scientific reports

Check for updates

OPEN

Research on agricultural disease recognition methods based on very large Kernel convolutional network-RepLKNet

Guoquan Pei[1], Xueying Qian[1], Bing Zhou[2], Zigao Liu[3] & Wendou Wu[1]✉

Agricultural diseases pose significant challenges to plant production. With the rapid advancement of deep learning, the accuracy and efficiency of plant disease identification have substantially improved. However, conventional convolutional neural networks that rely on multi-layer small-kernel structures are limited in capturing long-range dependencies and global contextual information due to their constrained receptive fields. To overcome these limitations, this study proposes a plant disease recognition method based on RepLKNet, a convolutional architecture with large kernel designs that significantly expand the receptive field and enhance feature representation. Transfer learning is incorporated to further improve training efficiency and model performance. Experiments conducted on the Plant Diseases Training Dataset, comprising 95,865 images across 61 disease categories, demonstrate the effectiveness of the proposed method. Under five-fold cross-validation, the model achieved an overall accuracy (OA) of 96.03%, an average accuracy (AA) of 94.78%, and a Kappa coefficient of 95.86%. Compared with ResNet50 (OA: 95.62%) and GoogleNet (OA: 94.98%), the proposed model demonstrates competitive or superior performance. Ablation experiments reveal that replacing large kernels with 3×3 or 5×5 convolutions results in accuracy reductions of up to 1.1% in OA and 1.3% in AA, confirming the effectiveness of the large kernel design. These results demonstrate the robustness and superior capability of RepLKNet in plant disease recognition tasks.

Keywords Agriculture, Disease recognition, Large kernel, Convolution, RepLKNet

Agriculture remains a cornerstone of global industry and food security[1], yet plant diseases continue to pose severe threats to plant yield and quality. Traditional machine learning approaches for agricultural disease recognition often depend on manual feature engineering, which limits scalability and adaptability across diverse disease types and environmental conditions. In contrast, deep learning has emerged as a powerful paradigm, enabling automatic feature extraction and offering superior generalization and robustness in complex agricultural scenarios[2].

Recent advances have focused on refining classical deep learning architectures for improved plant disease and species classification. For example, Gulzar et al.[3] proposed PL-DenseNet, an optimized DenseNet[4] variant for pear leaf disease identification. By integrating Global Average Pooling, Batch Normalization, Dropout layers, transfer learning, and data augmentation, the model achieved 99.18% accuracy on the DiaMOS Plant dataset-outperforming EfficientNetB0[5] and ResNet50[6]. Similarly, Pacal[7] developed a vision transformer model based on MaxViT for maize leaf disease recognition, reaching 99.24% accuracy and outperforming over 60 deep learning models on a newly constructed large-scale dataset. Kunduracioglu et al.[8] further explored the use of CNNs and transformer architectures for grape leaf disease diagnosis and classification, reporting perfect classification results on the PlantVillage and Grapevine datasets. Umar et al.[9] proposed a modified YOLOv7 model enhanced with SimAM and DAiAM modules for tomato leaf disease detection. Their approach, which also incorporated CNN-based classification and feature extraction via SIFT, achieved an accuracy of 98.8%, highlighting the effectiveness of combining object detection and classification in field environments. Kaya et al.[10] introduced a multi-head CNN based on DenseNet that fuses RGB and segmented images for improved recognition. Their model was validated on the PlantVillage dataset, achieving an average accuracy of 98.17% and showing strong

[1]College of Big Data, Yunnan Agricultural University, Kunming 650201, China. [2]College of Science, Yunnan Agricultural University, Kunming 650201, China. [3]Yunnan Traceability Technology Co. Ltd., Kunming 650201, China. ✉email: wuwd2004@126.com

robustness across disease types. Furthermore, Thakur et al.[11] proposed a lightweight CNN architecture named VGG-ICNN, optimized for crop disease recognition across five different datasets.

Lightweight convolutional networks have also been adapted for agricultural tasks. Alkanan et al.[12] improved MobileNetV2 by adding Average Pooling, Flatten, and Dropout layers to classify corn seed diseases. With adaptive learning rate scheduling and model checkpointing, the model achieved 96% accuracy on a dataset of 21,662 images, outperforming traditional models in terms of precision and recall. Hybrid and ensemble architectures have gained attention for their enhanced robustness. Amri et al.[13] proposed MIV-PlantNet, a fusion of MobileNet, Inception, and VGG, for classifying native Saudi plant species. The model achieved 99% accuracy and a 98% F1-score, while incorporating SHAP-based explainable AI to improve interpretability. Gulzar[14] modified InceptionV3 by adding Dense and Dropout layers, achieving 98.73% accuracy across five categories. The combination of transfer learning and adaptive scheduling further boosted F1-scores above 0.98. Transfer learning has consistently proven effective in plant classification.Li et al.[15] proposed the GhostNet_Triplet_YOLOv8 s algorithm, which integrates the GhostNet architecture and Triplet Attention to improve maize disease detection. This model achieved a 0.3% increase in mAP and a 50.2% reduction in size, making it ideal for real-time agricultural applications. Wang et al.[16]. The ULEN architecture, designed with only 100,000 parameters, combines residual depth-wise convolutions with a spatial pyramid pooling layer to effectively classify plant diseases and pest infections. Gulzar et al.[17] compared MobileNetV3, InceptionV3, Xception, VGG19, DenseNet121, ResNet101, and EfficientNetB3 on a custom dataset of 1,214 alfalfa leaf images from three varieties (Bilensoy-80, Diana, and Nimet). DenseNet121 and EfficientNetB3 reached test accuracies up to 100%, underscoring the utility of transfer learning for limited datasets.

Similarly, Gulzar et al.[18] used a pre-trained Xception model for seed classification on a dataset of 3,018 images from 15 seed types, achieving perfect accuracy (1.0000) on both validation and test sets. The model also showed faster convergence and lower training loss, demonstrating its predictive efficiency. From a broader perspective, Seelwal et al.[19] reviewed 69 studies on rice disease detection from 2008 to 2023, highlighting the increasing use of deep learning-based hybrid models, improved datasets, and enhanced annotation practices for diagnosing diseases such as rice blast and brown spot. An even more comprehensive overview is provided by Pacal et al.[20], who systematically reviewed 160 studies published between 2020 and 2024. This study categorized research into three core tasks-disease classification, detection, and segmentation-and provided insights into frequently used deep learning architectures, datasets, and common challenges. The review emphasized the growing adoption of vision transformers and hybrid architectures in plant disease recognition.

Building on this trend, Pacal et al.[21] explored CNNs and advanced Vision Transformers (ViT), including MaxViT, DeiT3, and MViTv2, for corn leaf disease classification. Their hybrid approach, which incorporated image preprocessing, data augmentation, and adaptive ensemble learning, achieved perfect classification accuracy (100%) on the CD&S dataset and 99.83% on the PlantVillage dataset. These results highlight the superior capability of ViT-based models in capturing global feature representations compared to traditional CNNs. Pal et al.[22] introduced the AgriDet framework, which combines INC-VGGN and Kohonen-based deep learning networks for plant disease detection and severity classification. The framework improves accuracy by addressing occlusion and background issues through pre-processing and multi-variate grabcut segmentation, achieving superior performance over previous methods.

Together, these studies highlight key trends in architectural optimization, lightweight network design, and transfer learning, alongside increasing attention to model interpretability. However, a common limitation in most convolutional neural network (CNN)-based approaches is their reliance on small convolutional kernels, typically 3×3, 5×5, or 7×7. For instance, Kumar et al.[23] applied ResNet-50 with 3×3 kernels for tomato disease classification, while Saxena et al.[24] employed AlexNet and GoogLeNet, both utilizing small convolutional kernels, to detect general plant diseases[25].

While small kernels offer computational efficiency, they restrict the effective receptive field, necessitating deep stacking to capture long-range dependencies-an approach that may lead to excessive model depth, vanishing gradients, and reduced training efficiency. In contrast, very large convolutional kernels (e.g., 31× 31) significantly expand the receptive field, enabling the network to capture global contextual and shape-level features that are crucial for recognizing irregular disease symptoms. Traditionally, such kernels were avoided due to computational demands, but innovations such as depthwise separable convolutions and residual connections have made them practical in modern architectures[6,26,27].

Despite the remarkable progress in plant disease recognition, a key limitation persists in most deep learning models: the predominant reliance on small convolutional kernels, typically 3×3, 5×5, or 7×7. These small kernels, while computationally efficient, limit the effective receptive field, making it difficult for models to capture global contextual features. As a result, networks often require deep stacking of layers to model long-range dependencies, which increases training complexity, introduces vanishing gradient issues, and reduces overall efficiency. Moreover, the inability to capture large-scale, irregular patterns-such as scattered lesions, diffuse blights, or non-localized discolorations-can hinder accurate disease recognition in real-world agricultural imagery.

Large convolutional kernels offer a promising alternative by inherently expanding the receptive field, thereby enabling the model to directly extract global spatial and structural information from input images. This is particularly valuable in agricultural disease detection, where symptoms often exhibit complex spatial distributions. Although historically avoided due to high computational costs, recent architectural innovations-such as depthwise separable convolutions and residual connections-have made large kernels more practical and efficient.

Building on this idea, RepLKNet introduces a novel architectural design that utilizes very large convolutional kernels (e.g., 31×31) within a re-parameterizable framework. This approach allows the network to capture long-

range spatial dependencies without requiring excessive depth, making it especially well-suited for agricultural disease recognition tasks where global shape features are critical.

In this study, we propose a novel plant disease recognition framework based on RepLKNet and evaluate its effectiveness on a comprehensive plant disease image dataset.

The key contributions of this paper are summarized as follows:

- We propose the application of RepLKNet with large 31×31 convolutional kernels to capture global spatial and shape-related features essential for identifying complex plant disease symptoms.
- We construct and utilize a large-scale plant disease image dataset containing 95,865 annotated images across 14 plant species and 61 disease categories, covering both healthy and diseased samples.
- Our proposed method achieves competitive results, with an overall accuracy (OA) of 96.03%, average accuracy (AA) of 91.9%, and a Kappa coefficient of 93.3%, demonstrating the feasibility and effectiveness of very large kernel convolutions for real-world agricultural disease recognition.

Materials and methods

Data sources

The dataset used for this model is called the Plant Diseases Training Dataset, available on Kaggle at https://www.kaggle.com/code/gpiosenka/efficentnet-mobilenet-f1-s-93-93/input. The dataset covers various plants and their associated diseases and health conditions. Fig.1 presents a few example images from the dataset, showcasing selected plant types and their corresponding diseases.

The dataset has been pre-organized by category into separate folders for each plant and disease type. Each folder contains images corresponding to a specific disease or healthy state for the plants. The categorization has been done manually, and the images have been labeled using a Python script that automatically assigns the corresponding class label to each image based on the folder it resides in.

The dataset used in this study is a publicly available and well-curated agricultural dataset, comprising 95,868 images across 17 different plant species and 61 distinct disease categories, including healthy samples. One of the key advantages of this dataset lies in its broad plant diversity, which ensures that the model is exposed to a wide range of morphological and pathological variations. This diversity significantly enhances the model's ability to generalize to real-world agricultural scenarios involving different plant types.

Moreover, the large volume of annotated images provides a rich source of visual information for training deep learning models. The high number of samples not only contributes to improved feature representation and robustness but also reduces the risk of overfitting, particularly in complex classification tasks involving subtle disease symptoms. The inclusion of both healthy and diseased samples further facilitates the model's ability to distinguish between normal and abnormal plant conditions, improving classification accuracy and practical applicability in precision agriculture. The detailed categories and sample counts of the dataset are presented in Table.1.

Data splitting and preprocessing

To unify the data size and improve model training efficiency, we cropped all images to a size of 400px $\times$ 400px, as shown in Fig.2. This process was necessary to ensure consistent input dimensions for the model. Additionally, channel normalization was performed on the cropped data to avoid over-reliance on certain features due to inconsistent feature scales, thus enhancing the model's generalization ability.

The dataset was randomly divided into ten equal parts, and then split into training, validation, and test sets with an 8:1:1 ratio. This division was made in preparation for the subsequent 5-fold cross-validation experiment, ensuring a robust evaluation of the model's performance. The details of the data split are shown in Table.2.

RepLKNet

RepLKNet introduces large convolutional kernels-up to 31×31 in size-to enhance the model's capacity for global feature extraction, which differs fundamentally from traditional CNNs that rely on small kernels stacked in deep layers. While small kernels excel at capturing local features, they often require multiple layers to achieve a sufficient receptive field, potentially leading to information loss during spatial downsampling. In contrast, large kernels enable RepLKNet to capture long-range dependencies and contextual information directly within

(a) Apple **(b)** Cherry **(c)** Corn **(d)** Tomato

Fig. 1. Partial data preview.

Category	Image Count	Category	Image Count
Apple alternaria leaf spot	278	Bell pepper bacterial spot	997
Apple black rot	621	Bell pepper healthy	1478
Apple brown spot	215	Blueberry healthy	1502
Apple gray spot	395	Cassava bacterial blight	1087
Apple healthy	2570	Cassava brown streak disease	2189
Apple rust	1241	Cassava green mottle	2386
Apple scab	1222	Cassava healthy	2577
Bell pepper bacterial spot	997	Cassava mosaic disease	13158
Bell pepper healthy	1478	Cherry healthy	854
Blueberry healthy	1502	Cherry powdery mildew	1052
Cassava bacterial blight	1087	Corn common rust	1192
Cassava brown streak disease	2189	Corn gray leaf spot	513
Cassava green mottle	2386	Corn healthy	1162
Cassava healthy	2577	Corn northern leaf blight	985
Cassava mosaic disease	13158	Squash powdery mildew	1835
Cherry healthy	854	Soybean healthy	5090
Cherry powdery mildew	1052	Strawberry healthy	456
Corn common rust	1192	Strawberry leaf scorch	1109
Corn gray leaf spot	513	Tomato bacterial spot	2127
Corn healthy	1162	Tomato early blight	1000
Corn northern leaf blight	985	Tomato healthy	1591
Squash powdery mildew	1835	Tomato late blight	1909
Soybean healthy	5090	Tomato leaf curl	5357
Strawberry healthy	456	Tomato target spot	1404
Strawberry leaf scorch	1109		

Table 1. Image statistics for each disease and healthy category.

(a) Before cutting: 800px × 600px **(b)** After cutting: 400px × 400px

Fig. 2. Before and after cropping.

a single layer, which is particularly beneficial for identifying subtle and spatially dispersed disease patterns in agricultural images[28].

Compared to other large-kernel or attention-based models, RepLKNet is specifically advantageous in agricultural scenarios due to its unique ability to balance model complexity and global feature perception. Its design allows for efficient and direct receptive field expansion without incurring excessive computational

Set	Training Set	Validation Set	Test Set
Total Samples	76,694	9,586	9,588

Table 2. Number of training, validation, and test samples.

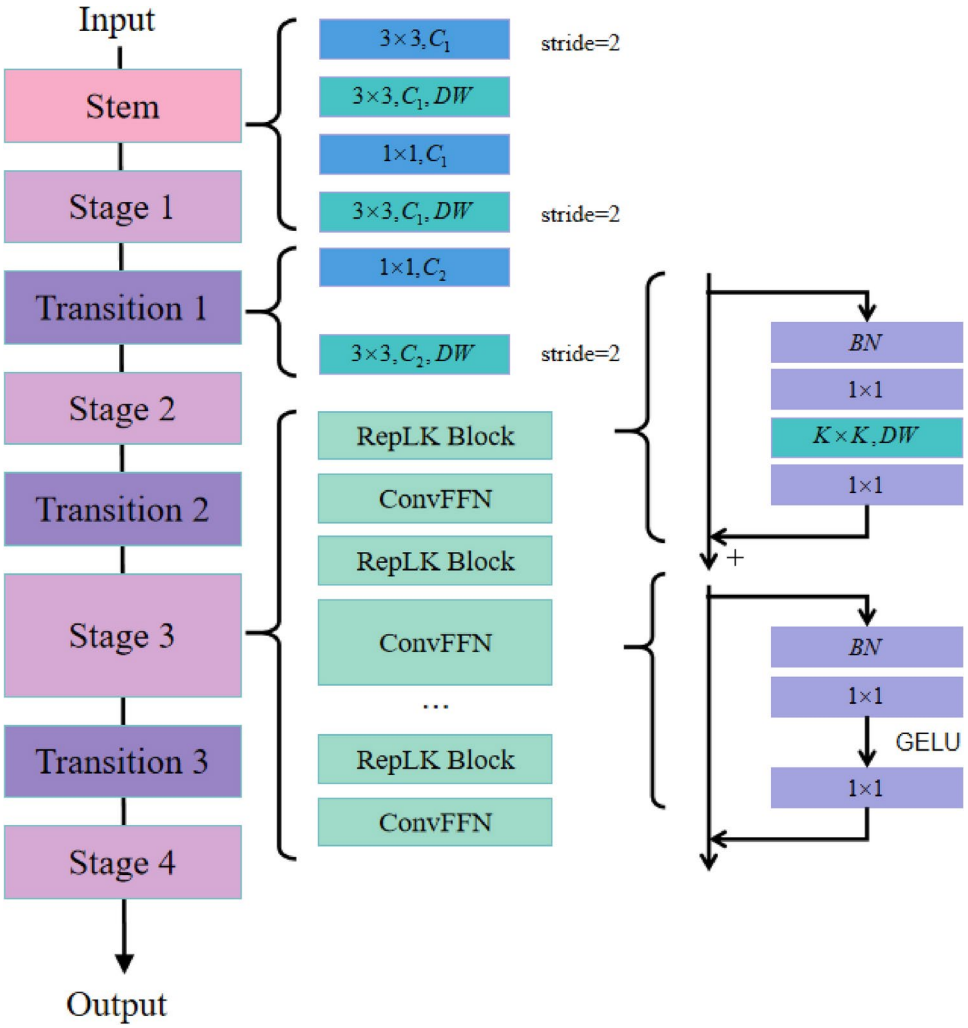


Fig. 3. The structure of RepLKNet.

overhead from self-attention mechanisms or overly deep architectures. This characteristic makes it particularly well-suited for handling high-resolution plant imagery with diverse and spatially variable disease features.

Moreover, recent studies have demonstrated the effectiveness of RepLKNet in agricultural applications. For instance, Liu et al.[29] adopted RepLKNet to enhance a pest recognition model in cotton fields. By integrating large kernel convolutions, their optimized model achieved a 2.8% improvement in average precision, reaching 96.2%, compared to conventional CNN architectures. This practical success further validates the suitability of RepLKNet for tasks that require both local detail sensitivity and global structural understanding, such as plant disease and pest identification.

Therefore, we adopt RepLKNet in this study to leverage its strong global feature extraction capability and proven advantages in agricultural visual recognition tasks.

Although large kernel convolutions traditionally suffer from high computational cost-due to a quadratic increase in FLOPs and parameter count-RepLKNet addresses this limitation through architectural innovations. Specifically, it incorporates structural re-parameterization and depthwise separable convolutions, which significantly reduce the computational overhead while retaining the expressive power of large kernels.

The overall architecture of RepLKNet consists of one Stem, four Stages, and three Transition layers, as illustrated in Fig.3.

Most existing deep learning frameworks do not provide effective support for large convolutional kernels in terms of efficient computation, primarily due to the large number of parameters and floating-point operations

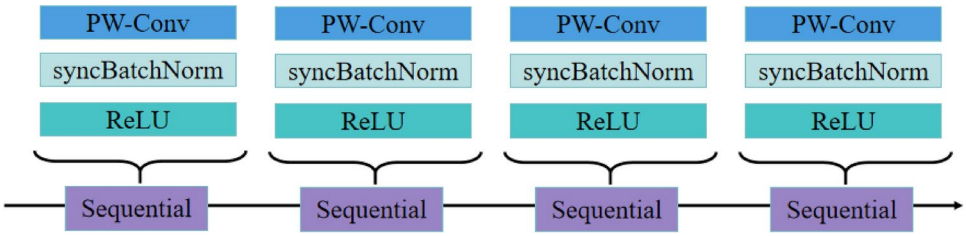


Fig. 4. Stem.

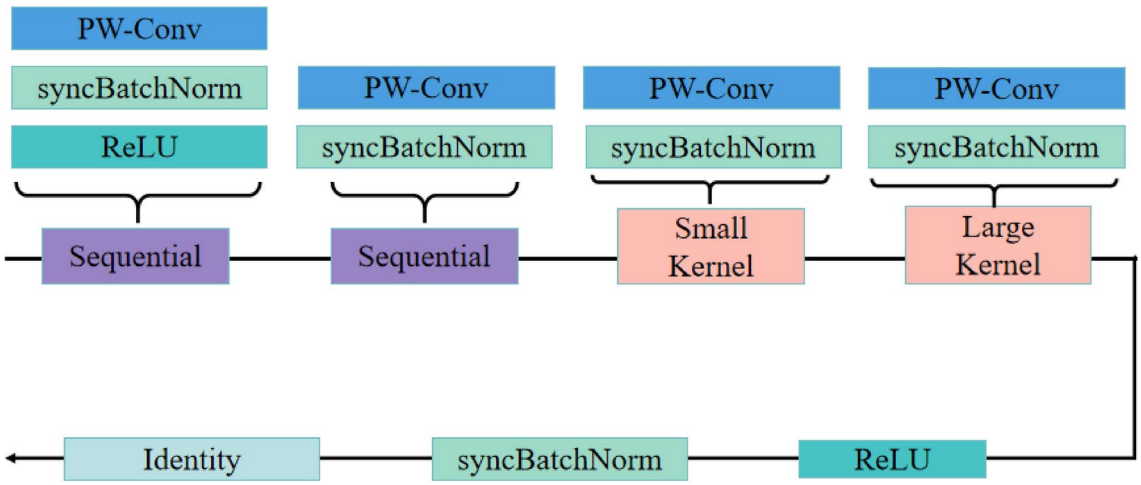


Fig. 5. Stage.

required. However, the RepLKNet network, which employs depthwise separable convolutions and small kernel convolutions for structural re-parameterization[30], significantly alleviates the computational burden caused by large kernel convolutions.

The Stem layer in RepLKNet is designed to perform initial feature extraction and spatial downsampling while maintaining computational efficiency. It consists of a sequence of four convolutional operations: a standard convolution, followed by a depthwise convolution, then another standard convolution, and finally a second depthwise convolution, as shown in Fig.4.

Specifically, the first 3×3 convolution with a stride of 2 reduces the spatial resolution of the input while increasing the feature dimensionality. This is followed by a depthwise convolution with 3×3 kernels to extract local features in a channel-wise manner, enhancing spatial encoding with minimal additional computation. A 1×1 convolution is then applied to refine inter-channel interactions, and a second depthwise convolution with stride 2 is used for further downsampling. This layered design enables the model to efficiently learn low-level representations with rich spatial and channel-wise features, laying a strong foundation for deeper stages in the network.

The Stage layer is the most critical layer in RepLKNet, and there are four Stage layers in the model. The Stage layer consists of a RepLK Block and ConvFNN. The RepLK Block mainly consists of a normalization layer, a 1×1 convolution, a depthwise separable convolution, and a residual connection. This design greatly reduces the computational load of RepLKNet and improves operational efficiency. The RepLK Block contains the most important oversized convolutional kernels in the model. In these four stages, the kernel sizes are 31×31, 29×29, 27×27, and 13×13, respectively, significantly improving the receptive field compared to using small kernel convolutions with multiple levels. The structure is shown in Fig. 5.

Feed-Forward Network (FFN)[31] is the most commonly adopted operation in Transformer Networks. FFN is primarily used as a location-aware feature extractor and as a hidden layer between the encoder and decoder. There is one layer of FFN in each of the encoder and decoder: the first one uplifts the dimension of the input vectors, and the second downsizes the vectors. The FFNs are connected to each other through residual connections and normalization layers, which perform a weighted transformation of the input positional encoding to capture both local and global information of the input data, as well as the global relationships between different positions. The ConvFFN module in the Stage layer is inspired by the FFN operation in Transformers, but it is unique in that it uses a 1×1 convolution instead of the traditional fully connected layer. The structure of ConvFFN is shown in Fig.6.

The Transition layer is located between different stages of the model. It consists of three Sequential blocks, each containing two convolutional layers. The first Sequential block contains a 1×1 convolutional layer, which

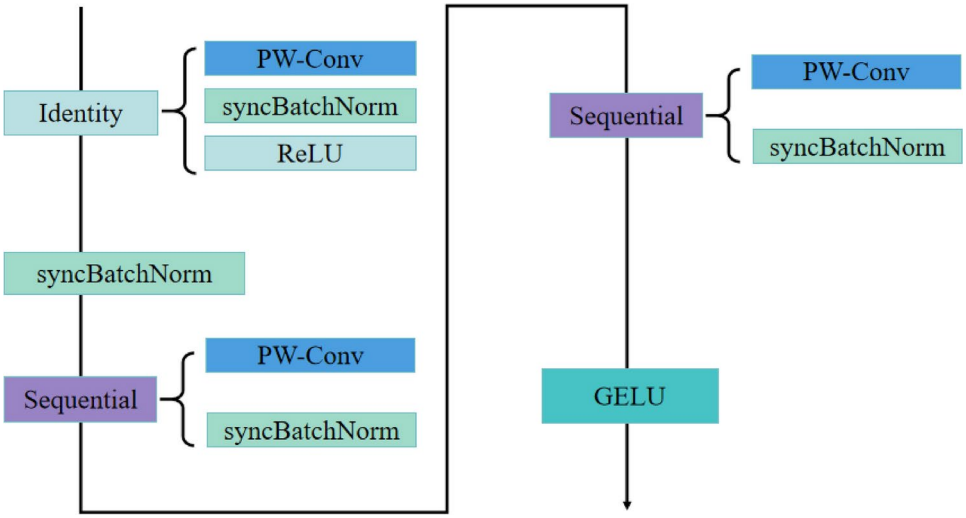


Fig. 6. ConvFNN.

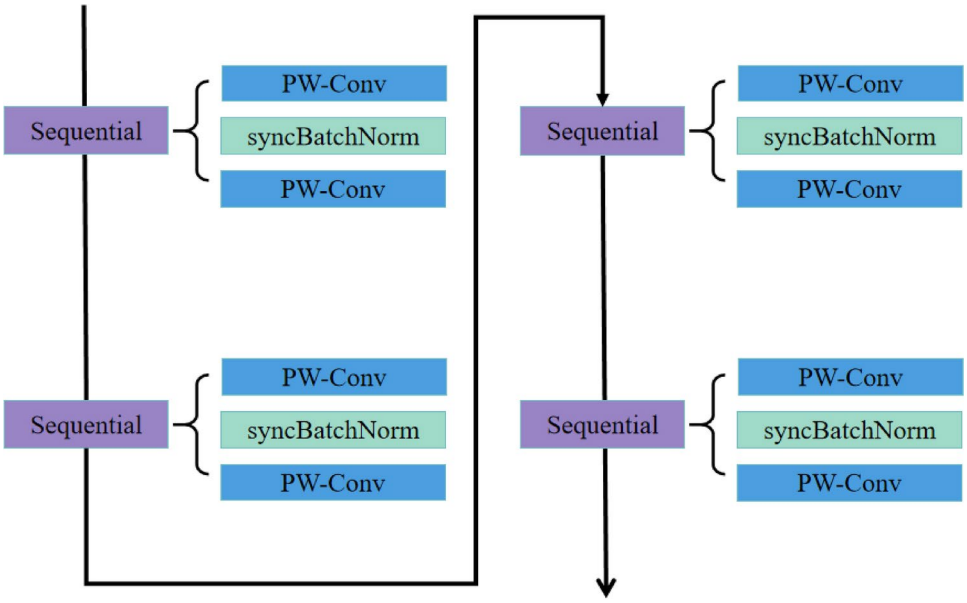


Fig. 7. Transition.

converts 128 input channels to 256 output channels while adjusting the feature resolution and the number of channels. The second Sequential block is a 3×3 grouped convolutional layer with a stride of 2, which is used to downsample the feature map and is a depthwise separable convolution, utilizing the ReLU activation function. The structure is shown in Fig. 7.

Experimental environment

To ensure the validity of the experiments, all experiments were conducted in an environment running on Windows Server 2022 with Intel Xeon Silver 4214R CPUs and 128 GB of RAM. The graphics card used is an Nvidia Geforce RTX 4090. The programming language is Python 3.9.7, the package manager is Conda 4.10.3, the deep learning framework is PyTorch 2.2.1, and the acceleration framework is CUDA 12.2. The specific configurations are shown in Table 3.

To ensure the rigor and fairness of the experiments, all models in this study-including RepLKNet GoogleNet, ResNet50, VGG16, and other baseline networks-were trained using the same set of hyperparameters. These parameters were kept consistent across both the frozen and unfrozen stages of training and were uniformly applied throughout the 5-fold cross-validation process.

The parameter settings were carefully chosen based on empirical tuning and prior literature to balance training stability and convergence speed. A batch size of 100 was used to fully utilize GPU memory without

Hardware	Version Number	Software	Version Number
Operating System	Windows Server 2022 Standard	Programming Language	Python 3.9.7
CPU	Intel(R) Xeon(R) Silver 4214R	Conda	Conda 4.10.3
RAM	128G	Deep Learning Framework	Pytorch 2.2.1
Graphics Card	Nvidia Geforce RTX 4090	Acceleration Framework	CUDA 12.2

Table 3. Environment of the experiment.

Freeze Parameter	Parameter Value	Unfreezing Parameter	Parameter Value
Batch Size	100	Batch Size	100
Epoch	10	Epoch	10
Learning Rate	0.003	Learning Rate	0.0005
Weight Decay	0.0002	Weight Decay	0.0002
Optimizer	AdamW	Optimizer	AdamW
Scheduler	LambdaLR	Scheduler	LambdaLR

Table 4. Parameters related to freeze training.

compromising training efficiency. The learning rate was set to 0.003 during the freezing stage to allow rapid convergence of the newly added classification head, and then reduced to 0.0005 in the unfreezing stage to fine-tune the entire model more delicately. AdamW was selected as the optimizer due to its strong generalization performance and ability to mitigate overfitting through decoupled weight decay, which was set to 0.0002. A LambdaLR scheduler was adopted to adaptively adjust the learning rate throughout training.

The detailed training configurations for both the freezing and unfreezing stages are summarized in Table.4.

Indicators for experimental evaluation

To comprehensively evaluate the performance of different models, we adopted several widely used evaluation metrics, including Overall Accuracy (OA), Average Accuracy (AA), Kappa coefficient, and Precision (P). These metrics are derived from the confusion matrix and collectively reflect the model's effectiveness and reliability in multiclass classification tasks.

OA measures the proportion of correctly classified samples among all samples and reflects the general classification performance of the model. It is calculated as:

$$\mathrm{OA} = \frac{\sum_{i=1}^{C} x_{ii}}{\sum_{i=1}^{C} \sum_{j=1}^{C} x_{ij}} \tag{1}$$

where x_{ii} is the number of correctly predicted samples for class *i*, x_{ij} represents the number of samples whose true label is class *i* but predicted as class *j*, and *C* is the total number of classes.

AA is the mean of the individual accuracies for each class, calculated as:

$$\mathrm{AA} = \frac{1}{C} \sum_{i=1}^{C} \frac{x_{ii}}{\sum_{j=1}^{C} x_{ij}} \tag{2}$$

This metric accounts for class imbalance by equally weighting each class, regardless of its sample size.

P measures the proportion of correctly predicted samples among all samples predicted to belong to a specific class. It evaluates the model's ability to avoid false positives and is defined as:

$$\mathrm{P} = \frac{1}{C} \sum_{i=1}^{C} \frac{x_{ii}}{\sum_{j=1}^{C} x_{ji}} \tag{3}$$

where $\sum_{j=1}^{C} x_{ji}$ is the total number of samples predicted as class *i*, and x_{ii} is the number of correct predictions for class *i*. This metric helps to assess the model's accuracy in assigning correct labels to its predictions.

Kappa is used to evaluate the agreement between predicted and true labels while accounting for the possibility of random agreement. It is defined as:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{4}$$

where $P_o = \frac{1}{N}\sum_{i=1}^{C} x_{ii}$ is the observed agreement, $P_e = \frac{1}{N^2}\sum_{i=1}^{C}(x_{i+} \cdot x_{+i})$ is the expected agreement by chance, x_{i+} is the total number of actual samples in class *i*, x_{+i} is the total number of predicted samples in class *i*, and *N* is the total number of samples.

Results

Analysis of RepLKNet experimental results

By leveraging the pre-trained ImageNet-1 K-224 weights for transfer learning, RepLKNet demonstrated exceptional performance in agricultural disease recognition, achieving a peak OA of 96.03% in Fold 1, as shown in Table.5. This high OA value is indicative of the model's ability to effectively identify diverse disease categories, an essential capability for real-world agricultural applications where timely and accurate diagnosis is critical to minimizing plant losses.

RepLKNet's superior performance can be attributed to its use of large kernel convolutions, which allow the model to capture both fine-grained local textures and broader spatial patterns. This characteristic enhances the model's ability to recognize complex patterns in high-resolution agricultural images, making it particularly well-suited for tasks like disease classification. In fact, the OA values across all five folds consistently remain high, ranging from 95.55% to 96.03%, with a minimal standard deviation of 0.17%. This indicates not only strong accuracy but also the robustness of the model, as it can maintain stable performance across different data subsets, as summarized in Table.5.

Moreover, the model achieved strong AA values between 94.72% and 95.31%, and Kappa values between 95.35% and 95.86%, further underscoring the model's effectiveness in balancing classification across different disease categories. The P values, varying from 95.59% to 95.97%, reflect the model's consistency in making reliable predictions, further confirming the benefits of the large kernel convolutions in ensuring both accuracy and stability.

The use of large kernel convolutions in RepLKNet expands the effective receptive field, enabling the model to better capture global patterns and contextual information. This capability is especially crucial for recognizing agricultural diseases, where spatial relationships between different regions of the image play a key role in classification. By maintaining a balance between local texture features and broader spatial context, RepLKNet leverages its architecture to improve both classification accuracy and robustness, making it an ideal choice for practical applications in agricultural disease detection.

In conclusion, the strong and stable performance of RepLKNet across all folds, coupled with its minimal variance in both OA and precision metrics, emphasizes the significant advantages of large kernel convolutions in enhancing the model's ability to detect and classify agricultural diseases. These results suggest that RepLKNet's architectural choice not only enables precise disease identification but also ensures that the model performs reliably across diverse real-world scenarios.

We further analyzed the progression of OA accuracy during the unfreezing stage. As shown in Fig. 8, the OA curves for all five folds exhibit a clear upward trend at the beginning of this phase, typically rising from around 92% to approximately 95%. This notable improvement indicates that enabling end-to-end model training during the unfreezing stage significantly enhanced the network's ability to learn domain-specific features crucial for accurate agricultural disease classification.

As training proceeded, the OA values gradually plateaued, indicating that the models had reached convergence with both high accuracy and stable performance. The uniform improvement and eventual stabilization across all folds further validate the effectiveness of the fine-tuning strategy and the strong generalization capability of the RepLKNet model.

Ablation study

To further validate the effectiveness of large kernel convolutions in agricultural disease recognition, an ablation study was conducted where the large kernel convolutions in the RepLKNet model were replaced with commonly used smaller kernel sizes, specifically 3x3 and 5x5 convolutions. The purpose of this experiment was to assess how the choice of kernel size influences the model's performance in recognizing agricultural diseases.

The results of the ablation study are shown in Table 6. For each fold, we compared the original RepLKNet model with the versions using 3x3 and 5x5 convolutions, referred to as RepLKNet-3x3 and RepLKNet-5x5, respectively.

From the results presented in Table.6, it is clear that the RepLKNet model with large convolution kernels consistently outperforms the models with smaller 3x3 and 5x5 convolutions across all five folds in terms of OA, AA, and Kappa coefficient. Specifically, the OA for RepLKNet ranged from 95.55% to 96.03%, while the models

Fold	Model	OA (%)	AA (%)	Kappa (%)	P (%)
1	RepLKNet-1	96.03	94.78	95.86	95.97
2	RepLKNet-2	95.96	94.84	95.80	95.84
3	RepLKNet-3	95.99	95.31	95.83	95.90
4	RepLKNet-4	95.97	95.20	95.80	95.87
5	RepLKNet-5	95.55	94.72	95.35	95.59
Std		0.17	0.24	0.17	0.14

Table 5. 5-fold cross-validation results of RepLKNet (%).

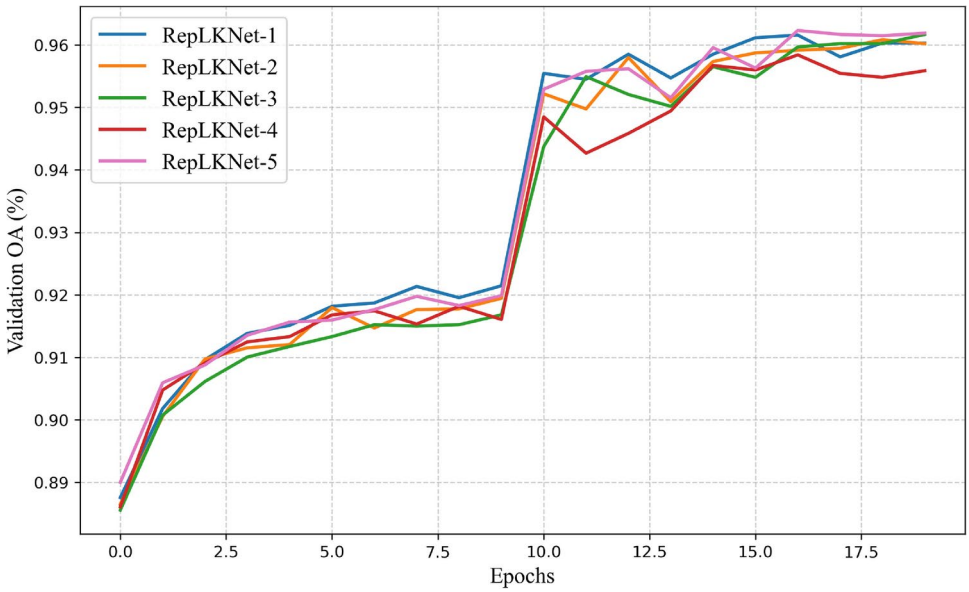


Fig. 8. OA accuracy curves of 5-fold cross-validation during the unfreezing stage.

Fold	Model	OA (%)	AA (%)	Kappa (%)	P (%)
1	RepLKNet	96.03	94.78	95.86	95.97
	RepLKNet-3x3	95.06	93.73	94.85	94.96
	RepLKNet-5x5	94.93	93.48	94.72	94.83
2	RepLKNet	95.96	94.84	95.80	95.84
	RepLKNet-3x3	94.96	93.68	94.75	94.84
	RepLKNet-5x5	95.17	93.61	94.97	94.92
3	RepLKNet	95.99	95.31	95.83	95.90
	RepLKNet-3x3	95.19	94.10	94.99	95.01
	RepLKNet-5x5	95.50	94.53	95.31	95.24
4	RepLKNet	95.97	95.20	95.80	95.87
	RepLKNet-3x3	95.11	94.43	94.90	94.93
	RepLKNet-5x5	94.98	94.44	94.77	94.90
5	RepLKNet	95.55	94.72	95.35	95.59
	RepLKNet-3x3	94.75	93.73	94.53	94.82
	RepLKNet-5x5	94.76	93.78	94.54	94.80

Table 6. Performance comparison of RepLKNet variants under 5-fold cross-validation (%).

with smaller kernels (RepLKNet-3x3 and RepLKNet-5x5) achieved lower OA values, ranging from 94.75% to 95.17% for 3x3 and from 94.72% to 95.31% for 5x5. Similarly, the AA and Kappa coefficient results follow the same pattern, with RepLKNet consistently achieving higher values than the smaller kernel models.

In addition, the P values further support this trend. The RepLKNet model consistently achieves the highest precision across all five folds, ranging from 95.59% to 95.97%, indicating that the model makes fewer false positive predictions compared to its 3x3 and 5x5 counterparts. In contrast, RepLKNet-3x3 and RepLKNet-5x5 yield slightly lower precision values, fluctuating between 94.82%–95.01% and 94.80%–95.24%, respectively. This further highlights the advantage of large kernels in maintaining more accurate and reliable predictions in multi-class disease classification tasks.

These results suggest that larger kernels are highly beneficial for improving the model's performance in agricultural disease recognition. The larger kernels allow the model to capture more complex patterns and dependencies in the data, which is particularly advantageous in agricultural applications where diseases often manifest in large, spatially distributed areas on plants. In contrast, smaller kernels, while commonly used in many deep learning models, seem to limit the model's ability to fully capture the spatial relationships within the image data, leading to lower performance in terms of classification accuracy, consistency, and precision.

In conclusion, the ablation study reinforces the importance of large kernel convolutions in the context of agricultural disease recognition, suggesting that the enhanced receptive field and improved feature extraction

offered by larger kernels are crucial for achieving high classification accuracy, stable performance metrics, and precise disease identification.

Comparative experimental

To comprehensively evaluate the effectiveness of the proposed RepLKNet model in agricultural disease recognition, we conducted comparative experiments with several widely used convolutional neural networks and a transformer-based model under 5-fold cross-validation. The selected baseline models include MobileNetV3, ResNet50, VGG16, GoogLeNet, and Swin Transformer. These models were chosen based on their popularity, architectural diversity, and relevance to the task.

Specifically, ResNet50 and VGG16 are classical CNN architectures that rely on small convolutional kernels. They were selected to serve as representative small-kernel models, allowing for a direct comparison with RepLKNet, which adopts large-kernel convolutions. This enables us to evaluate whether expanding the receptive field through large kernels can improve feature extraction in agricultural disease images. MobileNetV3 and GoogLeNet are lightweight models often used in real-time or resource-constrained scenarios, providing insight into the trade-off between accuracy and model complexity. Meanwhile, Swin Transformer is a recent vision transformer model that leverages hierarchical attention mechanisms and global context modeling, making it suitable for comparison with convolutional methods from a different architectural perspective.

The performance of each model across the five folds is summarized in Table.7.

From the experimental results, RepLKNet consistently achieves top-level performance, with OA ranging from 95.55% to 96.03% and Kappa values between 95.35% and 95.86%. This demonstrates its strong capability in accurately recognizing diverse disease categories across different folds. The Swin Transformer shows comparable results, with slightly higher OA in Fold 1, reflecting the power of attention-based mechanisms in capturing global contextual information. However, RepLKNet remains more consistent across all folds.

In terms of P, RepLKNet also achieves the highest or near-highest values in each fold, with scores ranging from 95.59% to 95.97%. This indicates that the model not only maintains high overall and balanced accuracy but also excels at minimizing false positive predictions. Compared to other models such as ResNet50 and Swin

Fold	Model	OA (%)	AA (%)	Kappa (%)	P (%)
1	MobileNetV3	94.40	93.54	94.17	94.28
	ResNet50	95.38	94.72	95.19	95.22
	VGG16	93.60	90.91	93.33	93.49
	GoogLeNet	94.98	94.19	94.77	94.75
	Swin Transformer	96.12	94.98	95.96	95.83
	RepLKNet	96.03	94.78	95.86	95.97
2	MobileNetV3	94.84	93.37	94.62	94.51
	ResNet50	95.62	94.17	95.44	95.26
	VGG16	93.65	91.21	93.39	93.41
	GoogLeNet	95.08	93.86	94.87	95.52
	Swin Transformer	95.85	94.68	95.68	94.95
	RepLKNet	95.96	94.84	95.80	95.84
3	MobileNetV3	94.41	93.70	94.17	94.17
	ResNet50	95.62	94.67	95.44	95.35
	VGG16	94.17	91.83	93.92	93.70
	GoogLeNet	95.18	94.25	94.98	94.85
	Swin Transformer	95.84	94.85	95.66	95.46
	RepLKNet	95.99	95.31	95.83	95.90
4	MobileNetV3	94.71	94.15	94.49	94.50
	ResNet50	95.35	94.35	95.15	95.04
	VGG16	92.99	90.36	92.70	92.76
	ConvNeXt	95.76	94.98	95.59	95.68
	GoogLeNet	94.82	94.16	94.60	94.66
	Swin Transformer	95.96	95.00	95.79	95.73
	RepLKNet	95.97	95.20	95.80	95.87
5	MobileNetV3	94.13	93.26	93.87	94.13
	ResNet50	95.44	94.29	95.25	95.22
	VGG16	93.66	91.42	93.38	93.56
	GoogLeNet	94.88	93.86	94.66	94.38
	Swin Transformer	95.61	94.60	95.42	94.72
	RepLKNet	95.55	94.72	95.35	95.59

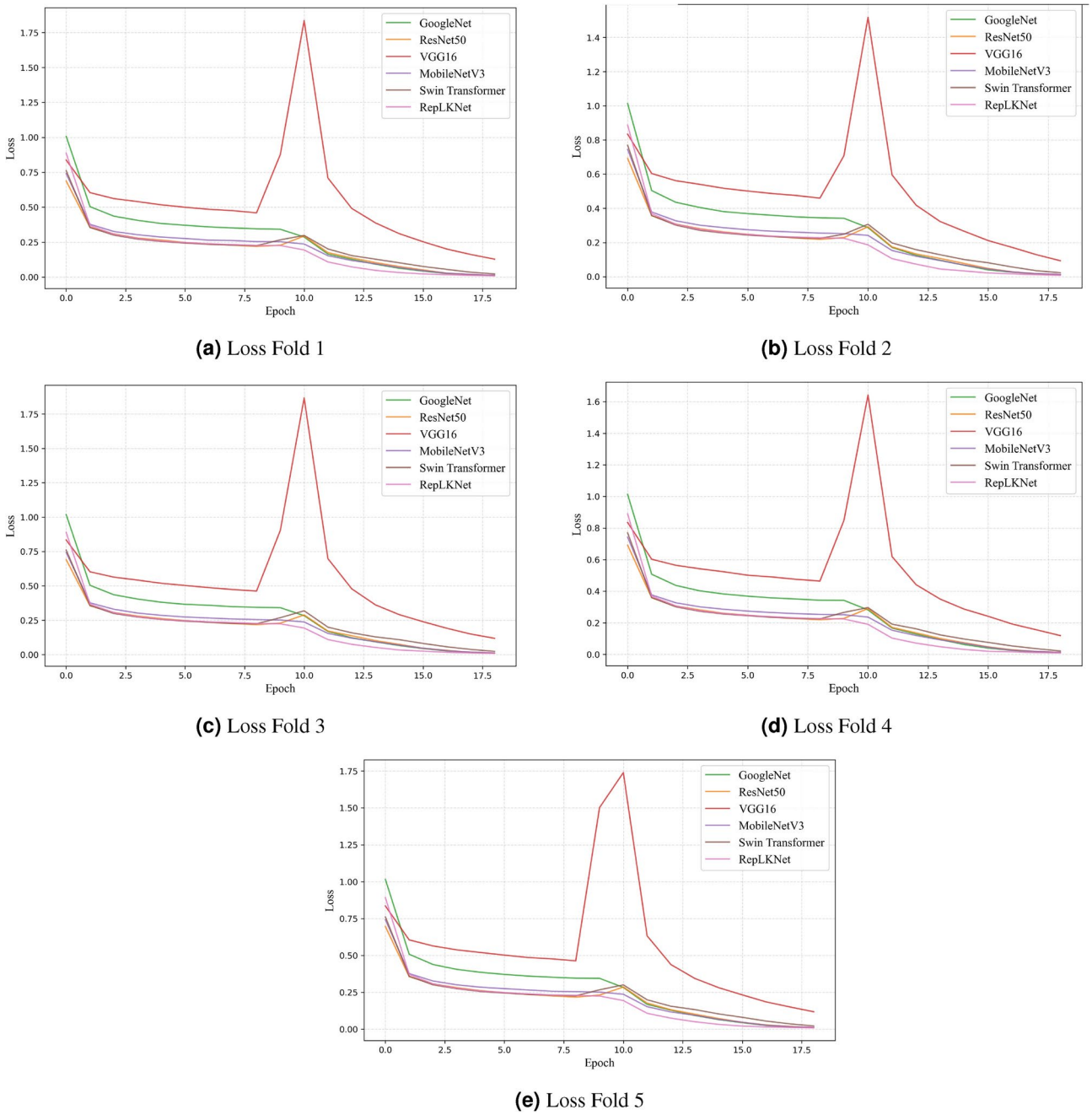
Table 7. Performance comparison of different models under 5-fold cross-validation (%).

Transformer, which show competitive performance in some folds, RepLKNet maintains a slight but consistent edge in precision, highlighting its robustness and reliability in correctly identifying disease classes without over-predicting.

To assess the models' training stability, we also analyzed the loss curves across all five folds, as shown in Fig.9. The loss curves provide valuable insights into how each model converges during training. Lower and stable loss values indicate better training stability and efficient learning.

From the loss plots, we observe that, apart from VGG16, all other models show a clear downward trend in their loss values, eventually converging towards zero. This indicates that these models are effectively learning and minimizing their prediction errors as training progresses. Specifically, RepLKNet demonstrates a smooth and steady decline in loss, consistent with its stable performance observed in the OA curves. The loss steadily decreases towards zero, reflecting the model's ability to efficiently reduce error and improve predictions over time. This is indicative of RepLKNet's robustness in learning complex patterns in agricultural disease recognition tasks.

In contrast, VGG16 exhibits significant fluctuations in its loss curve, with frequent increases in loss throughout the training process. This instability further corroborates its poorer performance in terms of OA, as



(a) Loss Fold 1

(b) Loss Fold 2

(c) Loss Fold 3

(d) Loss Fold 4

(e) Loss Fold 5

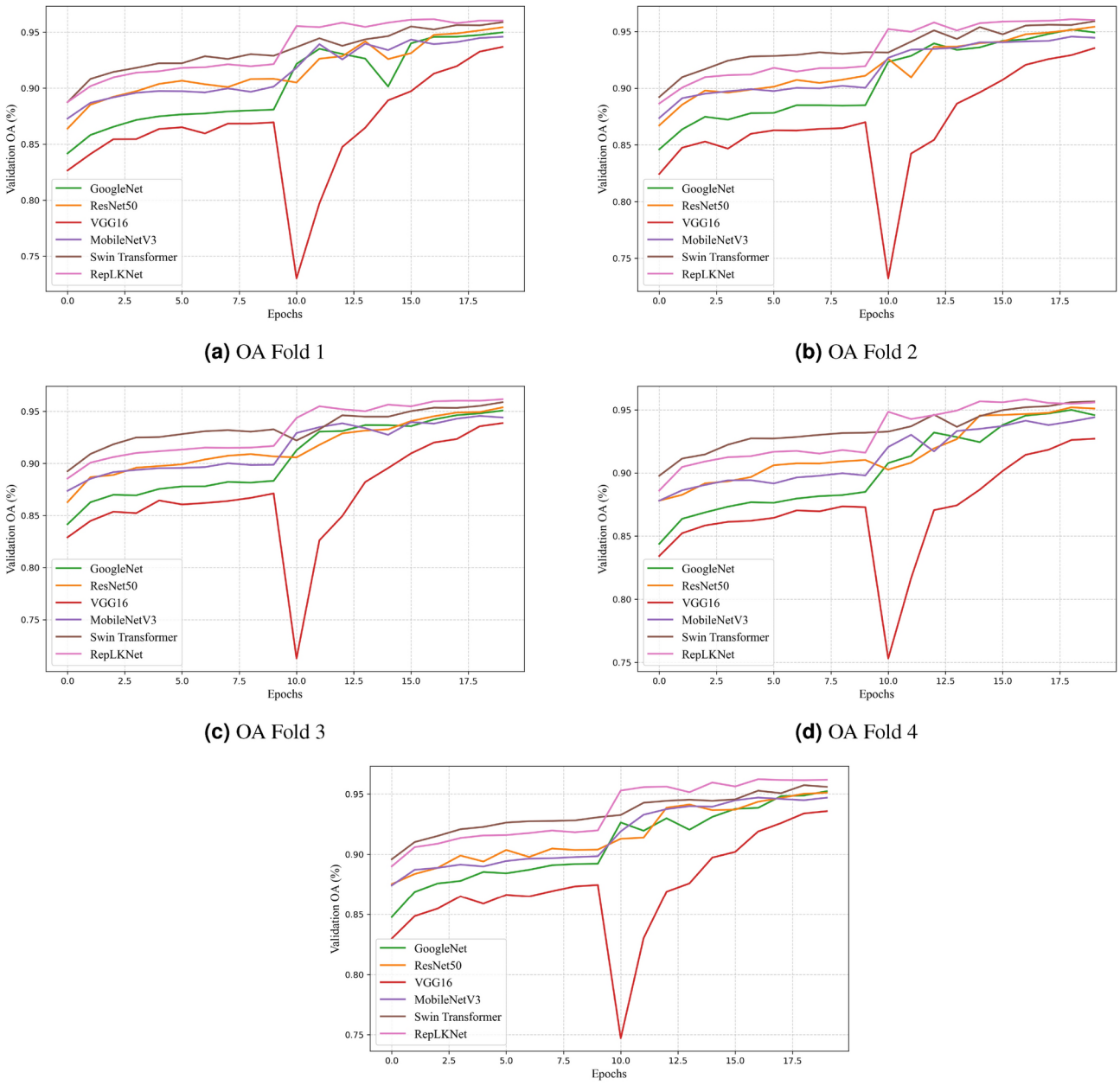
Fig. 9. Loss curves for each fold in 5-fold cross-validation.

the model struggles to effectively learn and adapt to the complex patterns within the agricultural dataset. The loss fluctuations suggest that VGG16 may face challenges in capturing the rich variability of disease symptoms, likely due to its shallow architecture and limited feature representation capabilities.

The loss curves also indicate that Swin Transformer, ResNet50, and GoogLeNet experience relatively smooth and steady decreases in loss, which aligns with their stable and competitive performance in OA accuracy.

To further visualize the performance trends, we plotted the OA accuracy curves for each model across all five folds, as shown in Fig. 10. These plots illustrate the dynamic learning behavior and stability of each model during training. It can be observed that RepLKNet maintains a consistently high OA throughout training, with relatively smooth and convergent curves across all folds. In contrast, models such as VGG16 not only start from a lower baseline but also exhibit more fluctuations and slower convergence, indicating instability and limited feature representation during the training process.

Swin Transformer displays stable and competitive performance, particularly evident in Fold 1, where its OA briefly surpasses that of RepLKNet. However, its performance in subsequent folds varies slightly more, suggesting that while it excels in capturing global dependencies, its generalization may be less consistent than that of RepLKNet on this specific dataset.



(a) OA Fold 1

(b) OA Fold 2

(c) OA Fold 3

(d) OA Fold 4

(e) OA Fold 5

Fig. 10. OA accuracy curves for each fold in 5-fold cross-validation.

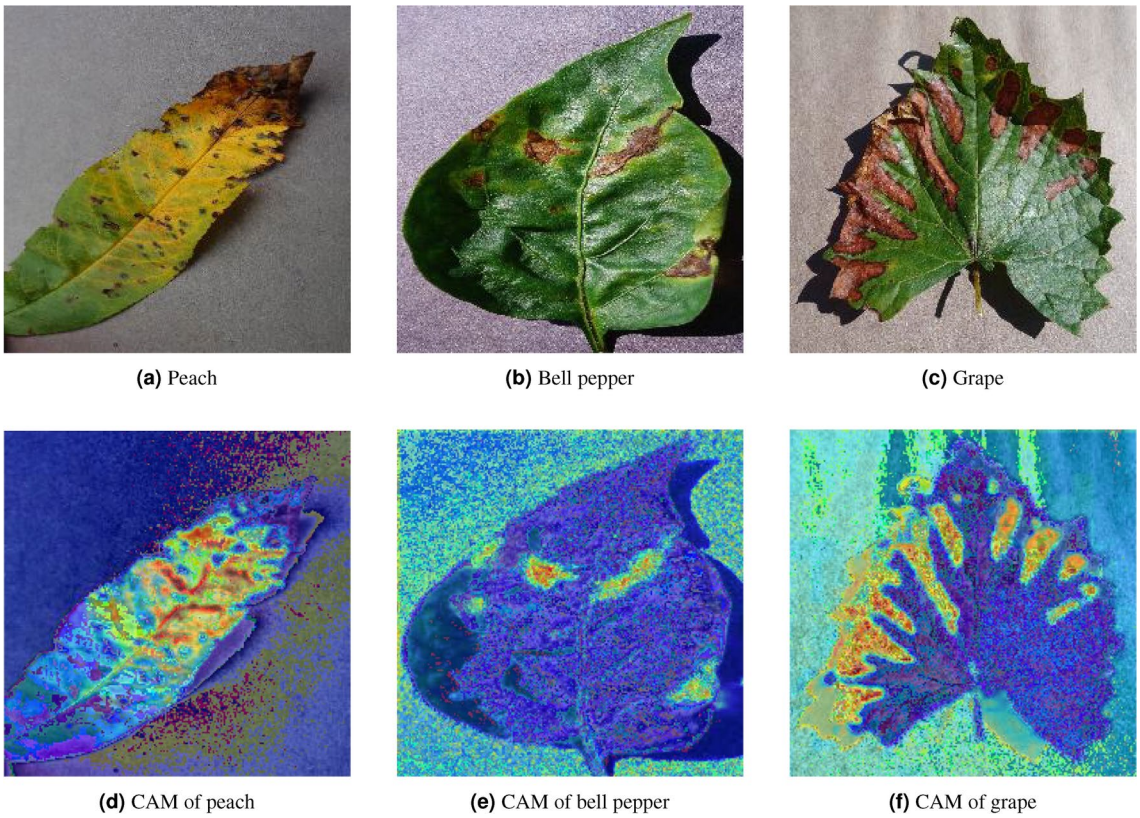


Fig. 11. Grad-CAM images of three different plants, with the warmer areas being those attacked by the disease.

In comparison, ResNet50 and GoogLeNet perform well among traditional CNNs but still fall short of RepLKNet, especially in terms of average accuracy and Kappa values. These models, which rely on small convolution kernels and deep structures, are effective to a certain extent but may lack the broader receptive fields necessary for capturing complex patterns in high-resolution leaf and plant imagery.

VGG16 demonstrates the weakest performance overall, with OA dropping as low as 92.99% and consistently lower Kappa scores. Its shallow architecture and reliance on simple stacking of small convolutional layers make it less suitable for the rich variability present in agricultural disease datasets. MobileNetV3, although lightweight and efficient, also lags behind due to limited representational capacity.

The results demonstrate that the use of large kernel convolutions in RepLKNet provides a clear advantage in capturing both fine-grained local textures and broader spatial patterns, which are essential for accurate disease classification. By expanding the effective receptive field, RepLKNet enhances the model's discriminative capability without relying on attention mechanisms, achieving a strong balance between accuracy, robustness, and architectural simplicity.

Grad-CAM heat map visualisation

Grad-CAM[32] is a powerful technique for interpreting deep learning models by generating heat maps that highlight important regions of an input image that contribute to the model's decision. To evaluate whether RepLKNet effectively learns relevant features from plant disease images, we applied Grad-CAM to three different types of plant disease images and visualized the heat maps.

The Grad-CAM heat maps highlight regions in the image that the model considers important for its decision. Warmer colors (red and yellow) correspond to areas of high activation, while cooler colors (blue and green) indicate regions with low activation. The red areas are typically concentrated in the parts of the plant that are severely affected by the disease, demonstrating the model's ability to focus on the regions that are most relevant for classification. This is particularly important in agricultural disease recognition, where identifying the disease-affected regions is crucial for accurate diagnosis.

As shown in Fig.11, which displays the Grad-CAM visualizations for peach, bell pepper, and grape leaves, we observe that the warm-colored regions in the heat maps closely align with the visually apparent disease symptoms, such as spots or discoloration, on the plant leaves. These visualizations were generated using Fold 1 of the 5-fold cross-validation. Specifically, in the original leaf images-peach (Fig.11a), bell pepper (Fig.11b), and grape (Fig.11c)-distinct disease symptoms are clearly visible. Correspondingly, the Grad-CAM visualizations-Fig.11d (peach), Fig.11e (bell pepper), and Fig. 11f (grape)-highlight regions with high model attention that strongly coincide with the diseased areas in the original images. This consistency reinforces the model's effectiveness in identifying disease-related features.

The Grad-CAM visualizations further highlight the distinct advantages of using large kernel convolutions in RepLKNet for plant disease recognition. By leveraging larger receptive fields, RepLKNet can capture broader spatial patterns in the images, which are essential for identifying complex disease symptoms. The warm-colored regions in the Grad-CAM heat maps are larger and more comprehensive compared to models with smaller convolution kernels, reflecting the ability of RepLKNet's large kernel convolutions to aggregate contextual information from a wider area of the plant leaf. This allows the model to detect subtle changes in texture and color that might be overlooked by smaller kernels. In contrast, models using smaller kernels, such as VGG16 or ResNet50, may struggle to capture these larger-scale patterns, potentially leading to less accurate disease identification.

These Grad-CAM visualizations substantiate RepLKNet's effectiveness by showing that the model focuses on the most relevant and critical regions affected by disease, demonstrating the capability of large kernel convolutions to capture both fine-grained local details and broader contextual patterns. This combination enhances the model's robustness in recognizing agricultural diseases, making it a valuable tool for real-world applications where accurate and timely disease detection is vital.

Discussion

This study demonstrates the potential of large kernel convolutional networks, specifically RepLKNet, in improving the accuracy and robustness of agricultural disease recognition across multiple plant types. By achieving an overall accuracy of 96.03%, with an average accuracy of 94.78% and a Kappa coefficient of 95.85%, our approach outperforms many traditional CNN-based methods, particularly in scenarios requiring the recognition of spatially distributed and subtle disease patterns.

Compared with recent transformer-based approaches, our method presents both strengths and trade-offs. For instance, Pacal[7] proposed a MaxViT-based maize disease detection framework and achieved 99.24% accuracy by combining multiple datasets and optimizing transformer modules. Similarly, Kunduracioglu et al.[8] demonstrated that fine-tuned vision transformer (ViT) models like Swinv2-Base could achieve 100% accuracy on small and clean grape disease datasets. Furthermore, Pacal et al.[21] explored ViT models and soft ensemble strategies for corn disease detection, reaching up to 100% accuracy on the curated CD&S dataset.

While these transformer-based models achieve impressive results on relatively constrained datasets, they often come with high computational costs and lack interpretability in spatial pattern learning, especially when the symptoms are subtle and scattered. In contrast, our approach based on RepLKNet strikes a balance by employing large kernel convolutions to directly capture long-range dependencies without the complexity and resource overhead of transformer architectures. This is especially advantageous in agricultural scenarios where high-resolution imagery and localized symptoms require both global and fine-grained feature understanding. Similar findings have been reported in the context of non-agricultural domains, where large kernels were shown to improve global shape recognition in medical and industrial tasks[28,29].

Nevertheless, our method has several limitations. First, the model's adaptability to unseen plants or disease types remains a challenge. Unlike Pacal[7], who combined PlantVillage, PlantDoc, and CD&S to create a large-scale benchmark dataset, our dataset-while diverse-still lacks full coverage of agricultural disease diversity. Generalization to plants beyond the training set may be limited. This reflects a broader trend in plant disease recognition research, where many models demonstrate high accuracy on known categories but struggle with cross-species transferability[16,19]. Future research should consider expanding the dataset and employing domain adaptation techniques-such as adversarial training or style transfer[1,13]-to improve cross-plant robustness.

Secondly, the fixed-size image cropping approach used in preprocessing may cause important contextual information to be lost, particularly when disease symptoms appear on the edges or in irregular patterns. Methods such as dynamic region of interest (ROI) extraction or adaptive resizing, as seen in object detection pipelines like YOLOv7[9] and AgriDet[22], may provide better localization capabilities in future implementations. Additionally, hybrid preprocessing pipelines that incorporate semantic segmentation or attention-guided cropping[10] may help preserve critical disease-relevant regions.

Moreover, although large kernels effectively extract global features, they may overlook fine-grained local patterns essential for detecting early-stage or mild diseases. Hybrid architectures that combine large kernels with smaller receptive fields or attention mechanisms-such as GhostNet-Triplet-YOLOv8 s[15] or ULEN[16]-could enhance performance by integrating both macro and micro spatial cues. This balance is particularly important in field conditions where symptoms can vary in size and visual prominence across growth stages.

Another critical challenge is the real-time applicability of our model. While RepLKNet shows strong accuracy, the use of large kernels, even with depthwise separable convolution, still results in high training and inference costs. In contrast, models such as MobileViT used in Pacal et al.[21] are designed for lightweight deployment on mobile and edge devices. Recent works have successfully employed model compression strategies such as pruning, quantization, and knowledge distillation to reduce computational burdens while maintaining accuracy[12,17]. Future research should explore similar strategies to improve deployment feasibility in field conditions, including integration with drones, smartphones, or agricultural robots[22].

Finally, practical implementation also depends on environmental variability, including lighting conditions, leaf occlusion, background clutter, and growth stages of plants. While our model demonstrated good performance under controlled conditions, robustness under these real-world factors requires further evaluation. Synthetic data augmentation, such as Generative Adversarial Networks (GANs), and domain generalization methods have been suggested as effective techniques to simulate such variabilities and improve model resilience[19,20].

Conclusions

In this study, we proposed a plant disease recognition framework based on RepLKNet, which utilizes large convolutional kernels (31×31) to capture both local and global features from agricultural disease images. The model demonstrated strong performance, achieving an overall accuracy of 96.03%, an average accuracy of 94.78%, and a Kappa coefficient of 95.86%. These results underscore the potential of large kernel convolution in enhancing agricultural disease recognition, providing an important advancement in the application of deep learning to this domain.

The main contribution of this research lies in the effective use of RepLKNet's large kernels to improve the extraction of spatial and long-range dependencies in agricultural disease images. The incorporation of depthwise separable convolutions helped mitigate the computational burden, while transfer learning significantly accelerated model convergence and improved performance.

However, despite these advances, challenges remain, particularly in terms of real-time disease recognition. While large kernel convolutions offer better feature extraction, the model's training time is still relatively long, limiting its practical deployment for real-time applications. Future research can explore methods such as model pruning, optimization techniques, and enhanced transfer learning strategies to improve both training efficiency and inference speed, making the model more suitable for real-time disease detection in agricultural environments.

Additionally, while the model was tested on a diverse set of 61 agricultural disease categories, the variety of agricultural diseases is vast, and the model's generalization ability could be further improved. Future work should focus on enhancing the model's robustness to handle a broader range of disease types and adapt to different datasets, which could include more varied plant species and environmental conditions.

Data availability

The data supporting the findings of this study are publicly available and were sourced from an open dataset on Kaggle. The dataset can be accessed at https://www.kaggle.com/code/gpiosenka/efficentnet-mobilenet-f1-s-93-93/input.

Received: 21 January 2025; Accepted: 7 May 2025
Published online: 15 May 2025

References

1. Zhang, Y. & Diao, X. The changing role of agriculture with economic structural change-the case of china. *China Economic Review* **62**, 101504. https://doi.org/10.1016/j.chieco.2020.101504 (2020).
2. Kashyap, D., Subramanyam, N. *et al.* Robustness to augmentations as a generalization metric. *arXiv preprint*arXiv:2101.06459 https://doi.org/10.48550/arXiv.2101.06459 (2021).
3. Gulzar, Y. & Ünal, Z. Optimizing pear leaf disease detection through pl-densenet. *Applied Fruit Science* **67**, 40 (2025).
4. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708 (2017).
5. Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, 6105–6114 (PMLR, 2019).
6. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. 770–778 (2016).
7. Pacal, I. Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model. *Expert Systems with Applications* **238**, 122099 (2024).
8. Kunduracioglu, I. & Pacal, I. Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases. *Journal of Plant Diseases and Protection* **131**, 1061–1080 (2024).
9. Umar, M. *et al.* Precision agriculture through deep learning: Tomato plant multiple diseases recognition with cnn and improved yolov7. *IEEE Access* (2024).
10. Kaya, Y. & Gürsoy, E. A novel multi-head cnn design to identify plant diseases using the fusion of rgb images. *Ecological Informatics* **75**, 101998 (2023).
11. Thakur, P. S., Sheorey, T. & Ojha, A. Vgg-icnn: A lightweight cnn model for crop disease identification. *Multimedia Tools and Applications* **82**, 497–520 (2023).
12. Alkanan, M. & Gulzar, Y. Enhanced corn seed disease classification: leveraging mobilenetv2 with feature augmentation and transfer learning. *Frontiers in Applied Mathematics and Statistics* **9**, 1320177 (2024).
13. Amri, E. et al. Advancing automatic plant classification system in saudi arabia: introducing a novel dataset and ensemble deep learning approach. *Modeling Earth Systems and Environment* **10**, 2693–2709 (2024).
14. Gulzar, Y. Enhancing soybean classification with modified inception model: A transfer learning approach. *Emirates Journal of Food & Agriculture (EJFA)* **36** (2024).
15. Li, J., Liu, Z. & Wang, D. A lightweight algorithm for recognizing pear leaf diseases in natural scenes based on an improved yolov5 deep learning model. *Agriculture* **14**, 273. https://doi.org/10.3390/agriculture14020273 (2024).
16. Wang, B. et al. An ultra-lightweight efficient network for image-based plant disease and pest infection detection. *Precision Agriculture* **24**, 1836–1861 (2023).
17. Gulzar, Y., Ünal, Z., Kızıldeniz, T. & Umar, U. M. Deep learning-based classification of alfalfa varieties: A comparative study using a custom leaf image dataset. *MethodsX* **13**, 103051 (2024).
18. Gulzar, Y., Ünal, Z., Ayoub, S. spsampsps Reegu, F. A. Exploring transfer learning for enhanced seed classification: pre-trained xception model. In *International Congress on Agricultural Mechanization and Energy in Agriculture*, 137–147 (Springer, 2023).
19. Seelwal, P. et al. A systematic review of deep learning applications for rice disease diagnosis: current trends and future directions. *Frontiers in Computer Science* **6**, 1452961 (2024).
20. Pacal, I. et al. A systematic review of deep learning techniques for plant diseases. *Artificial Intelligence Review* **57**, 304 (2024).
21. Pacal, I. & Işık, G. Utilizing convolutional neural networks and vision transformers for precise corn leaf disease identification. *Neural Computing and Applications* **37**, 2479–2496 (2025).
22. Pal, A. & Kumar, V. Agridet: Plant leaf disease severity classification using agriculture detection framework. *Engineering Applications of Artificial Intelligence* **119**, 105754 (2023).
23. Kumar, S., Pal, S., Singh, V. P. & Jaiswal, P. Performance evaluation of resnet model for classification of tomato plant disease. *Epidemiologic Methods* **12**, 20210044. https://doi.org/10.1515/em-2021-0044 (2023).

24. Saxena, O., Agrawal, S. & Silakari, S. Disease detection in plant leaves using deep learning models: Alexnet and googlenet. 1–6, https://doi.org/10.1109/TRIBES52498.2021.9751620 (publisherIEEE, 2021).
25. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **25**, https://doi.org/10.1145/3065386 (2012).
26. Chollet, F. Xception: Deep learning with depthwise separable convolutions. 1251–1258 (2017).
27. Howard, A. G. *et al.* Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint*arXiv:1704.04861 https://doi.org/10.48550/arXiv.1704.04861 (2017).
28. Ding, X., Zhang, X., Han, J. & Ding, G. Scaling up your kernels to 31x31: Revisiting large kernel design in cnns. 11963–11975 (2022).
29. Liu, Q., He, P., Zhao, Z. & Li, C. Cotton field pest recognition based on gradient descent mechanism and large convolutional kernel network. In *2024 IEEE 4th International Conference on Power, Electronics and Computer Applications (ICPECA)*, 515–520 (IEEE, 2024).
30. Dong, X. *et al.* Cswin transformer: A general vision transformer backbone with cross-shaped windows. 12124–12134 (2022).
31. Vaswani, A. *et al.* Attention is all you need. *Advances in neural information processing systems* **30** (2017).
32. Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. 618–626 (2017).

Author contributions

G.P. and W.W. wrote the main manuscript text, B.Z. and Z.L. completed the experimental part, G.P. and Z.L. collected data, and the pictures were uniformly processed according to the experimental requirements. X.Q. finished making the pictures in the article.

Funding

The present study was funded by the Major Science and Technology Special Programs in Yunnan Province, 202302 AE090020, and the Basic Research Special Program in Yunnan Province, 202401 AS070006.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to W.W.