RESEARCH ARTICLE

Identification of Maize Diseases Based on Dynamic Convolution and Tri-Attention Mechanism

FEILONG TANG [1,2], **ROSALYN R. PORLE** [1], (Member, IEEE),
HOE TUNG YEW [1], (Member, IEEE), AND **FARRAH WONG** [1], (Senior Member, IEEE)
[1]Faculty of Engineering, Universiti Malaysia Sabah, Kota Kinabalu, Sabah 88400, Malaysia
[2]College of Mechanical and Electrical Engineering, Xichang University, Xichang 615013, China

Corresponding author: Rosalyn R. Porle (rlyn39@ums.edu.my)

ABSTRACT Accurate, non-destructive classification of maize diseases is crucial for efficiently managing agricultural losses. While existing methods perform well in controlled environment dataset like PlantVillage, their accuracy often declines in real-world scenarios. In this work, ResNet50 is enhanced by integrating a dynamic convolution module and triplet attention modules. This method adaptively recalibrates the convolution kernel weights, establishing dependencies across spatial and channel dimensions through tensor rotation and residual transformations. The proposed method surpasses state-of-the-art alternatives, reaching 98.79% validation accuracy on the PlantVillage maize dataset and 97.47% on the Corn Leaf Disease Dataset through cross-validation. Even with complex backgrounds, it attains an average accuracy of 88.33% for classifying six types of maize diseases. Experimental results confirm its effectiveness in maize disease detection.

INDEX TERMS Attention mechanism, dynamic convolution, fine-grained visual classification, maize leaf disease, residual network.

I. INTRODUCTION

Maize contributes significantly to global food security and is widely cultivated across the Americas, Africa, and Asia. The intensification of maize cultivation requires rapid, accurate, and non-destructive methods for classifying maize diseases, which is crucial for effective loss management. Deep learning has proven to be an effective method for plant disease classification due to its strong ability to extract features and recognize patterns [1], [2].

However, most existing models rely on samples collected under controlled laboratory conditions. Unlike laboratory environments, field conditions introduce complexities such as varied backgrounds, inconsistent lighting, and diverse feature stages. Consequently, these models must recognize the variable disease symptoms of size, color, and shape and capture the correlations between spatial and channel features. In the field, two main challenges arise in identifying

The associate editor coordinating the review of this manuscript and approving it for publication was Jeon Gwanggil.

maize leaf diseases: 1) Disease symptoms on maize leaves vary widely in size, shape, and severity, making accurate classification by models challenging. 2) The features of diseased maize leaves are often influenced by varying backgrounds, which introduce visual noise and make it more difficult for models to identify key disease features, reducing their accuracy.

To address these challenges, models must effectively capture the correlations among spatial and channel features. Researchers have proposed attention mechanisms, such as the Convolutional Block Attention Module (CBAM) and the Squeeze-and-Excitation (SE) module, to enhance the ability to capture dependencies in spatial and channel features. Additionally, the increase in the model's width or depth has been proven to enable the recognition of more complex features. However, handling cross-dimensional interactions remains challenging, and increasing model width and depth often results in higher computational complexity.

This study presents a novel ResNet50-based model variant designed to improve the accuracy of classifying maize

diseases from field images. The proposed model improves the representation and generalization ability of the original ResNet50 by integrating dynamic convolution and triplet attention modules. Specifically, dynamic convolution [3] provides a superior non-linear representation by adapting the kernel weights for each image without expanding the network's architecture. Additionally, the triplet attention module [4] strengthens the detection of cross-dimensional relationships by employing rotational and averaging strategies to combine spatial attention, thus reinforcing the interaction between channel and spatial dimensions. As a result, the enhanced model performs robustly in detecting complex disease features in real-world backgrounds, even with limited data availability.

1) A new structure is presented by enhancing the ResNet50 model, which is used to identify healthy and five-class maize diseases. This model achieves higher classification accuracy than all other compared models.

2) To enhance the representational abilities of the model, dynamic convolution modules have been integrated within the residual connections of the ResNet50 model. This adaptation allows for the convolutional kernel weights to be fine-tuned responsively to the image content, thereby enhancing model performance without expanding the network's depth or width.

3) The DY-Tri-ResNet50 model also features a triplet attention module, which increases classification accuracy by efficiently capturing cross-dimensional interaction data.

4) The proposed model was tested on both public and custom datasets. The results demonstrate that the model performs consistently well across a range of background conditions, outperforming state-of-the-art models in accuracy.

II. RELATED WORK

Deep learning is widely used in plant disease classification, with models such as Inception [5], ResNet [6], and DenseNet [7]. These methods empower farmers to make timely decisions during the early stages of disease development through effective classification and pattern recognition. These models generally adopt a two-module architecture, comprising a feature extractor and a classification head, as shown in [8] and [9]. Zhao et al. [10] employed the ResNet50 architecture as the base model and proposed an image-based Plant Disease Severity Estimation Network, achieving an average accuracy of 98% across all classes in the PlantVillage dataset. Waheed et al. [11] conducted experiments on selected grape diseases from the PlantVillage dataset using AlexNet, MobileNet, and ShuffleNet. Their results showed that AlexNet achieved the highest accuracy of 99.01%, while MobileNet, which has fewer parameters and is easier to train, achieved an accuracy of 95.34%. Wang et al. [12] proposed a convolutional neural network (CNN) that integrates Inception and residual structures, along with an enhanced Convolutional Block Attention Module (CBAM). Experimental results showed classification accuracies of 97.96%, 99.43%, and 95.20% for maize, potato, and tomato

diseases, respectively, in the PlantVillage dataset. Fan et al. [13] introduced an optimized DenseNet architecture for classifying maize leaf diseases. Their study reported an average accuracy of 98.06% for identifying three maize leaf disease categories and one healthy category in the PlantVillage dataset.

In practice, the performance of these models may significantly decrease when applied to field environments. For instance, Xiong et al. [14] reported that an enhanced ResNet101's classification accuracy decreased from 98.91% to 75.06% when applied outside laboratory conditions. The decline in accuracy can be attributed to environmental factors that obscure the disease symptoms in images. Additionally, plant diseases exhibit both large-scale features and fine-grained which complicate the detection process [15]. Attention mechanisms can help the network to focus on disease-relevant features while suppressing irrelevant ones, thereby enhancing its robustness and classification accuracy in challenging environmental conditions.

Qi et al. [16] extended the public dataset PlantVillage by adding real-world disease samples from the internet and actual agricultural sites. The authors proposed an improved MobileNet network. Experimental results indicated that the model achieved an accuracy of over 80% on the PlantVillage dataset mixed with field samples. Masood et al. [17] proposed an improved version of the EfficientNet architecture for classifying cucumber leaves into four categories using field-collected data. The modified network achieved a mean average Precision (mAP) of 85.52% on the test set. Feng et al. [18] proposed an optimized ResNet architecture, SE-ResNet, and evaluated it on a late blight potato dataset containing 15,360 images. The proposed architecture achieved an accuracy of 73.9% on the test dataset.

Recent studies have focused on enhancing the ResNet architecture for plant disease classification, successfully mitigating the vanishing and exploding gradient problems commonly encountered in deep neural networks. Notably, models such as MaizeNet [19], ResNet-CBAM [20], and PDICNet [21] mainly focus on modifying the fundamental components of ResNet to customize their designs for specific classification tasks. Incorporating attention mechanisms has proven advantageous for ResNet in terms of efficiency and performance, as evidenced by studies on CS-ResNet [22], FA-ResNet [23], and ResNet-Dropped [24].

III. MATERIALS AND METHOD
A. BASE FRAMEWORK
In this work, an enhanced ResNet50 model is proposed for the identification of maize leaf diseases. A key feature of the ResNet [6] architecture lies in its "residual blocks." Each of these modules consists of at least two convolutional layers and includes skip connections that enable the input to bypass intermediate layers and flow directly to the output, creating an effective "shortcut." This innovative design not only improves feature reuse but also mitigates the vanishing gradient problem. ResNet50, which consists of 50 layers,
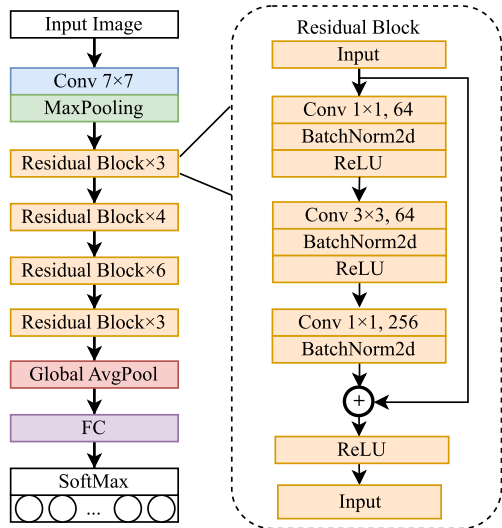
FIGURE 1. The network architectures for ResNet50.

FIGURE 2. Dynamic convolutional layer.

enables more effective gradient flow through the network, a crucial factor for training deeper networks efficiently. Recent studies have demonstrated that ResNet50 and its variants achieve exceptional performance in a variety of plant disease classification tasks. A detailed visual representation of the ResNet50 architecture is provided in Fig 1.

B. DYNAMIC CONVOLUTIONAL

In traditional Convolutional Neural Networks, the convolution operation is commonly defined as $y = g\left(W^T x + b\right)$. W represents the convolutional weight matrix, b represents the convolutional bias vector, and represents the activation function. In such networks, the weights W and biases b of the convolutional kernels are learned during the training process and remain constant throughout the forward and backward propagation of the network. This means that the convolutional kernel's weights and biases are static and do not change in response to variations in the input data.

Dynamic convolution, also called to as conditional convolution. In this type of convolution, the weights of the convolutional kernel can dynamically change based on input data. An example of these methods is Dynamic Convolutional Networks [8]. Unlike static convolutional kernels, dynamic convolution adapts its convolutional kernel by dynamically altering both the weights $\hat{W}_k(x)$ and biases $\hat{b}_k(x)$ according to variations in each input x. The overall process of dynamic convolution is illustrated, as in (1).

$$y = g\left(\hat{W}^T(x)x + \hat{b}(x)\right)$$

$$\hat{W}_k(x) = \sum_{i=1}^{k} A_i W_i, \quad \hat{b}_k(x) = \sum_{k=1}^{k} A_i \hat{b}_i$$

$$\sum_{1}^{i=k} A_i = 1 \quad A_i \in [0, 1] \tag{1}$$

To simplify calculations, the impact of the bias term is neglected in this paper. Therefore, the simplified dynamic
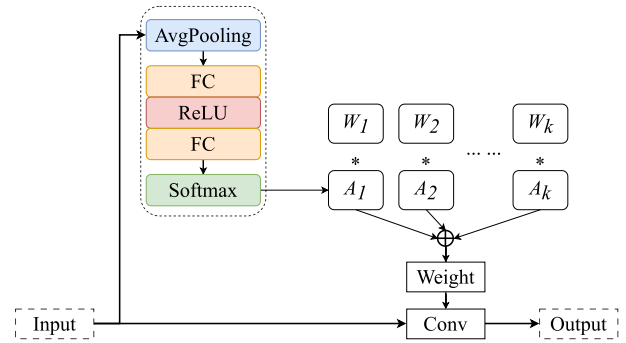
convolution is expressed, as in (2).

$$y = g\left(\hat{W}^T(x)x\right) \tag{2}$$

Here, the convolutional kernel weights for dynamic convolution are defined by aggregating multiple (k) linear functions, as in (3).

$$\hat{W}_k(x) = \sum_{i=1}^{k} A_i W_i \left(\sum_{1}^{i=k} A_i = 1, A_i \in [0, 1]\right) \tag{3}$$

The aggregation of dynamic convolutional kernels represents an effective strategy to enhance the network's expressive capability while maintaining control over the model size. This approach does not introduce additional parameters by increasing the depth or width of the network. Instead, it enhances the model's representational power without significantly adding to the computational burden, by employing parallel kernel aggregation and shared output channels. The structure is illustrated in Fig 2.

C. TRIPLET ATTENTION

Triplet attention mechanism consists of three simultaneous branches, two of which are responsible for understanding interactions across different dimensions between the channel dimension, C, and any other spatial dimension. The remaining branch, similar to CBAM [23], is used for creating focus on spatial features. Fig 3 illustrates the structure.

Given an input tensor $\chi \in R^{(C \times H \times W)}$, which is fed into three branches referred to as $\chi = \chi_1^{in} = \chi_2^{in} = \chi_3^{in}$, the first branch aims to create a relationship between the height dimension H, and the channel dimension C. To this end, χ_1^{in} is initially rotated 90° counterclockwise along the H axis to yield $\chi_1' \in R^{(W \times H \times C)}$. χ_1' is then dimensionally reduced through Z-pool to obtain $\chi_1'' \in R^{(2 \times H \times C)}$. Subsequently, χ_1'' is passed through a standard convolutional layer with a kernel size of $k \times k$, followed by a Batch Normalization (BN) layer, resulting in $\chi_1''' \in R^{(1 \times H \times C)}$. Tensor χ_1''' is then transmitted through a sigmoid activation layer, $\sigma\left(\chi_1'''\right)$, to produce attention weights for tensor χ_1', producing $\chi_1^* \in R^{(W \times H \times C)}$. Finally, χ_1^* is rotated 90° clockwise along the H axis to achieve $\chi_1^{out} \in R^{(C \times H \times W)}$.

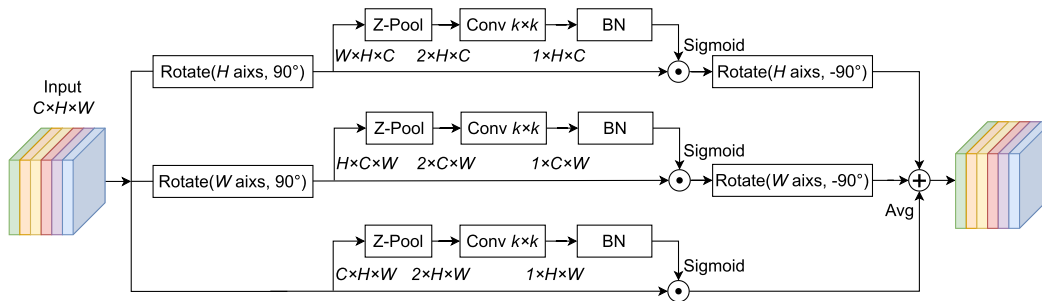In the second branch, the objective is to build interaction between the width dimension W, and the channel dimension

FIGURE 3. Illustration of the triplet attention which has three branches.

C. χ_2^{in} is rotated 90°counterclockwise along the W axis to obtain $\chi_2' \in R^{(H \times C \times W)}$. χ_2' is then reduced in dimension via Z-pool to get $\chi_2'' \in R^{(2 \times C \times W)}$. Afterward, χ_2'' is passed through a standard convolutional layer with a kernel size of $k \times k$, followed by a BN layer, to obtain $\chi_2''' \in R^{(1 \times C \times W)}$. Tensor χ_2''' is subsequently sent through a sigmoid activation layer $\sigma\left(\chi_2'''\right)$, to produce attention weights for tensor χ_2', resulting in $\chi_2^* \in R^{(H \times C \times W)}$. Lastly, χ_2^* is rotated 90°clockwise along the W axis to derive $\chi_2^{out} \in R^{(C \times H \times W)}$.

In the third branch, χ_3^{in} is reduced in dimension by Z-pool to obtain $\chi_3'' \in R^{(2 \times H \times W)}$. Then χ_3'' is sent through a standard convolutional layer with a kernel size of $k \times k$, followed by a BN layer, resulting in $\chi_3''' \in R^{(1 \times H \times W)}$. Next, tensor χ_3''' is passed through a sigmoid activation layer $\sigma\left(\chi_3'''\right)$, to generate attention weights for tensor χ_3', leading to $\chi_3^{out} \in R^{(C \times H \times W)}$.

The Z-pool process is articulated in (4):

$$Z - \text{pool}(\chi) = \left[\text{MaxPool}_{0d}(\chi), \text{AvgPool}_{0d}(\chi)\right] \quad (4)$$

D. PROPOSED DY-TRI-RESNET50

ResNet50 is a widely recognized and powerful architecture in plant disease classification tasks, known for its outstanding performance in classification. Its modular design allows for flexible modifications, making it a solid foundation for further enhancements. Furthermore, Chen et al. [24] demonstrated the efficacy of dynamic convolutional modules in enhancing MobileNetV2, MobileNetV3, and ResNet-18 models, showcasing their ability to aggregate information in a non-linear fashion for improved image recognition. Jiang et al. [25] further leveraged dynamic convolution modules and self-attention mechanisms to refine BERT and its derivatives, achieving comparable performance metrics while significantly reducing model complexity and training expenses. Huang et al. [26] introduced an enhanced SD-CNN model incorporating dynamic convolutions for brain disorder identification. This modification resulted in a 3% and 2% improvement in classification accuracy on two real-world datasets compared to models without dynamic convolutions. Huang et al. [26] optimized the DGLNet model by incorporating dynamic convolutions, achieving a classification accuracy of 99.62% on the PlantVillage dataset, representing a 0.6% improvement over the previous model. Building on these advancements, dynamic convolutions are integrated into the ResNet50 architecture to enhance the recognition of complex features, without increasing its depth or width. By replacing static convolutions with dynamic ones, the model can adaptively adjust the convolutional weights for each input sample, thereby improving its ability to capture intricate patterns. This innovation enables the model to more effectively handle a diverse range of features, thereby improving the accuracy of disease classification in field conditions.

Additionally, the proposed improved model introduces a triplet attention module after each stage in ResNet50. This module is specifically designed to capture complex dependencies in the spatial and channel dimensions, aiding the model in better distinguishing between different types of plant diseases. This enhancement leads to an overall improvement in classification accuracy. To avoid increasing the depth of the model, the number of residual blocks in ResNet50 has been reduced. Specifically, in this work, the original ResNet-50 architecture, which consists of residual blocks in a 3, 4, 6, 4 configuration, was modified to a lighter version with a 3, 4, 4, 3 structure. This adjustment reduces the number of residual blocks in each stage while maintaining the overall hierarchical design of the network. Such a modification aims to balance model complexity. Key improvements have been introduced to the residual blocks of ResNet50. Specifically, static convolutions are replaced with dynamic convolutions, allowing the convolutional kernels to adapt to the input data. This incorporation of non-linear information enhances the model's feature representation capabilities. By addressing the complex characteristics of maize diseases in real-world conditions, the model becomes more adaptable to diverse features, leading to improved classification accuracy. The framework of the proposed Dy-Tri-ResNet50 model is illustrated in Fig 4.

In summary, the DY-Tri-ResNet50 model is designed for maize disease classification. By integrating dynamic convolution and triplet attention mechanisms into the residual blocks, the model enhances its generalization capabilities, facilitating more accurate identification of maize diseases, particularly in real-field conditions. This approach allows the model to capture local features while preserving sensitivity to global information, making it highly effective for precise plant disease diagnosis in agricultural applications.
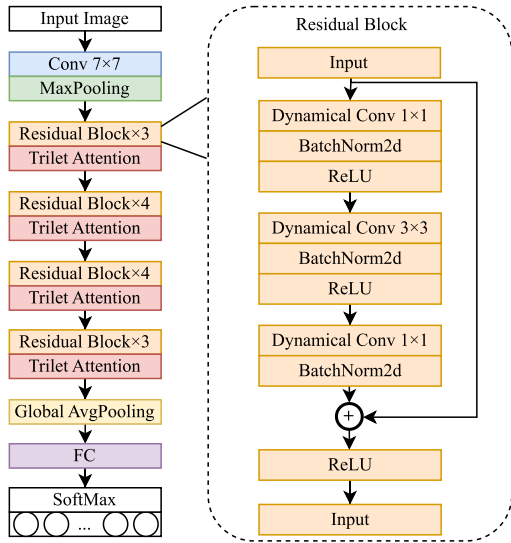
FIGURE 4. The proposed Dy-Tri ResNet50 architecture.

E. MODEL EVALUATION INDEXES

Numerous metrics can be utilized to assess the performance of models. In this study, Precision (P), Recall, and F1 Score were selected to evaluate the models for Maize disease and pest identification. Precision (P) refers to the percentage of correctly predicted samples relative to all samples in the model, as shown in (5):

$$P(\%) = \frac{TP}{TP+FP} \times 100 \tag{5}$$

Here TP (True Positive) represents the number of samples correctly classified as positive, TN (True Negative) as negative, FP (False Positive) as incorrectly classified as positive, and FN (False Negative) as incorrectly classified as negative. Recall measures the proportion of positive samples that are correctly classified. The formula for calculating Recall is shown in (6):

$$Recall(\%) = \frac{TP}{TP+FN} \times 100 \tag{6}$$

In certain contexts, such as medical diagnosis or anomaly detection, Recall is more critical than Accuracy because the primary concern is to accurately identify the positive samples rather than just the proportion of correct classifications of the model. F1 Score is the harmonic mean of Precision and Recall. It takes into account both the accuracy and recall of the model. The formula for calculating the F1 Score is shown in (7):

$$F1(\%) = \frac{2TP}{2TP+FP+FN} \times 100 \tag{7}$$

IV. RESULTS AND ANALYSIS

A. TEST PLATFORM AND PARAMETER SETTING

The experimental platform used in this study is composed of the following hardware and software environments: the hardware includes an Intel Core i7-12700H processor and an NVIDIA RTX A2000 8G Laptop GPU, while the software

environment is based on the WIN11 operating system, Python 3.7, and the PyTorch deep learning framework.

B. DATASET

In this section, three datasets are used for the experiments: PlantVillage, maize Leaf Disease, and a self-constructed dataset. PlantVillage [28] is a publicly available database widely recognized as a benchmark for testing algorithms in plant disease classification. The maize Leaf Disease dataset [29], provided by Sandi Indika Saputra (Kaggle, 2023), contains images of maize leaves categorized into four classes. Additionally, a self-constructed dataset, characterized by challenges such as data imbalance and insufficient samples, is included. These datasets collectively encompass six categories of maize leaf diseases. Among them, Cercospora Leaf Spot (CLS), Northern Leaf Blight (NLB), and Common Rust (CR) are the most prevalent. NLB, caused by the pathogen responsible for maize leaf blight, presents as yellow to brown spots on the leaves that gradually enlarge, potentially leading to leaf death. CLS, caused by the fungus Cercospora zeaemaydis, initially appears as round to oval pale yellow or brown spots, which darken to purple or brown as the disease progresses. CR, caused by the fungus Puccinia sorghi, results in small, circular rust-colored or dark brown spots on maize leaves, which may cause leaf fall in severe cases. In addition, Phaeosphaeria Leaf Spot (PLS), a relatively uncommon fungal disease in maize, causes round to oval lesions on maize leaves, often surrounded by distinct yellow halos as the disease progresses. Zinc Deficiency (ZD), a common nutrient deficiency in maize leaves, results in interveinal chlorosis on maize leaves, particularly at the tips and margins.

C. EXPERIMENTS ON THE PUBLIC DATASET

1) EXPERIMENTS ON THE MAIZE DATASET FROM THE PLANTVILLAGE

To evaluate the performance of the proposed method, we extracted maize-related data from the PlantVillage dataset, forming a maize dataset with four categories: CLS, NLB, CR, and healthy leaves. However, the extracted dataset exhibited an imbalance, with a significantly lower number of CLS images compared to the other categories. To address this issue, standard offline augmentation techniques were employed to balance the dataset.

Specifically, a portion of the original images from all four categories was randomly selected as a validation set for model evaluation. Then, 50 CLS images and 20 NLB images were randomly chosen and augmented using techniques such as flipping, rotation, cropping, random erasing, and resizing to increase the number of images in these categories. Following augmentation, the total maize dataset consisted of 5,796 images, including 1,349 CLS images, 1,421 CR images, 1,528 NLB images, and 1,498 healthy leaf images. All images were resized to a uniform dimension of 256 × 256 pixels. A detailed overview of the dataset is provided in Table 1.

TABLE 1. Detailed dataset the maize dataset from the PlantVillage.

Classes	Original counts	Augmented counts	Train counts	Test counts
CLS	513	1349	997	352
NLB	985	1421	1069	352
CR	1192	1528	1176	352
Healthy	1162	1498	1146	352

Model training and validation were conducted on the Maize dataset using the proposed Dy-Tri-ResNet50 architecture. Additionally, to further verify the effectiveness of the proposed method, we conducted comparative experiments with six influential CNN architectures, such as DenseNet [5], InceptionV3 [3], Xception [30], InceptionResNetV2 [31], MobileNetV3 [32], and ResNet [4]. Moreover, The MaizeNet model, an enhanced version of ResNet50 proposed by Masood et al. [17], was also included for comparison.

To optimize performance based on the experimental hardware, a batch size of 8 was used. The model was trained for 20 epochs, with the best-performing model being saved automatically. The cross-entropy loss function was employed, and a learning rate of 0.001, commonly used for small to medium-sized datasets, was applied. The training and validation loss curves are shown in Fig 5.

Figures 5(f) and 5(h) show that the proposed model and InceptionResNetV2 demonstrate the most stable and consistent performance, with closely aligned training and validation loss curves and minimal signs of overfitting. This demonstrates the proposed model's strong ability to capture the nonlinear features of plant diseases, validating its effectiveness in addressing complex plant disease classification tasks. The model is particularly effective at identifying the intricate features of maize disease images, which often have subtle inter-class differences and intra-class similarities. Similarly, InceptionV3 and MaizeNet showed relatively stable validation loss trends and balanced training progress.

In contrast, the ResNet50, DenseNet121, and Xception models demonstrated clear signs of overfitting, as indicated by the divergence between their training and validation loss curves after 10 epochs. While training losses continued to decrease, validation losses fluctuated or increased, indicating strong performance on training data but poor generalization to unseen validation data. This overfitting can be attributed to the complex nature of maize disease images, which feature overlapping characteristics across different disease categories and considerable variability within the same category. ResNet50 and DenseNet121 appeared to focus on memorizing specific details of the training data, hindering their ability to differentiate subtle inter-class variations and maintain consistency within classes. Although Xception initially performed well, the instability in its validation loss suggests difficulty in learning the fine-grained texture and structural features of disease images, likely due to its reliance on deep separable convolutions, which are highly sensitive to such details.

TABLE 2. The results of different model training and testing on the public dataset.

Model	Training Accuracy (%)	Training Loss	Testing Accuracy (%)	Testing Loss
MobileNetV3	**99.82**	0.0053	93.11	0.2685
Xception	98.53	0.0451	92.68	0.4103
ResNet50	98.57	0.0483	95.53	0.1465
DenseNet121	97.43	0.0746	97.87	0.0658
InceptionResNetV2	99.45	0.0286	97.8	0.0607
InceptionV3	98.71	0.0485	98.58	0.0538
MaizeNet [17]	97.52	0.0739	96.52	0.0891
RIC-Net [10]	98.43	0.0432	97.96	0.0675
Proposed Model	98.85	0.0414	**98.79**	0.0468

Additionally, a comparison was made with methods reported for maize disease classification on the PlantVillage dataset. Liu et al. [15] proposed the RIC-Net network, which utilizes three simplified RI-Block modules, achieving a classification accuracy of 97.69% on the maize disease subset of the PlantVillage dataset. Based on the experimental results obtained from the extracted maize dataset and the reported performance of the RIC-Net network, the proposed model demonstrates superior test accuracy compared to all previously reported models. The test accuracies of different methods are summarized in Table 2.

The results presented in Table 2 indicate that the proposed model achieved a training accuracy of 98.85% and a testing accuracy of 98.79%, with both values closely aligned. It outperformed all other compared models on the unseen test set, achieving the highest accuracy of 98.79%. This demonstrates the model's strong generalization capability. The minimal gap between training and testing accuracy further emphasizes its stability and ability to adapt effectively to unseen data.

2) CROSS-VALIDATION ON THE CORN LEAF DISEASE DATASET

In this subsection, K-fold cross-validation is employed to evaluate the performance of these models. This method effectively mitigates the issue of model overfitting while providing a more reliable performance assessment for the proposed model. The same cross-validation folds and training-validation splits are applied consistently across all models in this study.

The Corn Leaf Disease dataset, used for evaluation, contains images of healthy leaves and three common disease categories: CLS, NLB, and CR, with 1,000 samples per class, totaling 4,000 images. All images were resized to 256×256 pixels and normalized to ensure consistency in the model input. Data augmentation techniques, including random rotation, flipping, and cropping, were applied during training to enhance the model's generalizability.

Commonly used values for K in cross-validation are 5 or 10. Considering the dataset size and hardware limitations, the data was divided into five equally sized folds. Each fold was sequentially used as the validation set (20% of the data), with the remaining 80% reserved for training. For each fold,
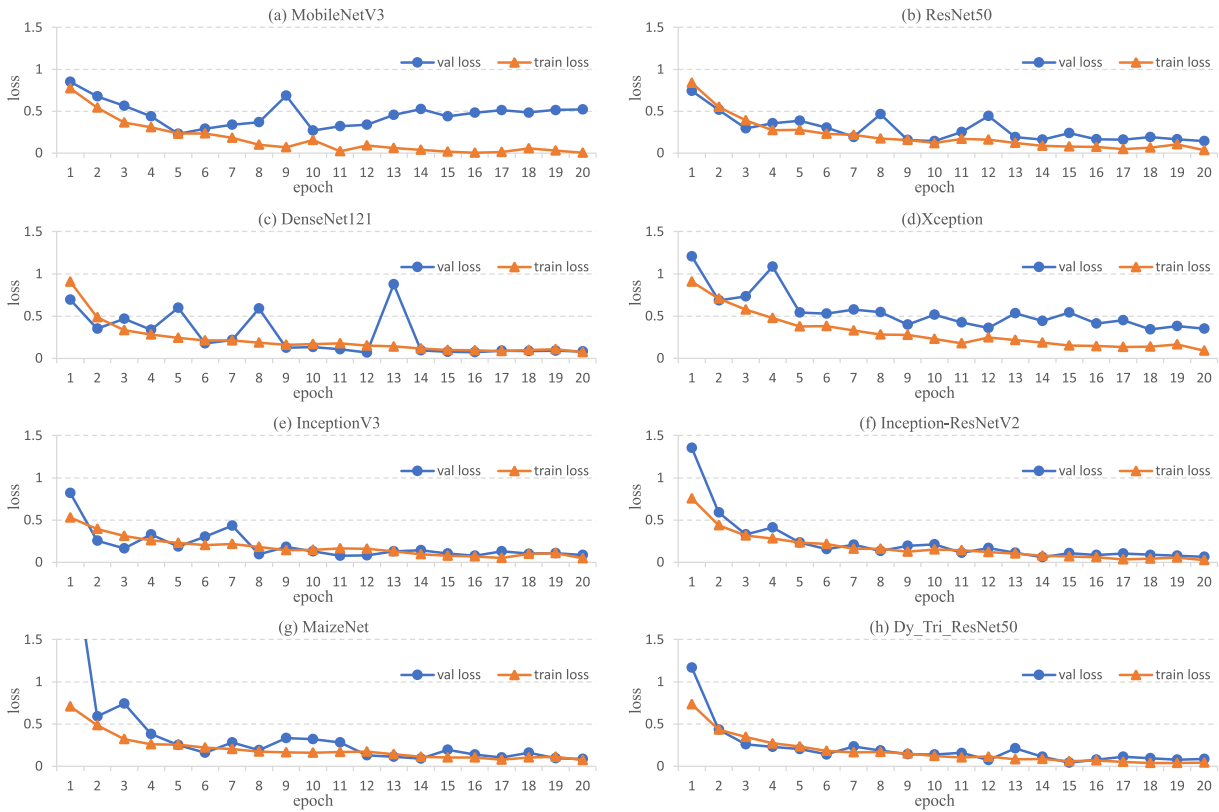
FIGURE 5. Training and validation loss curves of 8 trained models.

the model was trained for 20 epochs with a learning rate of 0.001, and a weight decay of 0.0001 was applied to mitigate overfitting. A cosine annealing learning rate scheduler was used to dynamically adjust the learning rate during training. The model's performance was evaluated after each epoch using the corresponding validation fold. The cross-validation results were averaged across the five folds to provide a final assessment of the model's performance, ensuring consistency across different data subsets. The test results for the various methods are summarized in Table 3.

TABLE 3. Results of cross-validation on the Corn Leaf Disease dataset. Here, 'Av' represents the average best accuracy, and all values in the table are rounded to two decimal places. The values are in %.

Model	Fold1	Fold2	Fold3	Fold4	Fold5	Average
MobileNetV3	95.50	96.38	96.38	96.13	97.13	96.30
Xception	94.50	93.00	94.13	93.88	95.00	94.10
ResNet50	97.00	96.63	96.13	96.50	96.25	96.50
DenseNet121	96.38	96.25	96.00	96.25	96.88	96.35
InceptionResNetV2	96.25	97.13	96.88	97.13	97.50	96.98
InceptionV3	96.88	97.25	97.00	97.63	97.13	97.17
MaizeNet[17]	96.13	96.88	96.63	97.25	96.63	96.70
Proposed model	97.25	97.50	97.25	97.38	98.00	**97.47**

As shown in Table 3, the proposed model outperformed all other models, achieving the highest average accuracy of 97.47% across five folds. This demonstrates the model's strong generalization ability and robustness across different

data splits. Its best performance was in Fold 5, with an accuracy of 98.00%, highlighting its stability and reliability. While InceptionV3, InceptionResNetV2, and MaizeNet showed competitive results with average accuracies of 97.17%, 96.98%, and 96.70%, respectively, their performance was less consistent. ResNet50, MobileNetV3, and DenseNet121 achieved moderate accuracy but fell short of the proposed model. The lowest average accuracy was recorded for Xception at 94.10%, reflecting its limitations in handling the complexities of the Corn Leaf Disease dataset. These results confirm that Dy-Tri-ResNet50 is highly effective in capturing essential features and managing the variability of the dataset, making it the most robust and reliable model for maize disease classification.

D. EXPERIMENTS ON THE COLLECTED DATASET
The effectiveness of the proposed model on real-world field samples was evaluated using data samples for five types of maize diseases, which were collected from online sources. These include 78 samples for NLB, 77 for CLS, 144 for PLS, 63 for CR, and 166 for ZD. Additionally, 149 images of healthy maize leaves were gathered to form the collected dataset. The combined dataset contains images of NLB, CLS, and CR from both laboratory and field environments, whereas the samples for the PLS, ZD, and healthy categories are exclusively derived from field images. This dataset highlights

three common challenges encountered in field applications of plant disease classification tasks: (1) data imbalance, (2) insufficient sample numbers per class, and (3) complex backgrounds and lighting conditions. Examples of these categories are illustrated in Fig 6.
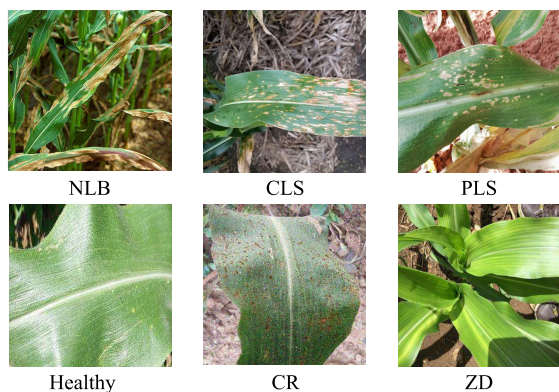

FIGURE 6. Images of maize leaf diseases.

Given the limitations of the dataset, this study deviates from the commonly used 80% training and 20% testing split. Instead, 50 raw images were randomly selected from each test category to form the validation set for model evaluation. This balanced test set approach is commonly used in disease classification tasks [33], [34]. Offers several advantages. First, it minimizes bias by ensuring each category contributes equally to performance evaluation, thus reducing the influence of categories with larger sample sizes. Second, it allows for a fair comparison by maintaining consistent representation across categories, leading to an unbiased assessment of the model's performance. Table 4 shows the distribution of maize disease types and their respective sample sizes in both the training and test sets.

TABLE 4. Detailed information of the self-built dataset. Here, the data in parentheses represents the number of samples with real field backgrounds.

Classes	Total counts	Training counts	Testing counts
NLB	222 (78)	172 (60)	50 (18)
CLS	219 (77)	169 (62)	50 (15)
PLS	161 (161)	101 (101)	50 (50)
Healthy	149 (149)	99 (99)	50 (50)
CR	170 (63)	120 (45)	50 (18)
ZD	166 (166)	116 (116)	50 (50)

Data augmentation can be classified into online and offline methods. Online augmentation, which modifies data during input to the model without increasing the original dataset size, is typically used for large datasets. In contrast, offline augmentation, which is more appropriate for smaller datasets, increases the data volume by applying various geometric transformations. In this study, offline augmentation techniques were employed to address the limitations of the sample size. These techniques included random brightness adjustments to simulate varying lighting conditions, thereby

improving the model's robustness to lighting variations. Random erasing was also applied to reduce reliance on specific image details, enhancing the model's ability to handle occlusions and missing segments. Furthermore, image flipping was used to increase the diversity of the dataset. These augmentation methods expanded the dataset size. The experimental parameters in this section were consistent with those used in Section III-C, and all methods, including the proposed model and comparison models, were tested under the same conditions. The results are summarized in Table 5.

TABLE 5. Comparison of the results of different models on the collected dataset.

Model	Precision (%)	Recall (%)	F1 (%)
MobileNetV3	77.33	75.33	75.18
Xception	72.76	71.67	71.42
ResNet50	84.69	83.67	83.85
DenseNet121	87.49	86.33	86.43
InceptionResNetV2	87.86	87.67	87.67
InceptionV3	88.53	88.00	88.09
MaizeNet[17]	87.75	87.67	87.67
Proposed model	**88.63**	**88.33**	**88.28**

The superior performance of the proposed model, achieving 88.63% Precision, 88.33% Recall, and 88.28% F1-score, can be attributed to its strong generalization, enhanced by dynamic convolution and tri-attention mechanisms. Dynamic convolution enables adaptive focusing on key spatial features, while tri-attention effectively captures spatial and channel dependencies, addressing subtle inter-class differences and complex backgrounds. Furthermore, data augmentation techniques, including random brightness adjustments, random erasing, and image flipping, enhance the model's robustness to lighting variations and occlusions. The higher Recall, compared to InceptionV3, reflects improved sensitivity to true positives, which is essential for accurate disease detection. In contrast, models like ResNet50, DenseNet121, and Xception face challenges in handling fine-grained details and complex backgrounds. Overall, the proposed model outperforms others in real-world maize disease classification due to its advanced architectural components and the effective use of data augmentation strategies.
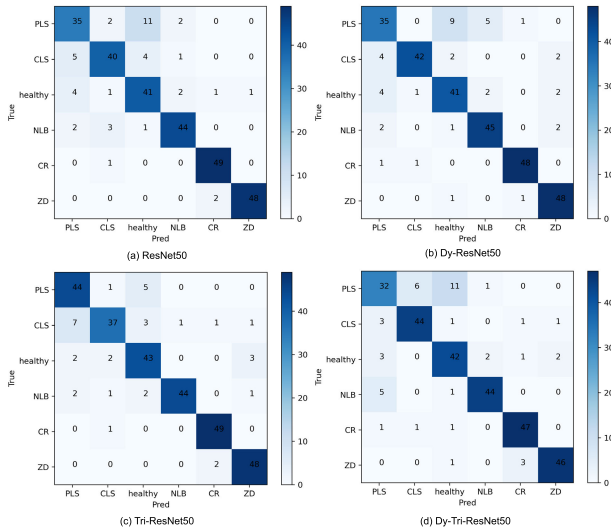
E. ABLATION STUDY
To better assess the contribution of each component in the proposed model, we progressively incorporated the triplet attention mechanism and Dynamic Convolution into the original ResNet50 architecture. As described in Section III-D, we reduced the number of residual blocks in ResNet50 when incorporating the triplet attention layers to avoid increasing the model's depth. This ensures that the observed performance improvements are not attributed to changes in the model's depth or width. The detailed experimental results are presented in Table 6.

The confusion matrix is an essential tool for evaluating the performance of classification models, as it provides insights into the model's ability to correctly classify different

TABLE 6. Table 1: Detailed configurations of different model variants. Here, Dy represents dynamic convolution, and Tri represents the triplet attention mechanism.

Model	Precision (%)	Recall (%)	F1 (%)
ResNet50	85.09	85.00	84.92
Dy + ResNet50	85.85	85.67	85.66
Tri + ResNet50	87.49	86.33	86.43
Proposed model	**88.63**	**88.33**	**88.28**

FIGURE 7. Confusion matrices of 4 models.

FIGURE 8. Accuracy of each class.

The Dy-Tri-ResNet50 model employs dynamic convolution to adaptively adjust its convolutional kernels in response to varying inputs, thereby enhancing its ability to focus on critical areas of the image. This adaptability is crucial for disease detection in complex backgrounds, as it reduces background noise and enables the model to focus on the key features of maize disease images. Moreover, the triplet attention mechanism refines the model's focus by controlling attention allocation across different regions and analyzing relationships between various image features. This mechanism helps the model avoid learning from irrelevant areas, enabling it to allocate more computational resources to extracting disease-relevant features.

Overall, the proposed Dy-Tri-ResNet50 model, which integrates dynamic convolution with multi-dimensional attention, excels in identifying and diagnosing maize disease areas, particularly in complex backgrounds. This combined approach significantly improves recognition precision by reducing errors caused by background noise and enhancing the model's ability to differentiate disease-specific features, thus boosting overall performance compared to models that rely on a single strategy.

categories and identifies areas of strength and error [35]. In the context of maize disease classification, each row of the matrix represents the true class, while each column corresponds to the predicted class. The diagonal values indicate the percentage of correctly classified instances, while the off-diagonal values represent misclassifications. By comparing the confusion matrices of models with and without the Dy and Tri components, we can visually assess the performance of each model on the test dataset, as shown in Fig 7 (a) to (d).

In this ablation study, we assessed the performance of ResNet50 and its variants across six categories. Fig 8 shows the classification accuracy for each category under identical experimental conditions.

The baseline ResNet50 demonstrated strong performance, particularly in CLS, CR, and ZD, with accuracy values of 88%, 94%, and 92%, respectively. The Dy-ResNet50 model improved performance across most classes, achieving 98% in CR and 96% in ZD. The Tri-ResNet50 model showed enhanced results in the Healthy and ZD categories, with accuracy reaching 86% and 96%, respectively. The combined Dy-Tri-ResNet50 approach achieved the highest accuracy in PLS at 88%, along with strong performance in healthy and CR at 86% and 98%, respectively. However, it showed a slight decrease in CLS 74%, indicating that the integration of both dynamic convolution and triplet attention may complicate feature extraction in some cases.

V. CONCLUSION
This work presents a novel network for maize leaf disease classification, built on residual connection blocks and incorporating dynamic convolution and a triplet attention mechanism to enhance the model's ability to capture fine-grained features. This approach facilitates effective cross-dimensional interactions, thereby improving classification accuracy and training stability. Dynamic convolution enables the model to adjust kernel weights for each image, providing a nonlinear representation that outperforms static convolutions. Furthermore, the triplet attention mechanism strengthens the detection of cross-dimensional relationships, improving synergy between channel and spatial dimensions.

To evaluate the model's performance, we used maize images from the PlantVillage dataset. The data was split into training and testing sets, with data augmentation applied. Compared to seven models, including InceptionV3, Xception, ResNet50, and Inception-ResNetV2, our model achieved an accuracy of 98.79%, surpassing the reported

accuracy of 97.96% for RIC-Net under the same experimental conditions. Cross-validation on the Corn Leaf Disease dataset yielded an average best accuracy of 97.47% across five folds. Additionally, testing on a real-world maize dataset with realistic background conditions showed that our model outperformed the comparison models, achieving an accuracy of 88.33%.

In future work, we will focus on three primary objectives: optimizing data collection methods to ensure diverse and representative disease samples, developing the framework for multi-disease collaborative detection to improve overall performance, and enhancing fine-grained disease region localization to precisely identify infected areas. These efforts aim to improve detection accuracy and facilitate the broader application of models in real-world agricultural settings.

REFERENCES

[1] A. Abade, P. A. Ferreira, and F. de Barros Vidal, "Plant diseases recognition on images using convolutional neural networks: A systematic review," *Comput. Electron. Agricult.*, vol. 185, Jun. 2021, Art. no. 106125, doi: 10.1016/j.compag.2021.106125.

[2] T. A. Shaikh, T. Rasool, and F. R. Lone, "Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming," *Comput. Electron. Agricult.*, vol. 198, Jul. 2022, Art. no. 107119, doi: 10.1016/j.compag.2022.107119.

[3] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[5] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2016, *arXiv:1608.06993*.

[6] M. Azadbakht, D. Ashourloo, H. Aghighi, S. Radiom, and A. Alimohammadi, "Wheat leaf rust detection at canopy scale under different LAI levels using machine learning techniques," *Comput. Electron. Agricult.*, vol. 156, pp. 119–128, Jan. 2019, doi: 10.1016/j.compag.2018.11.016.

[7] R. Cui, J. Li, Y. Wang, S. Fang, K. Yu, and Y. Zhao, "Hyperspectral imaging coupled with dual-channel convolutional neural network for early detection of apple valsa canker," *Comput. Electron. Agricult.*, vol. 202, Nov. 2022, Art. no. 107411, doi: 10.1016/j.compag.2022.107411.

[8] Q. Liang, S. Xiang, Y. Hu, G. Coppola, D. Zhang, and W. Sun, "PD2SE-Net: Computer-assisted plant disease diagnosis and severity estimation network," *Comput. Electron. Agricult.*, vol. 157, pp. 518–529, Feb. 2019, doi: 10.1016/j.compag.2019.01.034.

[9] Z. Tang, J. Yang, Z. Li, and F. Qi, "Grape disease image classification based on lightweight convolution neural networks and channelwise attention," *Comput. Electron. Agricult.*, vol. 178, Nov. 2020, Art. no. 105735, doi: 10.1016/j.compag.2020.105735.

[10] Y. Zhao, C. Sun, X. Xu, and J. Chen, "RIC-Net: A plant disease classification model based on the fusion of inception and residual structure and embedded attention mechanism," *Comput. Electron. Agricult.*, vol. 193, Feb. 2022, Art. no. 106644, doi: 10.1016/j.compag.2021.106644.

[11] A. Waheed, M. Goyal, D. Gupta, A. Khanna, A. E. Hassanien, and H. M. Pandey, "An optimized dense convolutional neural network model for disease recognition and classification in corn leaf," *Comput. Electron. Agricult.*, vol. 175, Aug. 2020, Art. no. 105456, doi: 10.1016/j.compag.2020.105456.

[12] D. Wang, J. Wang, Z. Ren, and W. Li, "DHBP: A dual-stream hierarchical bilinear pooling model for plant disease multi-task classification," *Comput. Electron. Agricult.*, vol. 195, Apr. 2022, Art. no. 106788, doi: 10.1016/j.compag.2022.106788.

[13] X. Fan, P. Luo, Y. Mu, R. Zhou, T. Tjahjadi, and Y. Ren, "Leaf image based plant disease identification using transfer learning and feature fusion," *Comput. Electron. Agricult.*, vol. 196, May 2022, Art. no. 106892, doi: 10.1016/j.compag.2022.106892.

[14] Y. Xiong, L. Liang, L. Wang, J. She, and M. Wu, "Identification of cash crop diseases using automatic image segmentation algorithm and deep learning with expanded dataset," *Comput. Electron. Agricult.*, vol. 177, Oct. 2020, Art. no. 105712, doi: 10.1016/j.compag.2020.105712.

[15] C. Liu, H. Zhu, W. Guo, X. Han, C. Chen, and H. Wu, "EFDet: An efficient detection method for cucumber disease under natural complex environments," *Comput. Electron. Agricult.*, vol. 189, Oct. 2021, Art. no. 106378, doi: 10.1016/j.compag.2021.106378.

[16] C. Qi, M. Sandroni, J. C. Westergaard, E. H. R. Sundmark, M. Bagge, E. Alexandersson, and J. Gao, "In-field classification of the asymptomatic biotrophic phase of potato late blight based on deep learning and proximal hyperspectral imaging," *Comput. Electron. Agricult.*, vol. 205, Feb. 2023, Art. no. 107585, doi: 10.1016/j.compag.2022.107585.

[17] M. Masood, M. Nawaz, T. Nazir, A. Javed, R. Alkanhel, H. Elmannai, S. Dhahbi, and S. Bourouis, "MaizeNet: A deep learning approach for effective recognition of maize plant leaf diseases," *IEEE Access*, vol. 11, pp. 52862–52876, 2023, doi: 10.1109/ACCESS.2023.3280260.

[18] S. Feng, D. Zhao, Q. Guan, J. Li, Z. Liu, Z. Jin, G. Li, and T. Xu, "A deep convolutional neural network-based wavelength selection method for spectral characteristics of Rice blast disease," *Comput. Electron. Agricult.*, vol. 199, Aug. 2022, Art. no. 107199, doi: 10.1016/j.compag.2022.107199.

[19] S. R. G. Reddy, G. P. S. Varma, and R. L. Davuluri, "Resnet-based modified red deer optimization with DLCNN classifier for plant disease identification and classification," *Comput. Electr. Eng.*, vol. 105, Jan. 2023, Art. no. 108492.

[20] H. Zhang, L. Jiang, and C. Li, "CS-ResNet: Cost-sensitive residual convolutional neural network for PCB cosmetic defect detection," *Expert Syst. Appl.*, vol. 185, Dec. 2021, Art. no. 115673, doi: 10.1016/j.eswa.2021.115673.

[21] L. Zhan, W. Li, and W. Min, "FA-ResNet: Feature affine residual network for large-scale point cloud segmentation," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 118, Apr. 2023, Art. no. 103259.

[22] M. Arunkumar, A. Mohanarathinam, and K. Subramaniam, "Detection of varicose vein disease using optimized kernel boosted ResNet-dropped long short term memory," *Biomed. Signal Process. Control*, vol. 87, Jan. 2024, Art. no. 105432.

[23] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 11211. Cham, Switzerland: Springer, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.

[24] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, and Z. Liu, "Dynamic convolution: Attention over convolution kernels," 2019, *arXiv:1912.03458*.

[25] Z.-H. Jiang, W. Yu, D. Zhou, Y. Chen, J. Feng, and S. Yan, "ConvBERT: Improving BERT with span-based dynamic convolution," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 12837–12848.

[26] J. Huang, M. Wang, H. Ju, Z. Shi, W. Ding, and D. Zhang, "SD-CNN: A static-dynamic convolutional neural network for functional brain networks," *Med. Image Anal.*, vol. 83, Jan. 2023, Art. no. 102679, doi: 10.1016/j.media.2022.102679.

[27] Y. Yang, G. Jiao, J. Liu, W. Zhao, and J. Zheng, "A lightweight Rice disease identification network based on attention mechanism and dynamic convolution," *Ecolog. Informat.*, vol. 78, Dec. 2023, Art. no. 102320, doi: 10.1016/j.ecoinf.2023.102320.

[28] D. P. Hughes and M. Salathe, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," 2015, *arXiv:1511.08060*.

[29] Sandi Indika Saputra. (2023). *Corn Leaf Disease Dataset*. Kaggle. [Online]. Available: https://www.kaggle.com/datasets/ndisan/corn-leaf-disease

[30] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.

[31] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI*, Feb. 2017, vol. 31, no. 1, pp. 4278–4284, doi: 10.1609/aaai.v31i1.11231.

[32] A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324, doi: 10.1109/ICCV.2019.00140.

[33] P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Deep feature based Rice leaf disease identification using support vector machine," *Comput. Electron. Agricult.*, vol. 175, Aug. 2020, Art. no. 105527, doi: 10.1016/j.compag.2020.105527.

[34] Y. Guo, Y. Lan, and X. Chen, "CST: Convolutional Swin transformer for detecting the degree and types of plant diseases," *Comput. Electron. Agricult.*, vol. 202, Nov. 2022, Art. no. 107407, doi: 10.1016/j.compag.2022.107407.

[35] X. Zhao, K. Li, Y. Li, J. Ma, and L. Zhang, "Identification method of vegetable diseases based on transfer learning and attention mechanism," *Comput. Electron. Agricult.*, vol. 193, Feb. 2022, Art. no. 106703, doi: 10.1016/j.compag.2022.106703.

FEILONG TANG received the B.S. degree in mechanical engineering from Sichuan University of Science and Engineering, Sichuan, China, in 2014, and the M.S. degree in mechanical engineering from Xihua University, Sichuan, in 2017. He is currently pursuing the Ph.D. degree in computer engineering with the University of Sabah, Sabah, Malaysia. Since 2020, he has been a Lecturer in mechanical and electrical engineering wit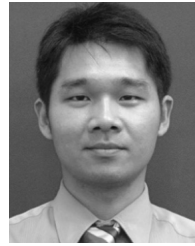h Xichang University. His research interests include the development of image processing techniques, classification of plant disease and pest images, and the design and manufacturing of intelligent harvesting robots.

HOE TUNG YEW (Member, IEEE) received the B.Eng. degree in electrical and electronic engineering from the University of Lincoln, Lincoln, U.K., in 2003, the M.Sc. degree in microelectronic and communications engineering from Northumbria University, Newcastle, U.K., in 2004, and the Ph.D. degree in biomedical engineering from Universiti Teknologi Malaysia, Skudai, Johor, Malaysia. He is currently a Senior Lecturer with the Faculty of Engineering, Universiti Malaysia Sabah. He is a Chartered Engineer (C.Eng.) with the Institute of Engineering and Technology, U.K., and a Professional Engineer with Malaysia Board of Engineers. He has co-authored more than 50 technical peer-reviewed journals and conference papers. His research interests include wireless communications, vertical handover, wireless sensor networks, and the Internet of Things. He served as an Executive Committee Member (Treasurer) for the IEEE Malaysia Sabah Subsection, from 2018 to 2020. He served as a TPC Member for the IEEE Malaysia Sabah Subsection conferences.

ROSALYN R. PORLE (Member, IEEE) received the B.Eng. degree in electronic and telecommunication engineering from Universiti Malaysia Sarawak, in 2001, and the M.Sc. degree in electrical and electronic engineering and the Ph.D. degree in computer engineering from Universiti Malaysia Sabah, in 2005 and 2011, respectively. She is currently a Senior Lecturer with the Faculty of Engineering, Universiti Malaysia Sabah. Her research interests include vision systems, artificial intelligence, and data visualization.

FARRAH WONG (Senior Member, IEEE) received the B.Eng., master's, and Doctor of Philosophy degrees from Universiti Malaysia Sabah, in 1999, 2001, and 2004, respectively. She has joined the Faculty of Engineering, Universiti Malaysia Sabah, as a Lecturer, since 2004. Her research interest includes intelligent and vision systems.

• • •