-nitrogen stress in lettuce have been established using phenotypic indicators and image data. Additionally, models using image data have been developed for the object detection of water-nitrogen stress and pest damage traits. These models facilitate the accurate identification and grading diagnosis of water-nitrogen stress conditions in lettuce, as well as early pest prevention. The results indicate that the DNN network model, established with the feature phenotype indicator dataset FPD, can efficiently and rapidly diagnose water-nitrogen stress conditions in lettuce. This provides a feasibility basis for monitoring and diagnosing water-nitrogen stress using non-image data. The ResNet50Evo-SE model, created using image data combined with the SE channel attention mechanism, achieves a high accuracy rate in identifying and grading the water-nitrogen stress conditions in lettuce. Moreover, compared to CNNs that focus on local texture features, the ViT model demonstrates superior capability in capturing global dependencies of leaf stress, validating the complementary advantages of different architectural backbones in this task. The YoloV8-CBAM object detection model accurately detects predefined visual proxies (leaf color and physiological symptoms) of water-nitrogen stress. These proxies can be combined into diagnostic rules for stress assessment, providing fine-grained symptom localization that complements the image-level classification achieved by ViT and ResNet50Evo-SE. Because the output spaces differ (stress grades vs. symptom labels), the detection metrics are not directly comparable to the classification accuracies; each model serves a distinct role in a multi-modal diagnostic system. Furthermore, the YoloV8-CBAM model effectively identifies and detects leafminer pest damage in lettuce leaves. It should be noted, however, that all results are limited to the specific greenhouse conditions (controlled lighting, black background, fixed imaging protocol) employed in this study. External validation under open-field or variable conditions is needed before generalizing these findings.

5.2. Prospects

The deep learning models developed in this study achieve the recognition and diagnosis of water-nitrogen stress and pest damage in lettuce, providing effective solutions for intelligent cultivation and whole-life-cycle intelligent monitoring of lettuce. However, there are still several research directions to be explored:

1. The current models were trained and validated exclusively under controlled greenhouse conditions with a black background and fixed imaging protocol. They have not yet been tested in open-field environments with variable lighting, cluttered backgrounds, or occlusions. External validation using diverse field datasets is therefore necessary before these models can be considered for general agricultural deployment. Future work will focus on collecting such datasets and applying domain adaptation techniques to improve robustness.
2. Deep learning technology has shown promising outcomes; however, its high computational demands and the opaqueness of its models pose challenges. Advancements in creating more efficient, accessible, and interpretable models are crucial to enhance its adoption in agricultural applications.

Acknowledgements

This research was funded by:

- Doctoral Research Startup Fund of Mudanjiang Medical University (Grant No. 2024-MYBSKY-004)
- Basic Scientific Research Project of Heilongjiang Provincial Education Department (Grant No. 2024-KYYWF-0494)
- Outstanding Master/Doctoral Dissertation Award of Heilongjiang Province (Grant No. LJYXL2024-038).
- Key Research and Development Program of Mudanjiang Municipal Science and Technology Bureau (Grant No. HA25BR075)

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The author declares that thereis no conflict of interest.

Declaration of generative AI in scientific writing

The author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

1. Ahmed, N., Xinagyu, G., Alnafissa, M., Khan, A., Rehman, S., Ali, F. Evaluating the Impact of Sustainable Technology, Resource Utilization, and Climate Change on Soil Emissions: A CS-ARDL Analysis of Leading Agricultural Economies. Cleaner Engineering and Technology, 2025, 24, 100869. https://doi.org/10.1016/j.clet.2024.100869
2. Chamara, N., Bai, G., Ge, Y. AICropCAM: Deploying Classification, Segmentation, Detection, and Counting Deep-Learning Models for Crop Monitoring on the Edge. Computers and Electronics in Agriculture, 2023, 215, 108420. https://doi.org/10.1016/j.compag.2023.108420

3. Chandel, N. S., Jat, D., Chakraborty, S. K., Singh, R., Sharma, A., Kumar, P. Deep Learning Assisted Real-Time Nitrogen Stress Detection for Variable Rate Fertilizer Applicator in Wheat Crop. Computers and Electronics in Agriculture, 2025, 237, 110545. https://doi.org/10.1016/j.compag.2025.110545
4. Chen, F., Qiu, H., Zhao, Y., Wang, X., Zhang, Y., Li, J. Impact of New Maize Variety Adoption on Yield and Fertilizer Input in China: Implications for Sustainable Food and Agriculture. Agricultural Systems, 2024, 218, 104004. https://doi.org/10.1016/j.agsy.2024.104004
5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N. An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale. In, Proceedings of the International Conference on Learning Representations (ICLR), 2021. Available at, https://openreview.net/pdf?id=YicbFdNTTy
6. Du, S.-J., Ye, F.-Y., Xu, S.-Y., Wan, W.-J., Guo, J.-Y., Yang, N.-W., Liu, W.-X. Thelytokous Diglyphus wani: A More Promising Biological Control Agent Against Agromyzid Leafminers Than Its Arrhenotokous Counterpart. Journal of Integrative Agriculture, 2023, 22(12), 3731-3743. https://doi.org/10.1016/j.jia.2023.05.012
7. Fuentes, A. F., Sook, Y., Lee, J., Park, S., Kim, D. High-Performance Deep Neural Network-Based Tomato Plant Diseases and Pests Diagnosis System with Refinement Filter Bank. Frontiers in Plant Science, 2018, 9, 1162. https://doi.org/10.3389/fpls.2018.01162
8. Haggag, M., Abdelhay, S., Mecheter, A., Elgamal, A., Hassan, M. An Intelligent Hybrid Experimental-Based Deep Learning Algorithm for Tomato-Sorting Controllers. IEEE Access, 2019, 7, 106890-106898. https://doi.org/10.1109/ACCESS.2019.2932730
9. Hinton, G. E., Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks. Science, 2006, 313(5786), 504-507. https://doi.org/10.1126/science.1127647
10. Hu, J., Shen, L., Sun, G. Squeeze-and-Excitation Networks. In, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, 7132-7141. https://doi.org/10.1109/CVPR.2018.00745
11. Jabamo, T., Ayalew, G., Goftishu, M., Wakgari, M. Integrated Effect of Insecticide and Sex Pheromone on the Tomato Leafminer, Tuta absoluta (Lepidoptera: Gelechiidae). Crop Protection, 2023, 171, 106285. https://doi.org/10.1016/j.cropro.2023.106285
12. Kumar, D., Kukreja, V. Image Segmentation, Classification, and Recognition Methods for Wheat Diseases: Two Decades' Systematic Literature Review. Computers and Electronics in Agriculture, 2024, 221, 109005. https://doi.org/10.1016/j.compag.2024.109005
13. Kumar, S., Singh, S. P., Rahman, M., Islam, S., Khan, M. A., Islam, M. T. Unravelling the Multifarious Role of Wood Vinegar Made from Waste Biomass in Plant Growth Promotion, Biotic Stress Tolerance, and Sustainable Agriculture. Journal of Analytical and Applied Pyrolysis, 2025, 185, 106851. https://doi.org/10.1016/j.jaap.2024.106851
14. Li, D., Li, C., Yao, Y., Wang, J., Zhang, X., Liu, H. Modern Imaging Techniques in Plant Nutrition Analysis: A Review. Computers and Electronics in Agriculture, 2020, 174, 105459. https://doi.org/10.1016/j.compag.2020.105459
15. Lin, T. Y., Goyal, P., Girshick, R., He, K., Dollár, P. Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2), 318-327. https://doi.org/10.1109/TPAMI.2018.2858826
16. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A. C. SSD: Single Shot MultiBox Detector. In, Proceedings of the European Conference on Computer Vision (ECCV), 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
17. Maji, A. K., Das, S., Marwaha, S., Kumar, S., Dutta, S., Roy Choudhury, M., Arora, A., Ray, M., Perumal, A., Chinusamy, V. Intelligent Decision Support for Drought Stress (IDSDS): An Integrated Remote Sensing and Artificial Intelligence-Based Pipeline for Quantifying Drought Stress in Plants. Computers and Electronics in Agriculture, 2025, 236, 110477. https://doi.org/10.1016/j.compag.2025.110477
18. Owomugisha, G., Mwebaze, E. Machine Learning for Plant Disease Incidence and Severity Measurements from Leaf Images. In, Proceedings of the IEEE International Conference on Machine Learning and Applications (ICMLA), 2016, 158-163. https://doi.org/10.1109/ICMLA.2016.0031
19. Preite, L., Vignali, G. Artificial Intelligence to Optimize Water Consumption in Agriculture: A Predictive Algorithm-Based Irrigation Management System. Computers and Electronics in Agriculture, 2024, 223, 109126. https://doi.org/10.1016/j.compag.2024.109126
20. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In, Proceedings of the IEEE Conference on Computer

Vision and Pattern Recognition (CVPR), 2016, 779-788. https://doi.org/10.1109/CVPR.2016.91

21. Ren, S., He, K., Girshick, R., Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6), 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

22. Sankaran, S., Mishra, A., Ehsani, R. A Review of Advanced Techniques for Detecting Plant Diseases. Computers and Electronics in Agriculture, 2010, 72, 1-13. https://doi.org/10.1016/j.compag.2010.02.007

23. Woo, S., Park, J., Lee, J. Y., Kweon, I. S. CBAM: Convolutional Block Attention Module. In, Proceedings of the European Conference on Computer Vision (ECCV), 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

24. Xie, C., Shao, Y., Li, X. Detection of Early Blight and Late Blight Diseases on Tomato Leaves Using Hyperspectral Imaging. Scientific Reports, 2015, 5, 16564. https://doi.org/10.1038/srep16564

25. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K. Aggregated Residual Transformations for Deep Neural Networks. In, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, 5987-5995. https://doi.org/10.1109/CVPR.2017.634

26. Xue, B., Zheng, Q., Li, Z., Wang, J., Mu, C., Yang, J., Fan, H., Feng, X., Li, X. Perturbation Defense Ultra High-Speed Weak Target Recognition. Engineering Applications of Artificial Intelligence, 2024, 138, 109420. https://doi.org/10.1016/j.engappai.2024.109420

27. Yu, H., Dong, M., Zhao, R., Zhang, L., Sui, Y. Research on Precise Phenotype Identification and Growth Prediction of Lettuce Based on Deep Learning. Environmental Research, 2024, 252, 118845. https://doi.org/10.1016/j.envres.2024.118845

28. Zhang, G.-F., Ma, D.-Y., Wang, Y.-S., Gao, Y.-H., Liu, W.-X., Zhang, R., Fu, W.-J., Xian, X.-Q., Wang, J., Kuang, M., Wan, F.-H. First Report of the South American Tomato Leafminer, Tuta absoluta (Meyrick), in China. Journal of Integrative Agriculture, 2020, 19(7), 1912-1917. https://doi.org/10.1016/S2095-3119(20)63165-3

29. Zhao, Q., Feng, J., Guo, S., Liu, Y., Zhang, H., Chen, X. Water-Fertilizer Regulation Improves Yield, Quality, and Water-Fertilizer Use Efficiency of Greenhouse Lettuce by Altering the Bacterial Community in Nutrient Solutions. Journal of Agriculture and Food Research, 2025, 21, 101976. https://doi.org/10.1016/j.jafr.2025.101976

30. Zheng, X., Zheng, J., Wang, X., Li, H., Sun, Y., Zhao, L. An Adaptive Recognition Method for Crop Row Orientation in Dry Land by Combining Morphological and Texture Features. Soil and Tillage Research, 2025, 252, 106576. https://doi.org/10.1016/j.still.2025.106576

31. Zhou, Z., Majeed, Y., Diverres Naranjo, G., Gambacorta, E. M. T. Assessment for Crop Water Stress with Infrared Thermal Imagery in Precision Agriculture: A Review and Future Prospects for Deep Learning. Computers and Electronics in Agriculture, 2021, 182, 106019. https://doi.org/10.1016/j.compag.2021.106019