IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 2, April 2026, pp. 1790~1803
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i2.pp1790-1803

A sequential attention-enhanced deep learning framework for robust potato leaf disease diagnosis under real field conditions

Watcharkorn Yoochomboon[1], Nithizethe Mhuadthongon[1], Piyaporn Krachodnok[2]
[1]Digital Technology, School of Science and Technology, Sukhothai Thammathirat Open University, Nonthaburi, Thailand
[2]School of Telecommunication Engineering, Institute of Engineering, Suranaree University of Technology, Nakhon Ratchasima, Thailand

Article Info

Article history:

Received Nov 26, 2025
Revised Feb 7, 2026
Accepted Mar 5, 2026

Keywords:

Attention mechanism
Deep learning
Efficient channel attention
Convolutional block attention module
Potato leaf disease
ResNeXt-50

ABSTRACT

Diagnosing potato leaf diseases from images collected in real-life field settings is challenging, mainly because of uneven lighting, complex backgrounds, and disease symptoms that are often subtle or visually inconsistent. In this study, a deep learning-based framework was developed to support potato leaf disease diagnosis, with particular attention given to improving generalization and interpretation. Several convolutional neural network (CNN) architectures were first examined under the same experimental conditions, and ResNeXt-*50* showed the most stable overall performance. The model was then extended by applying efficient channel attention (ECA), followed by spatial attention adapted from the convolutional block attention module (CBAM). Test results indicate that this sequential attention design performs better than the baseline model as well as variants using only a single attention mechanism. Additional evaluation using 300 real-field images collected under different field conditions suggests improved robustness, while visualization results from gradient-weighted class activation mapping (Grad-CAM) show clearer focus on lesion-related regions. Overall, the findings suggest that combining channel-wise and spatial attention can improve both prediction reliability and interpretability, making the approach suitable for practical agricultural use.

Corresponding Author:

Nithizethe Mhuadthongon
Digital Technology, School of Science and Technology, Sukhothai Thammathirat Open University
9/9 Mu 9, Chaengwattana Rd. Bangpood, Pakkret Nonthaburi, Thailand
Email: Nithizethe.Mhu@stou.ac.th

1. INTRODUCTION

The potato (*Solanum tuberosum* L.) is the world's most important food crop after wheat, rice, and maize. The potato is known worldwide for its nutritional value [1]. It is a versatile crop with many industrial uses. Moreover, it plays an important role in global food security. It is particularly important for developing economies. The consumption of processed potato products is increasing rapidly [2]. In Thailand, farmers in the northern and northeastern regions contribute to the national production of over 120,000 tons of potatoes every year across more than 42,000 rai of farmland. Although a highly nutritious and economically important crop, potato productivity is under constant threat from foliar diseases, namely late blight and early blight. According to research conducted in 2022, smallholder farmers may lose up to 70% of their harvest to *Phytophthora infestans* and *Alternaria solani*, driven by favorable humidity and temperature conditions [3]. A more recent study suggests that management factors related to field practices and integrated pest management can lower the severity of early blight and improve disease control efficiency [4], [5]. In an ideal

scenario, the detection of potato leaf diseases would be timely, accurate, and closely aligned with real field conditions, enabling farmers to respond before irreversible yield losses occur.

In an ideal agricultural ecosystem, productivity would be consistent with sustainability, producing high yields without excessive use of pesticides. In reality, frequent disease outbreaks have led Thai farmers to rely heavily on chemical fungicides. While these provide short-term protection, they increase production costs, degrade soil and water quality, and raise concerns over food safety and export potential [6]. Thailand produces over 4.5 billion-baht worth of potatoes each year, yet importing countries such as Japan, South Korea, and Singapore continue to tighten residue standards [7], [8]. Although sustainable pest-management practices can improve environmental performance [9], [10] their adoption in Thailand remains slow, largely due to limitations in disease diagnosis and technology [11]–[13]. Disease identification in rural areas is still based mainly on visual observation, which is often inconsistent and time-consuming [14], leading farmers to overuse fungicides as a precaution. This situation reflects a practical bottleneck, as the lack of reliable, field-ready diagnostic methods prevent sustainability measures from being effectively applied on farms.

Agro-based research on artificial intelligence and deep learning has witnessed increasing attention over the past decade for employing leaf images for disease diagnosis [15], [16]. Convolutional neural networks (CNN) are also effective in classifying many types of plant diseases, often achieving strong accuracy in experimental settings [17]. However, much of this development has been driven by models trained on public datasets and evaluated in controlled environments, such as PlantVillage or Kaggle. Although these datasets are commonly used for benchmarking image-based models, they do not capture the variability and complexity seen in field images [18], [19]. Consequently, models that display robustness in laboratory settings tend to perform poorly when faced with the visual uncertainty of real farm environments.

When models trained on controlled datasets are deployed in real fields, the structural limitations of the data become obvious quickly. The clarity of field images is impaired due to variations in lighting, background clutter, partial leaf occlusion, and disease symptoms. These symptoms do not conform to any specific pattern. Furthermore, they do not stand out distinctly from the natural textures found on the leaves themselves. The characteristics of these images are fundamentally different from those present in the model's training images. As a result, the internal prioritization of features is disrupted. The drop in classification accuracy is not the only issue. It also reflects the degradation of generalization ability when faced with unseen data [20]–[22]. This instability severely erodes user confidence in automated diagnostic systems, particularly when decisions are made under time and resource constraints.

To overcome these limitations, researchers in computer vision have integrated attention mechanisms into deep learning networks. Attention enables models to learn the importance of channels as well as the spatial relevance of features. This capability helps distinguish informative features, which improves model performance [23]–[25]. Nonetheless, the majority of existing works either utilize a single attention mechanism or incorporate attention modules into standard CNN architectures without exploring the interaction between different attention mechanisms [26]–[29]. Thus, these methods do not shed significant light on how attention affects feature selection behavior and decision consistency under real-field conditions.

The utilization of attention mechanisms is now proving increasingly popular in deep learning research. However, there are not many studies that provide a systematic investigation of the impact of attention modules with different structural properties. This limitation is especially observed in architectures with a higher degree of structural complexity, such as ResNeXt-50 [30], which enhances representational capacity through cardinality and parallel transformations [31]. Although ResNeXt-50 is capable of learning complex features from variable data, the existing literature does not provide empirical analyses of how, in practice, three-dimensional (channel-based) and two-dimensional (spatial-based) attention modules operate jointly within ResNeXt-50. The effective combination of efficient channel attention (ECA) and the convolutional block attention module (CBAM) within ResNeXt-50 has not yet been investigated for the classification of potato leaf diseases using real-field images. The absence of knowledge regarding how deep learning models attend to spatial evidence and form decisions in visually unstable agricultural scenarios helps explain this gap.

This study combines ECA and the CBAM in the ResNeXt-50 model for the classification of potato leaf diseases using real-field images. The efficiency of the model is evaluated quantitatively, and the spatial decision behavior is examined using gradient-weighted class activation mapping (Grad-CAM). The aim of this research is to ascertain whether a joint attention design improves feature-selection stability, clarifies the spatial evidence that enters the decision-making process, and enhances the deployability of deep learning models in real agricultural applications. This research work, from an academic perspective, offers insights into the understanding of decision behavior in attention-augmented deep learning models. From a practical perspective, it supports the development of diagnostic systems that better reflect real-world farming constraints and sustainability goals.

2. RESEARCH METHOD

The goal of this study was to create an automated system for assessing potato leaf diseases using deep learning, focusing on performance evaluation under a controlled environment and a real-field environment. The research process was conducted in three main stages. The CiRA CORE platform was used to select the initial baseline model in the first phase. The second phase focused on model refinement and analysis through the incorporation of attention mechanisms on the Google Colab platform. The final phase evaluated the practical applicability of the model using 300 images from a real-field image dataset. The research design was developed to meet the objectives of the study by assessing model stability and generalization ability under realistic agricultural conditions.

2.1. Data collection

The dataset used in this work consisted of a controlled dataset downloaded from a public source as well as a real-field image dataset. The controlled dataset used in this work was sourced from Kaggle, comprising a total of 7,128 potato leaf images classified into three classes, namely early blight, late blight, and healthy [32]. The dataset was split into 80:20 training and testing subsets. This dataset was selected because it is well known in the plant disease classification community, contains high-quality images, and has reliable and clearly annotated labels.

To further test the practical applicability of the developed models, the trained models were also evaluated using a real-field image dataset consisting of 300 images provided by experts from the Thai Department of Agriculture [33]. These images were collected from actual potato farms in Thailand and exhibited substantial variability in lighting conditions, background complexity, and leaf deformation. The high variability of this dataset made it suitable for assessing the models' generalization capability under real-world conditions. Table 1 summarizes the detailed characteristics of the datasets.

Table 1. Summary of datasets used in this study

Dataset source	Class	Number of images	Purpose
Public dataset (Kaggle)	Early blight	2,424	Training/testing
	Late blight	2,424	Training/testing
	Healthy	2,280	Training/testing
Field dataset (DOA, Thailand)	Early blight	100	External validation
	Late blight	100	External validation
	Healthy	100	External validation

2.2. Data preparation

The dataset was preprocessed beforehand to enhance image quality and consistency before model training. Each image was checked to remove the blurred, duplicated, or incorrectly labeled samples so that the noise in the sample could be removed. The quality-control step made sure that the dataset was reliable and trustworthy for developing deep learning models. The importance of preprocessing for improving robustness in vision-based recognition tasks has been widely reported. In practical systems operating with real captured images, preprocessing steps such as noise suppression, contrast enhancement, and structural refinement have been shown to significantly improve recognition reliability under variable lighting and background conditions [34]. The augmentation methods included the following.

Horizontal or vertical inclusion of flipped images as new samples in the dataset. Horizontal flipping is a left-to-right mirroring of an image, while vertical flipping is a top-to-bottom reflection of the image, as shown in Figure 1. To train the model to recognize disease features from different angles, images were rotated at fixed angles (90°, 180°, and 270°) and also random angles between 30° and +30°, as shown in Figure 2. Image brightness adjustment was performed to help with model stability and clearer discrimination of the effects of illumination between samples. The brightness levels of the images were changed to make the image lighter or darker, as in Figure 3.

Figure 1. Examples of data augmentation through horizontal and vertical image flipping

Figure 2. Examples of data augmentation through fixed-angle and random image rotation

Figure 3. Examples of data augmentation through brightness adjustment under different lighting conditions

2.3 Baseline model evaluation using CiRA CORE

The baseline models were developed using the CiRA CORE platform for training and evaluating deep learning architectures in a systematic and consistent manner due to its integrated environment. Five CNN architectures were trained and compared as baseline candidates in this study, namely AlexNet, MobileNet-V2, EfficientNet-B0, ResNet-50, and ResNeXt-50. Each model was deployed using transfer learning and trained under the same training settings. This strategy was adopted to ensure that performance variations were due to architectural differences rather than training conditions. Figure 4 illustrates the workflow of the baseline model selection process.

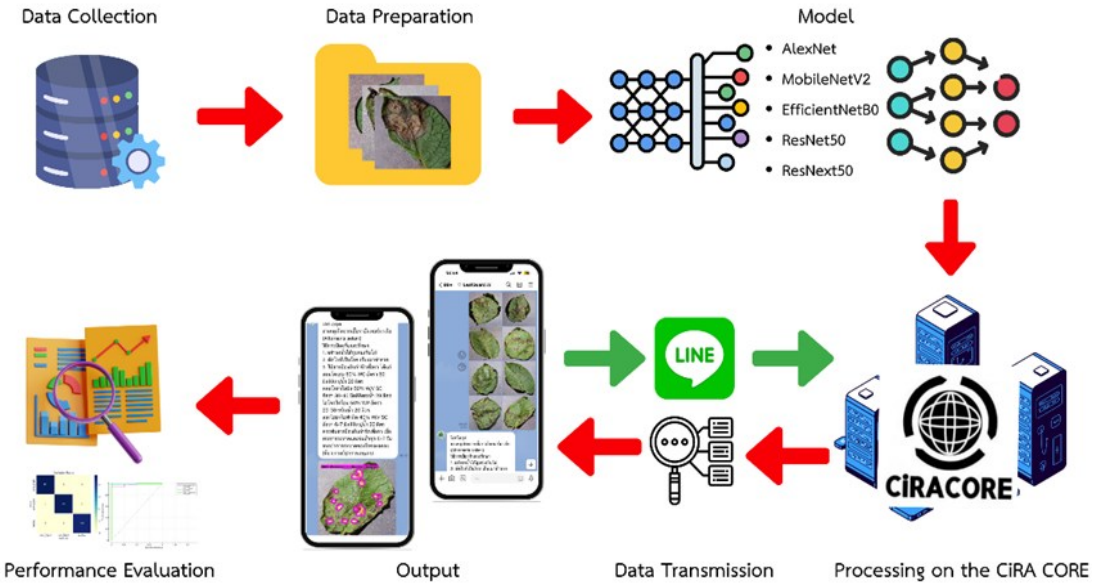

Figure 4. Workflow of baseline CNN model training and selection using the CiRA CORE platform

2.3.1. ResNeXt-50 architecture

ResNeXt-50 is a deep CNN built on the principle of residual learning. It is characterized by multi-branch transformations governed by cardinality, which enhance representational capacity by increasing feature diversity without increasing the depth or width of the network. In this study, the input image was first

processed through convolutional layers, followed by batch normalization and ReLU activation. The spatial dimensions were then reduced via max pooling before entering the deep feature extraction stages. The bottleneck residual blocks were arranged hierarchically to perform feature extraction. Each block consisted of 1×1, 3×3, and 1×1 convolutional layers. In addition, the 3×3 convolution employed grouped convolution with a cardinality of 32 and a base width of 4. This design enabled greater channel-wise feature diversity without increasing computational cost. After feature extraction, the resulting feature maps were passed through a global average pooling layer and a fully connected layer with a SoftMax function to classify images into three categories: early blight, late blight, and healthy. The overall structure of the ResNeXt-50 architecture is shown in Figure 5. ResNeXt-50 was selected as the backbone model because it provides semantically rich feature representations and serves as a suitable foundation for integrating attention mechanisms.

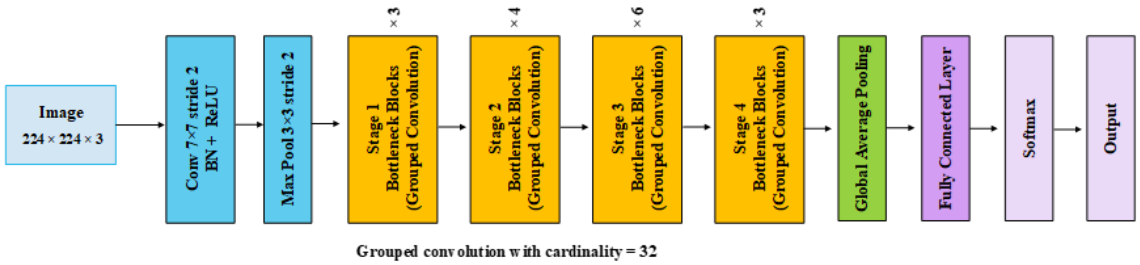

Figure 5. Architecture of the ResNeXt-50 backbone network

2.4. Integration of attention mechanisms into the model architecture

The design of attention mechanisms draws heavily from human visual perception, which allows humans to be selective in recognizing objects and activities. Thus, the human visual perception process can filter information and focus on a specific stimulus while ignoring others. Therefore, attention mechanisms draw inspiration from the ability of humans to filter irrelevant information to perform recognition tasks. Attention mechanisms have enabled deep learning models to emphasize useful visual features in more localized regions of the source images. In the field of computer vision research, attention mechanisms have been widely integrated with CNNs. In order to make channel-wise and spatial-wise features more discriminative, feature weights are automatically adapted.

This study integrated attention mechanisms into the ResNeXt-50 model to improve the model's capacity to learn disease pattern features from potato leaf images. The design incorporated attention in both the channel-wise and spatial domains within the backbone network. This encourages the model to focus on regions containing disease symptoms, thereby reducing noise from background clutter and environmental variation in field images. This technique was especially useful for differentiating two visually similar classes, early blight and late blight, which require more context-specific feature selection.

2.4.1. Integration of channel-wise attention using efficient channel attention

To improve channel-wise feature representation, ECA [24] was integrated into the ResNeXt-50 architecture. ECA is a lightweight attention mechanism that directly learns inter-channel relationships using a simple structure with a low number of parameters. This design allows the model to flexibly reweight channel importance according to the classification task while incurring minimal computational cost.

Within the ECA mechanism, a global average pooling operation is first applied to the feature maps generated by the backbone network to convert spatial information into a channel-wise descriptor. The resulting vector is then passed through a one-dimensional convolution to capture local cross-channel interactions directly, without dimensionality reduction or the use of fully connected layers. Channel attention weights are then generated using a sigmoid activation function and applied to the original feature maps through element-wise multiplication. This approach preserves semantic information in the feature maps and avoids information loss that may result from overly complex transformations. The conceptual architecture of the ECA module is illustrated in Figure 6.

Due to its simplicity and low computational cost, ECA was selected for application to real-field image data that are strongly affected by environmental variation. Previous studies have shown that ECA can be effectively integrated into CNNs to enhance disease-related channel feature representation. These

properties make ECA suitable for strengthening channel-wise discrimination in field-based potato leaf disease classification.

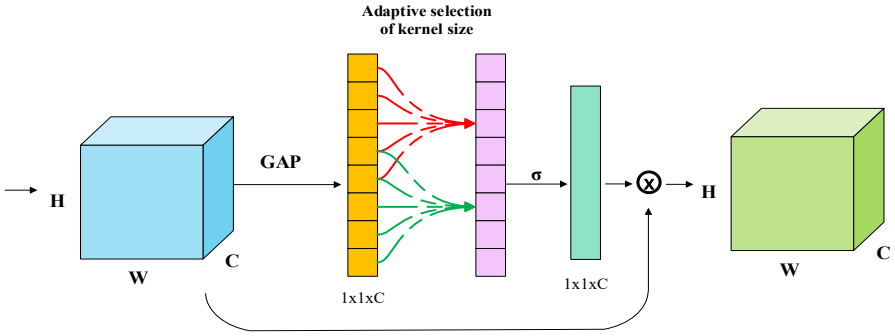

Figure 6. Conceptual architecture of the ECA module

2.4.2. Integration of channel-wise and spatial attention using the convolutional block attention module

To boost both channel-wise and spatial feature representation, the CBAM [25] was fused with the ResNeXt-50 architecture. The design of CBAM is such that channel attention is applied first, followed by spatial attention, to refine the feature maps sequentially. The use of this sequential attention strategy allowed the model to focus on informative features and disregard the background instead of giving all feature vectors equal importance. The structure of the CBAM used in this study follows the original design, as shown in Figure 7.

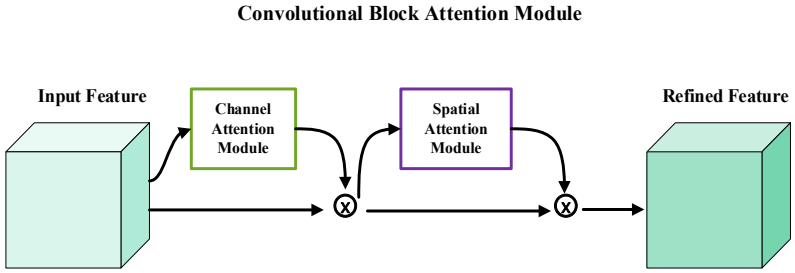

Figure 7. Architecture of the CBAM

The module in the channel attention stage is responsible for identifying which channels in the feature map are most important for the classification task. To capture complementary channel-wise statistics, feature maps were globally average-pooled and globally max-pooled along the spatial dimension. The resulting descriptors were passed through a shared multilayer perceptron and then combined, followed by a sigmoid activation function to generate the channel attention map. The resulting attention map was applied to the base feature map through element-wise multiplication, thereby enhancing channels relevant to disease characteristics. The structure of the channel attention component is shown in Figure 8.

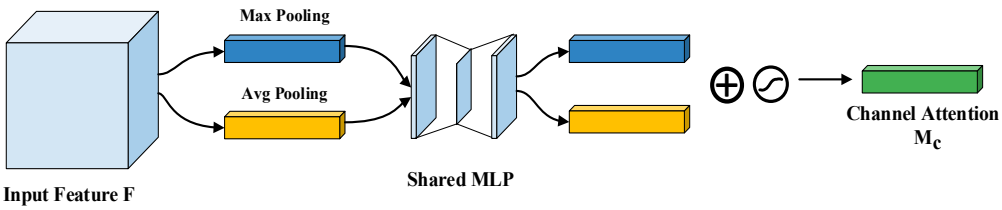

Figure 8. Architecture of the channel attention component in the CBAM

Subsequently, the channel-refined feature map was processed using spatial attention to identify important spatial locations within the feature map. The channel-refined feature map underwent average pooling and max pooling along the channel dimension to generate a spatial descriptor. This descriptor was then passed through a convolutional layer, followed by a sigmoid activation function, to produce the spatial attention map. By applying the spatial attention map to the feature map through element-wise multiplication, the model can emphasize regions indicative of disease symptoms while suppressing background areas.

Figure 9 illustrates the structure of the spatial attention component. By applying channel attention first and spatial attention second, the model was able to learn not only what features were important, but also where they were located. This design was particularly well-suited for field image classification tasks characterized by complex backgrounds and irregular distributions of disease symptoms.

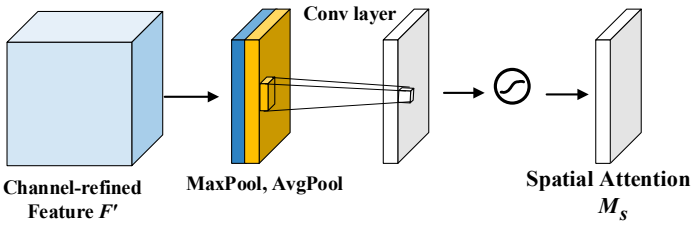

Figure 9. Architecture of the spatial attention component in the CBAM

2.4.3. Sequential integration of efficient channel attention and spatial attention

To enhance the capacity of the model to learn disease-related features in both channel-wise and spatial dimensions, a sequential integration of ECA and spatial attention was incorporated into the ResNeXt-50 architecture in this study. The attention mechanisms were organized from channel to spatial attention within the backbone feature extraction process. This design aimed to enhance feature importance hierarchy while limiting redundancy in attention weighting.

In the proposed structure, the ECA mechanism was applied at the early stage of feature extraction in the ResNeXt-50 backbone to calibrate channel-wise feature responses. By modeling inter-channel relationships directly and maintaining low computational complexity, ECA allows the model to selectively emphasize channels associated with disease characteristics. Without introducing significant architectural or computational overhead.

The feature maps produced by ECA were then forwarded to the spatial attention component of CBAM to identify spatial locations that were most relevant for disease classification. To avoid redundancy, only the spatial attention component was employed at this stage, as channel-wise weighting had already been performed by ECA. Through spatial refinement, the model was able to focus on lesion regions while minimizing the influence of background noise and environmental variation present in real-field images.

Organizing attention mechanisms in a sequential manner enabled the model to regulate learning of both which features were important and where they were located. This design was particularly suitable for analyzing potato leaf images captured under highly variable and complex field conditions. The proposed ECA+Spatial-CBAM+ResNeXt-50 architecture is illustrated in Figure 10.

2.5. Implementation environment

Due to the structural constraints imposed by the CiRA CORE platform that restrict direct modifications of internal components within deep neural network architectures (i.e., specifying and inserting attention mechanisms at the layer level), model development and training were carried out using Python with the PyTorch framework. The researchers opted to use PyTorch for its flexible architecture. It also allows implementation of attention-related adjustments that align with the goals of this research.

All experiments were conducted on a computing system consisting of an AMD Ryzen 7 7840HS processor, 16 GB of main memory, and an NVIDIA GeForce RTX 4050 graphics processing unit (GPU) with 6 GB of dedicated memory. The capacity of this configuration was sufficient to facilitate model training, evaluation, and testing throughout the experiment. To ensure that the training pipeline was run in a systematic manner and that the project was reproducible, a consistent experimental environment was selected.

In order to compare the models in an objective and fair manner, all settings were controlled. All models utilized the same ResNeXt-50 architecture as their backbone and used the same pretrained weights, datasets, data preparation methods, and hyperparameter configurations. Model variations arose solely due to differences in the design and integration of attention mechanisms. The experiment was controlled to create a setting that could measure the effect of attention on model behavior and performance.

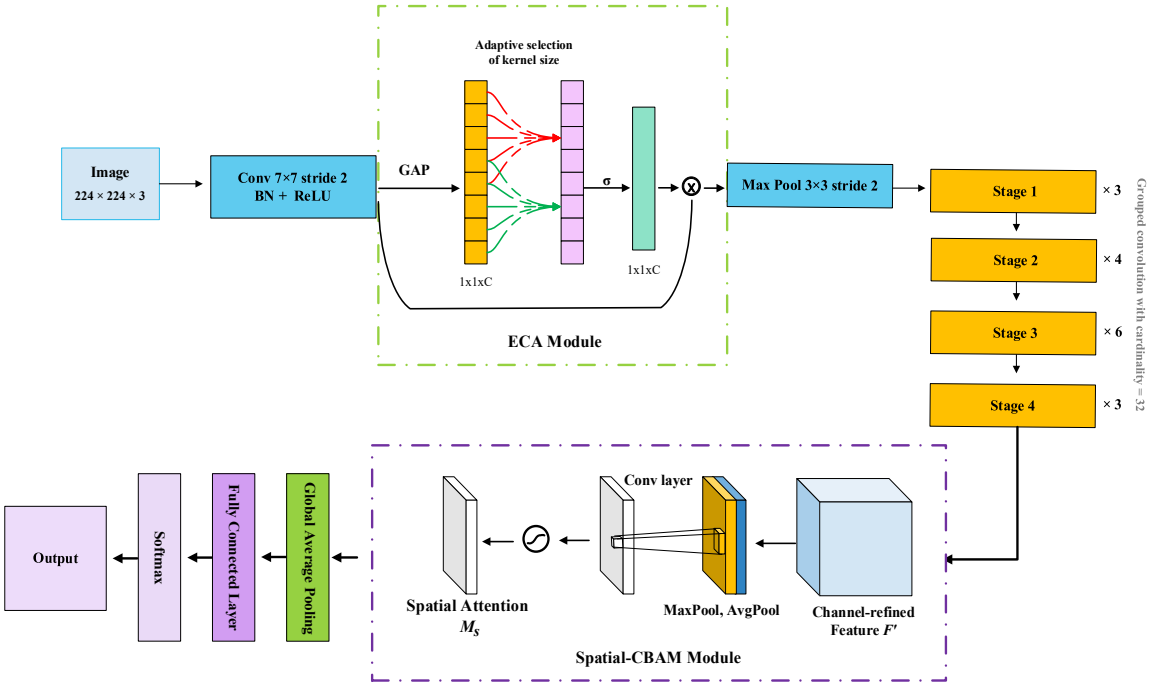

Figure 10. Architecture of the proposed ResNeXt-50 with sequential ECA and spatial attention integration

2.6. Model evaluation and real-field testing

The performance of the model was evaluated using standard metrics applied in image classification, including accuracy, precision, recall, and F1-score. Confusion matrix analysis was conducted to provide deeper insight into classification behavior. Accuracy was used to compute overall predictive correctness, while the remaining metrics evaluated the model's ability to distinguish between visually similar classes, which is particularly important for potato leaf disease classification.

Following model training and selection based on performance on the test dataset, the ResNeXt-50 model with integrated attention mechanisms was evaluated on an external dataset consisting of 300 real-field images. These images were not used during model training or hyperparameter tuning. The field images exhibited noticeable variation in brightness, background complexity, leaf deformation, and disease symptom appearance. The goal of this evaluation stage was to assess the model's ability to generalize to unseen data under practical conditions.

To enhance evaluation credibility and analyze qualitative model decision behavior, Grad-CAM was employed. Grad-CAM visualizes spatial activation maps indicating image regions that contribute most strongly to each classification decision. This analysis was used to examine whether the model focused on lesion regions consistent with biological disease characteristics, thereby supporting model interpretability and strengthening confidence in the reliability of the automated diagnostic system.

3. RESULTS AND DISCUSSION

This section presents the outcomes of the performance evaluation of the deep learning models created for the diagnosis of potato leaf disease. This section also describes the experimental results to facilitate understanding of model behavior under the given experimental settings. All experiments were performed under the same settings to control for possible confounding effects, thereby allowing a fair comparison of models based on genuine architectural differences rather than experimental confounding.

3.1. Baseline model comparison

Baseline model comparison was conducted under identical experimental conditions to ensure fair and unbiased evaluation. To mitigate the impact of confounding variables on performance outcomes, all models were trained and tested using the same dataset, data preparation procedures, and hyperparameters. Initially, five CNN architectures were trained and evaluated as baseline models, namely AlexNet, MobileNet-V2, EfficientNet-B0, ResNet-50, and ResNeXt-50. The assessment of these models, without attention mechanisms, served as a baseline for future performance improvement. Performance evaluation utilized standard classification metrics, including precision, recall, and F1-score for each class, as well as overall accuracy.

From the performance results presented in Table 2, ResNeXt-50 achieved the best performance with an overall accuracy of 99.30%, while ResNet-50 achieved an overall accuracy of 99.16%. Among the evaluated architectures, AlexNet exhibited the lowest performance. These results indicate that the enhanced capacity of the ResNeXt-50 architecture, which is based on grouped convolution and aggregated residual transformations, enabled more effective learning of diverse and discriminative feature representations. This architectural design proved advantageous for accurately classifying potato leaf diseases under controlled conditions. Based on these findings, ResNeXt-50 was selected as the backbone architecture for integrating attention mechanisms and further evaluation in subsequent stages.

Table 2. Performance comparison of the five deep learning models on the test set

Model	Early blight			Late blight			Healthy			Accuracy
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	
AlexNet	93.60	99.59	96.40	99.53	87.01	92.88	93.83	100	96.81	95.44
MobileNetV2	100	91.75	95.70	92.38	100	96.03	100	100	100	97.20
EfficientNetB0	100	98.35	99.16	97.19	100	98.57	99.77	97.15	98.44	98.94
ResNet-50	97.78	100	98.88	99.79	97.73	98.74	100	99.78	99.89	99.16
ResNeXt-50	99.79	99.18	99.52	99.17	98.76	99.00	98.92	100	99.46	99.30

3.2. External validation using real-field images

To test the model's ability to generalize to more realistic scenarios, an external dataset consisting of 300 real-field images was used for evaluation. These images were not incorporated in either training or model tuning. The dataset was sourced from real potato cultivation fields and exhibited significant variations in lighting, background complexity, leaf deformation, and disease symptom appearance. Figure 11 shows representative sample images from the field, where Figure 11(a) illustrates late blight, Figure 11(b) shows early blight, and Figure 11(c) presents healthy potato leaves.

(a) (b) (c)

Figure 11. Real-field potato leaf images from the Thai Department of Agriculture for (a) late blight, (b) early blight, and (c) healthy

After choosing ResNeXt-50 as the backbone model in the baseline study, the impact of employing different kinds of attention mechanisms on disease classification performance was examined. The testing involved four different variants of the model, including the baseline ResNeXt-50, ECA+ResNeXt-50, CBAM+ResNeXt-50, and ECA+Spatial-CBAM+ResNeXt-50. As presented in Table 3, quantitative performance comparisons are reported, and the confusion matrix of the best-performing model is shown in Figure 12.

According to Table 3, all attention-enhanced models consistently outperformed the baseline ResNeXt-50. The ECA module enhances performance by strengthening feature representation on a per-channel basis, whereas CBAM does this by jointly modeling channel and spatial attention. Of all the models studied in this paper, the ECA+Spatial-CBAM+ResNeXt-50 model achieved the highest accuracy

and F1-score. This indicates that the more effective feature representation can be attributed to the complementary roles of attention mechanisms, with channel attention supporting global feature selection and spatial attention facilitating lesion region localization.

The confusion matrix shown in Figure 12 indicates that the model comprising ECA+Spatial-CBAM+ResNeXt-50 achieves excellent classification performance across all classes with high accuracy. The majority of the misclassifications occurred between early blight and late blight, which share similar visual symptoms. This finding is consistent with an architectural analysis indicating that sequential attention integration enhanced the model's ability to discriminate between overlapping and complex disease features in agricultural environments.

Table 3. Quantitative comparison of attention-enhanced models

Model	Attention type	Accuracy	Precision	Recall	F1-score
ResNeXt-50	None	96.00	96.07	96.00	95.98
ECA+ResNeXt-50	Channel	97.33	97.32	97.33	97.32
CBAM+ResNeXt-50	Channel + spatial	96.33	96.44	96.33	96.32
ECA+Spatial-CBAM+ResNeXt-50	Optimized hybrid	98.67	98.68	98.67	98.67

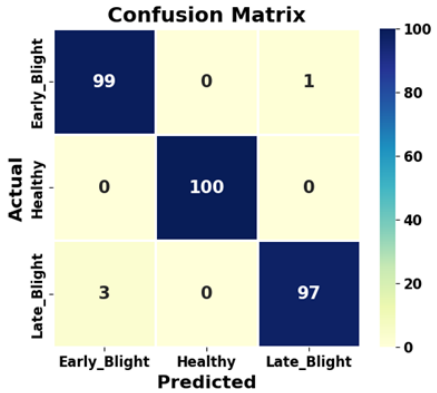

Figure 12. Confusion matrix of the ECA+Spatial-CBAM+ResNeXt-50 model evaluated on real-field images

3.3. Behavioral analysis of the model using grad-CAM

In this study, Grad-CAM was applied for qualitative interpretation in order to gain insights into the internal decision-making processes of the proposed models, as well as to analyze how consistent model predictions were with lesion locations on potato leaves. The main goal of this analysis was to visualize spatial areas that affect classification and to check whether the models are focusing on biologically relevant features for potato leaf diseases. Analyzing model interpretability and agreement for automated plant disease diagnosis is very important. The Grad-CAM activation maps in Figure 13 exhibit the behavior of the baseline ResNeXt-50 model and the attention-enhanced models under various configurations.

The baseline ResNeXt-50 model exhibited relatively blurred activation, with notable responses in background regions outside the lesion areas. The baseline model appears to have relied, in part, on visual cues unrelated to the disease during classification. Incorporating a single attention mechanism, such as ECA or the CBAM, concentrated the activation maps more on lesion areas. However, some background regions still retained activations, indicating that irrelevant features were not always completely suppressed.

In contrast, the ECA+Spatial-CBAM+ResNeXt-50 model produced highly focal and well-localized activation maps that closely matched lesion regions, especially in cases involving small or ill-defined lesions, which are commonly observed in both early blight and late blight under field conditions. The results reveal that the model can learn to focus on biologically meaningful regions within the image by separating attention mechanisms. ECA operates on the channel dimension to select semantic features, while Spatial-CBAM focuses on the spatial dimension for lesion localization

The qualitative results obtained from the Grad-CAM analysis align with the quantitative performance results reported above. In these results, ECA+Spatial-CBAM+ResNeXt-50 achieved the highest classification performance. This analysis demonstrates that attention-based optimization not only improves

numerical accuracy but also enhances interpretability and reliability, both of which are critical for deployment readiness, particularly in automated plant disease diagnosis systems.

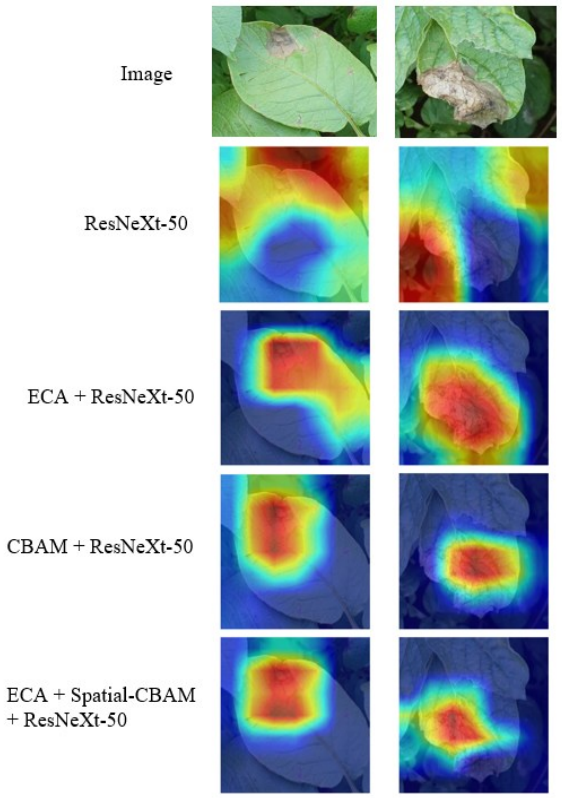

Figure 13. Grad-CAM visualization comparing baseline and CBAM models with different spatial kernels

3.4. Discussion and research implications

The various experiments conducted show that the design and sequential arrangement of the attention modules play an important role in improving the performance of deep neural networks specially designed for potato leaf disease diagnosis under complex and highly variable real-field conditions. The results indicate that using a single attention mechanism may not be sufficient. This is especially true when lesions are small, ill-defined, or embedded in background textures resembling natural patterns.

The proposed sequential integration of ECA with the spatial attention component of CBAM results in balanced learning of channel-wise global semantics and lesion localization in the spatial domain. This setup enables the model to focus on biologically relevant features while remaining robust to background noise and environmental fluctuations. Thus, the attention-enhanced model demonstrated superior generalization performance on real-field images, which is the primary goal of developing models intended for practical agricultural deployment.

Notably, the quantitative improvements in classification performance are mirrored by qualitative evidence from Grad-CAM analysis. The visualization results indicate that the proposed attention configuration guides the model to concentrate on lesion regions in accordance with potato leaf disease pathology. The proposed method is therefore convincing due to the consistency between quantitative and qualitative evidence presented in this study. Furthermore, interpretability-aware model design is both useful and important in agricultural computer vision applications.

From a theoretical perspective, this research advances understanding of the role of attention mechanisms as inductive biases in CNNs. By selectively focusing on complex visual scenarios, such systems identify distinctive features and salient spatial locations. In practice, the findings of this study can support the design of reliable deep learning architectures for plant disease diagnosis, where interpretability and robustness are as important as accuracy.

In summary, this research presents a systematic and effective approach for incorporating attention mechanisms into CNNs for plant disease diagnosis in real-field environments. The proposed

approach has important implications for the development of reliable and usable smart agricultural systems. It also provides a foundation for future research on attention-based feature learning in precision agriculture and decision-support systems.

4. CONCLUSION

This research presents a deep learning-based approach to develop an automated potato leaf disease diagnostic system. The main purpose of conducting the study is to enhance model performance and reliability under complex and highly variable field conditions. The main aim was to design and evaluate attention integration mechanisms that allow deep learning models to learn biologically relevant features and enable their deployment in practical agricultural scenarios. An exhaustive examination of various baseline CNN architectures through the CiRA CORE platform revealed that ResNeXt-50 was the best-performing model. Thus, this model was chosen as the backbone for further enhancement. Experimental results confirmed that the sequential integration of ECA and spatial attention from CBAM enhanced the model's capacity to discriminate potato leaf diseases. The ECA+Spatial-CBAM+ResNeXt-50 model achieved better quantitative performance than the baseline architecture and generalized better on the 300 real-field images that were not used for training or model tuning. Not only did the attention-enhanced models achieve better numerical performance, but qualitative analysis using Grad-CAM also showed that they focused on lesion regions consistent with potato leaf disease pathology. The attention-integrated architecture exhibited more distinct and localized activation patterns in disease areas compared with the baseline model. This improvement enhances model interpretability and increases reliability, which are essential for the deployment of automated plant disease diagnostic systems. Ultimately, this study demonstrates how attention mechanisms can enhance both classification accuracy and interpretability in deep learning models for agricultural images. The proposed attention integration framework can serve as a practical reference for developing reliable and field-ready plant disease diagnostic models. Additionally, this research contributes to the broader scope of intelligent agriculture and decision-support systems, demonstrating that carefully designed deep learning architectures can better align with real-world farming and operational constraints.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Watcharkorn Yoochomboon	✓	✓	✓	✓		✓		✓	✓		✓			
Nithizethe Mhuadthongon	✓	✓		✓	✓	✓	✓			✓		✓	✓	
Piyaporn Krachodnok	✓					✓	✓	✓		✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/mahibullahmudaser/potato-plantvillage.

REFERENCES

[1] FAO, "FAOSTAT-crops and livestock products," *Food and agriculture organization*, 2023. [Online]. Available: https://www.fao.org/faostat/en/#data

[2] Office of Agricultural Economics, "The table shows details about potatoes," *Agricultural Statistics of Thailand*, 2022. [Online]. Available: https://oae.go.th/home/article/505

[3] N. Srisawad, K. Petchaboon, S. Sraphet, P. Tappiban, and K. Triwitayakorn, "Possible reasons affecting different phytophthora infestans populations in tomato and potato isolates in Thailand," *Diversity*, vol. 15, no. 11, 2023, doi: 10.3390/d15111121.

[4] Å. Lankinen *et al.*, "Early blight infection and the influence of biocontrol agents on three wild potato relatives: implications for integrated pest management (IPM) in potato," *Potato Research*, vol. 68, no. 4, pp. 4181–4209, 2025, doi: 10.1007/s11540-025-09905-6.

[5] L. J. Stridh, G. Malm, Å. Lankinen, and E. Liljeroth, "Field and management factors can reduce potato early blight severity: an observational study on farms combined with field trials in Southern Sweden," *Potato Research*, vol. 67, no. 3, pp. 833–859, 2024, doi: 10.1007/s11540-023-09669-x.

[6] B. Kariyanna *et al.*, "Comprehensive insights into pesticide residue dynamics: unraveling impact and management," *Chemical and Biological Technologies in Agriculture*, vol. 11, no. 1, 2024, doi: 10.1186/s40538-024-00708-4.

[7] F. X. D. Verdadero *et al.*, "Pesticides in the environment: benefits, harms, and detection methods," *Sci*, vol. 7, no. 4, 2025, doi: 10.3390/sci7040171.

[8] J. M. M.-Bautista *et al.*, "Environmental and health impacts of pesticides and nanotechnology as an alternative in agriculture," *Agronomy*, vol. 15, no. 8, 2025, doi: 10.3390/agronomy15081878.

[9] W. Zhou, Y. Arcot, R. F. Medina, J. Bernal, L. C.-Zevallos, and M. E. S. Akbulut, "Integrated pest management: an update on the sustainability approach to crop protection," *ACS Omega*, vol. 9, no. 40, pp. 41130–41147, 2024, doi: 10.1021/acsomega.4c06628.

[10] R. S. Raghuvanshi *et al.*, "A review on sustainable plant disease management through integrated approaches," *Journal of Experimental Agriculture International*, vol. 47, no. 12, pp. 719–739, 2025, doi: 10.9734/jeai/2025/v47i123972.

[11] P. Mankong *et al.*, "Assessing life cycle impacts from toxic substance emissions in major crop production systems in Thailand," *Sustainable Production and Consumption*, vol. 46, pp. 717–732, 2024, doi: 10.1016/j.spc.2024.03.013.

[12] Y. Meechoovet and S. Siriwato, "Thailand's Smart Agriculture and its Impacts on Thai Farmers: a case study of smart agriculture in Ayutthaya, Thailand," *SSRN Electronic Journal*, vol. 7, no. 1, pp. 1–17, 2023, doi: 10.2139/ssrn.4540098.

[13] M. Aranguri, H. Mera, W. Noblecilla, and C. Lucini, "Digital literacy and technology adoption in agriculture: a systematic review of factors and strategies," *AgriEngineering*, vol. 7, no. 9, 2025, doi: 10.3390/agriengineering7090296.

[14] K. Minhans, S. Sharma, I. Sheikh, S. S. Alhewairini, and R. Sayyed, "Artificial intelligence and plant disease management: an agro-innovative approach," *Journal of Phytopathology*, vol. 173, no. 3, May 2025, doi: 10.1111/jph.70084.

[15] V. S. Dhaka *et al.*, "A survey of deep convolutional neural networks applied for prediction of plant leaf diseases," *Sensors*, vol. 21, no. 14, 2021, doi: 10.3390/s21144749.

[16] J. Zhao *et al.*, "A review of plant leaf disease identification by deep learning algorithms," *Frontiers in Plant Science*, vol. 16, 2025, doi: 10.3389/fpls.2025.1637241.

[17] T. D. Salka, M. B. Hanafi, S. M. S. A. A. Rahman, D. B. M. Zulperi, and Z. Omar, "Plant leaf disease detection and classification using convolution neural networks model: a review," *Artificial Intelligence Review*, vol. 58, no. 10, 2025, doi: 10.1007/s10462-025-11234-6.

[18] M. Bagga and S. Goyal, "Image-based detection and classification of plant diseases using deep learning: state-of-the-art review," *Urban Agriculture and Regional Food Systems*, vol. 9, no. 1, 2024, doi: 10.1002/uar2.20053.

[19] M. S. Krishna, P. Machado, R. I. Otuka, S. W. Yahaya, F. N. dos Santos, and I. K. Ihianle, "Plant leaf disease detection using deep learning: a multi-dataset approach," *J-Multidisciplinary Scientific Journal*, vol. 8, no. 1, 2025, doi: 10.3390/j8010004.

[20] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018, doi: 10.1016/j.compag.2018.01.009.

[21] M. Long, M. Hartley, R. J. Morris, and J. K. M. Brown, "Classification of wheat diseases using deep learning networks with field and glasshouse images," *Plant Pathology*, vol. 72, no. 3, pp. 536–547, 2023, doi: 10.1111/ppa.13684.

[22] J. G. A. Barbedo, "Factors influencing the use of deep learning for plant disease recognition," *Biosystems Engineering*, vol. 172, pp. 84–91, 2018, doi: 10.1016/j.biosystemseng.2018.05.013.

[23] A. El Hanafy, A. Hessane, and Y. Farhaoui, "Enhancing deep learning models with attention mechanisms for interpretable detection of date palm diseases and pests," *Technologies*, vol. 13, no. 12, 2025, doi: 10.3390/technologies13120596.

[24] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11531–11539, doi: 10.1109/CVPR42600.2020.01155.

[25] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Computer Vision – ECCV 2018*, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.

[26] L. Li and Y. Zhao, "Tea disease identification based on ECA attention mechanism ResNet50 network," *Frontiers in Plant Science*, vol. 16, 2025, doi: 10.3389/fpls.2025.1489655.

[27] X. Wang and J. Liu, "Multiscale parallel algorithm for early detection of tomato gray mold in a complex natural environment," *Frontiers in Plant Science*, vol. 12, May 2021, doi: 10.3389/fpls.2021.620273.

[28] J. Lu, X. Liu, X. Ma, J. Tong, and J. Peng, "Improved MobileNetV2 crop disease identification model for intelligent agriculture," *PeerJ Computer Science*, vol. 9, 2023, doi: 10.7717/peerj-cs.1595.

[29] S. Duhan *et al.*, "Investigating attention mechanisms for plant disease identification in challenging environments," *Heliyon*, vol. 10, no. 9, 2024, doi: 10.1016/j.heliyon.2024.e29802.

[30] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, pp. 5987–5995, doi: 10.1109/CVPR.2017.634.

[31] Z. Chen, H. Yu, S. Song, C. Bi, J. Guo, and X. Ling, "J-Rice-ResNeXt: a deep learning-enhanced framework for high-accuracy Japonica rice varietal classification in precision agriculture," *Industrial Crops and Products*, vol. 237, 2025, doi: 10.1016/j.indcrop.2025.122197.

[32] M. K. Elfouly, A. M. AbdelAziz, W. H. Gomaa, and M. Abdalla, "A deep learning-based framework for large-scale plant disease detection using big data analytics in precision agriculture," *Journal of Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-025-01265-9.
[33] Chiang Mai Royal Agricultural Research Center, "Plant disease and pet diagnostic resources," *Department of Agriculture Thailand*, 2023. [Online]. Available: https://www.doa.go.th/hc/cmrarc/?page_id=14568
[34] S. Phunklang *et al.*, "Automated postal sorting system using optical character recognition and image processing," *2025 IEEE 7th Symposium on Computers & Informatics (ISCI)*, 2025, pp. 310-315, doi: 10.1109/ISCI65687.2025.11167384.

BIOGRAPHIES OF AUTHORS

Watcharkorn Yoochomboon received a Bachelor of Science degree in Information Technology from Suan Sunandha Rajabhat University, Thailand, in 2012, and is a computer scientist at the Office of the Attorney General, Thailand. The area of his research is data processing, artificial intelligence, and the use of deep learning. He can be contacted at email: watcharkorn.y@gmail.com.

Nithizethe Mhuadthongon received his B.Eng., M.Eng., and D.Eng. degrees in Electrical Engineering from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, in 2008, 2010, and 2016, respectively. He currently holds the position of assistant professor at the School of Science and Technology, Sukhothai Thammathirat Open University, Nonthaburi, Thailand. His research interests include antenna design for wireless communications, telecommunication systems, cyber-physical systems (CPS) for intelligent learning, the internet of things (IoT), and automation applications. He is working on various educational projects related to IoT-based CPS platforms for smart agriculture and environmental monitoring. He can be contacted at email: nithizethe.mhu@stou.ac.th.

Piyaporn Krachodnok received her M.Eng. degree in Electrical Engineering from Chulalongkorn University, Thailand, in 2001, and her Ph.D. degree in Telecommunication Engineering from Suranaree University of Technology, Thailand, in 2008. Since 2001, she has been with the School of Telecommunication Engineering at Suranaree University of Technology. She is member of IEEE. Her expertise includes electromagnetic theory, microwave engineering, and antenna engineering, and she has extensive experience in both teaching and research in these areas. She can be contacted at email: priam@sut.ac.th.