

Designing a squeeze-and-excitation-capsule BiLSTM-transformer for plant leaf disease recognition

Aekkarat Suksukont, Ekachai Naowanich

Department of Digital Media Technology, Faculty of Science and Technology, Rajamangala University of Technology Suvarnabhumi, Nonthaburi, Thailand

Article Info

Article history:

Received Feb 21, 2025

Revised Sep 20, 2025

Accepted Oct 18, 2025

Keywords:

BiLSTM

Capsule network

Plant leaf disease recognition

SE-residual blocks

Transformer network

ABSTRACT

Deep learning (DL) is critical in plant disease recognition and classification with precision like those of expert human evaluators. However, development of effective systems is often disrupted due to the complexity and variability of disease pathogenesis. To address these challenges, this research applies to a hybrid DL architecture that integrates spatial encoding, sequential modelling, and attention for visual recognition. This proposed model can incorporate squeeze-and-excitation (SE) with residual blocks, capsule network (CapsNet), bidirectional long short-term memory (BiLSTM), and transformer network (TransNet)-based attention to realize spatial relationships and long-range dependencies for improving recognition accuracy. The proposed model is assessed on the corn leaf disease dataset (CLDD) and rice leaf diseases dataset (RLDD), and its performance is compared to leading-edge models. CLDD and RLDD achieved 99.88 and 99.10% training accuracy respectively. The area under the curve (AUC) reached almost ceiling recognition on CLDD, with 99.73, 99.96, 99.96, and 99.98% for blight (BL), common rust (CR), gray leaf spot (GL), and healthy (HE) result. RLDD results were also high, with 94.98, 93.70, 97.66, 84.57, 99.58, and 98.85% for bacterial leaf blight (BLB), brown spot (BS), HE, leaf blast (LB), leaf scald (LS), and narrow brown spot (NBS), respectively. The results of these tests show the remarkable promise and performance of the proposed model in plant disease recognition applications.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ekachai Naowanich

Department of Digital Media Technology, Faculty of Science and Technology

Rajamangala University of Technology Suvarnabhumi

Nonthaburi, Thailand

Email: ekachai.n@rmutsb.ac.th

1. INTRODUCTION

Plant disease classification and recognition have become a critical focus of precision agriculture, protecting crop yield and quality, and minimizing economic losses for farmers and stakeholders. Proactive and early detection of plant diseases can lead to timely interventions, with the resultant decrease of the spread of infections and optimized utilization of resources (e.g., pesticides and fertilizers), which can in turn contribute to sustainable agricultural practices [1]. However, diagnosing plant diseases by observing leaves is a highly complex task, and even experienced farmers often struggle to identify certain diseases, leading to inaccurate conclusions and delayed treatments. Consequently, traditional methods relying on visual observation are inadequate for modern agriculture. These challenges have spurred research and advancements in computer vision and machine learning (ML) to address these issues [2], [3].

Computer vision and ML have emerged as advanced and effective tools for diagnosing and identifying plant diseases [4], [5]. Traditional ML methods typically involve processes such as image segmentation, feature extraction, and classification [6]. While these networks can partially extract features and learn data autonomously, they often struggle to identify deep features. This limitation negatively impacts their recognition and generalization performance, especially when dealing with background noise or complexities associated with plant disease patterns [7], [8]. Therefore, the development of deep learning (DL), particularly convolutional neural network (CNN), has been enhanced to broaden its capabilities and improve results. For example, MobileNetV3Small was optimized by incorporating squeeze-and-excitation (SE) blocks to enhance the network's ability to capture critical features from the data. Additionally, Hard-Swish activation was introduced to boost learning efficiency and reduce complexity, creating a network with fewer parameters while maintaining high accuracy in plant leaf disease classification [9]. Similarly, the residual network (ResNet) was refined into a modified ResNet, and transfer learning was utilized to enhance the model's generalization and robustness, making it well-suited for plant disease detection [10]. The development of the modified balanced iterative reducing and clustering (MBIRCH) algorithm for segmentation and a hybrid model combining bidirectional long short-term memory (BiLSTM) and CNN for plant leaf disease classification has shown significant progress [11]. Additionally, the creation of a modified lightweight CNN model for fine-grained plant disease classification has focused on distinguishing diseases with similar characteristics. This approach integrates a coordinate attention mechanism to enhance the model's ability to capture the location and context of critical features in plant leaf diseases [12]. Despite these advancements, expanding functional layers in plant disease recognition networks has encountered challenges such as vanishing or exploding gradients, making it difficult for the model to learn effectively or update weights during training. Furthermore, issues like overfitting, computational complexity, and sustainability remain significant barriers, preventing these networks from being practical for real-time plant leaf disease recognition. The application of large networks, such as GoogLeNet [13], ResNet [14], Inception [15], EfficientNet [16], and others [17]–[19], has significantly advanced the leaf diseases classification and recognition.

As reported in the literature, current plant disease recognition models found in the literature are unable to solve the problems of capturing discriminative features, demonstrating reliable properties in real-world settings, and having efficient computational performance. These problems require lightweight design of architecture that compensates for vanishing gradients, noise robustness, and the increase in terms of recognition accuracy and reliability [4]–[19]. In the current work, a hybrid DL framework is proposed integrating a two-stream convolutional operator to model the complementary/multi-scale features of plant leaf diseases. The framework can support feature representation by using SE-residual blocks for channel recalibration, capsule network (CapsNet) for hierarchical spatial relationship representation, BiLSTM layers for sequential dependency modeling, and transformer network (TransNet) modules for attention-based feature optimization. With this model, the network has been able to integrate spatial, sequential, and contextual features to better recognize patterns of disease. Combined, these parts enable the network to learn deep and discriminative representations that can optimize its recognition performance. This architecture enhances the ability of the network to extract highly discriminative features, leading to increasing recognition performance. In this paper, the experimental framework has been evaluated on the corn leaf disease dataset (CLDD) as well as rice leaf disease dataset (RLDD), and the performance compared with the existing approaches for plant disease recognition to test the proposed framework's performance and to introduce the new concept for smart agriculture.

2. PROPOSED METHOD

The proposed network integrates a two-stream convolutional operator with varying kernel sizes to extract multi-scale features from plant leaf images. SE-residual blocks allow for channel recalibration and CapsNet for preserving spatial relations, and BiLSTM layers capture sequential dependencies and focus on globally relevant features with TransNet-based attention. In summary, the global average pooling layer aggregates it to formulate the resulting recognition output.

2.1. Convolutional operator

In Figure 1, images from the dataset are first processed to reduce noise and enhance the data preprocessing. This is achieved by implementing a background detection function to filter excessive color levels and setting a pixel ratio threshold of 0.8, equivalent to 80% of the image weight. The processed images are then passed to the network with parallel convolutional operators [6].

$$A_{ij} = (I * K)(i, j) \sum_m \sum_n I(i + m, j + n) \cdot K(m, n) \quad (1)$$

Where $I(i+m, j+n)$ represents the input pixel, A_{ij} is the output at location (i,j) , $K(m,n)$ is the kernel weight and, b is the bias term. In this study, the convolutional size is set to 32 and 64 in convolutional block 1, and 128 and 256 in convolutional block 2. The result feature maps are forwarded to the SE-residual block to enhance the significance of the extracted features.

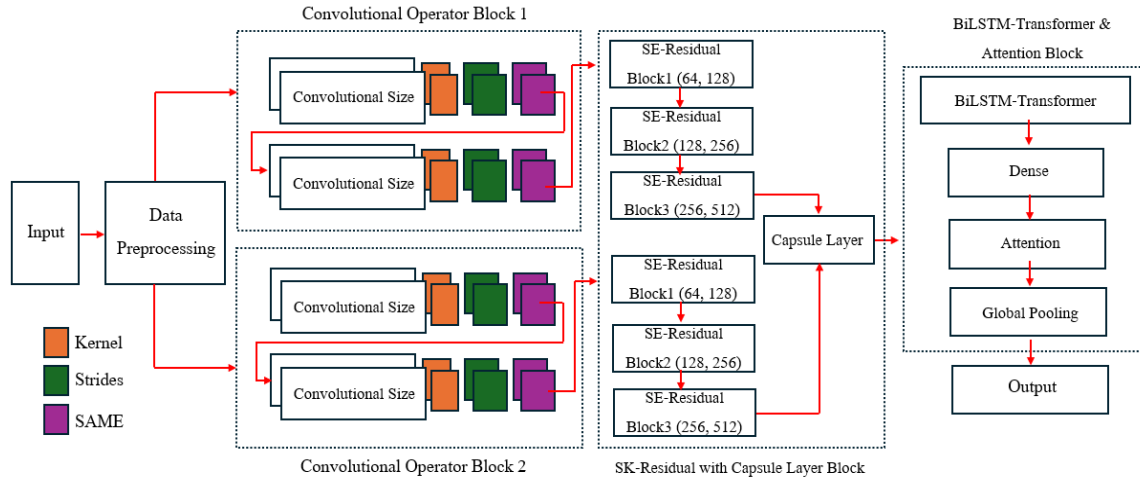


Figure 1. A two-stream SE-capsule BiLSTM-transformer for plant leaf disease recognition

2.2. SE-residual with capsule layer

In the next step, the feature map obtained from the convolutional operator is fed into the residual block integrated with the SE-residual block [20]. This integration enhances the significance of channel attention by disregarding less relevant multi-scale features, thereby emphasizing key specific features. In this study, stacked filters with filter sizes of 64, 128, 256, and 512, along with a stride at 2, are applied to improve the efficiency of feature aggregation across multiple channels as:

$$y = s \cdot F(x, W) + x \quad (2)$$

Where $F(x, W)$ is a linear and non-linear transformation function applied to the input x using the weights W , and s from the SE-residual block performs feature importance enhancement, the resulting output is then passed to the CapsNet [20], [21]. The CapsNet contains 32 convolutional capsules layer with 8-dimensional outputs vector, using a 3×3 kernel with stride 2 to preserve spatial hierarchies, followed by dynamic routing to aggregate feature relationships.

$$C = \text{Capsule}(X_N, W_{cap}, b_{cap}) \quad (3)$$

Where X_N is a feature vector or tensor of size $N \times d$, when N is the number of inputs and d is dimension of each vector. W_{cap} is the weight matrix used to compute the relationship between the capsule layer input and output, and b_{cap} is the bias of the capsule layer. The resulting output is then passed to the subsequent step.

2.3. BiLSTM-transformer process block

In the next step, The tensor output from the CapsNet is then fed into the BiLSTM [22], configured with 128 units per direction (forward and backward), resulting in a concatenated hidden state vector $h_t \in \mathbb{R}^{256}$. BiLSTM computes \vec{h}_t and \overleftarrow{h}_t using the forward LSTM and backward LSTM equations, combining the outputs from both directions as:

$$h_t = [\vec{h}_t, \overleftarrow{h}_t], h_t \in \mathbb{R}^{256} \quad (4)$$

Where h_t is the hidden state derived from the combination of forward and backward hidden states. When the input data $X = \{x_1, x_2, \dots, x_T\}$ is processed by the BiLSTM, the resulting calculation of H is determined as:

$$H = \{h_1, h_2, \dots, h_T\}, H \in \mathbb{R}^{T \times 256} \quad (5)$$

When $H \in \mathbb{R}^{T \times 256}$ is the output tensor of the BiLSTM, with T is the number of inputs and 256 representing the output dimension (as it combines 128 from the forward and 128 from the backward directions). When expressed in terms of evaluating H , when X is the input of size $T \times d_{in}$, with d_{in} is he input dimension, the result of H is determined as:

$$H = BiLSTM(X; W, U, b), H \in \mathbb{R}^{T \times 256} \quad (6)$$

When $X \in \mathbb{R}^{T \times d_{in}}$ is the input tensor, with d_{in} is the input feature dimension; W , U , and b are the parameters of the BiLSTM, and $H \in \mathbb{R}^{T \times 256}$ is the output tensor formed by merging the hidden states from the forward and backward LSTM. The output is then passed to a neural network (NN) with 128 nodes for learning. Subsequently, the resulting vector is fed into the TransNet [23], which utilizes multi-head self-attention and feed-forward NN to capture sequential and spatial relationships from the BiLSTM output. The process involves constructing query (Q), key (K), and value (V), and applying multi-head attention, where each head projects the input into different subspaces to capture diverse dependencies as:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_x}})V \quad (7)$$

When d_x is represents the dimension of the key, multi-head attention allows the model to analyze data features from multiple perspectives simultaneously. The position-wise feed-forward network is then applied to each position independently, further transforming the multi-head attention outputs and improving feature abstraction. Finally, the results are then processed through attention mechanism and global pooling [24], reducing dimensionality and highlighting critical elements to enhance learning and precision.

3. EXPERIMENTAL AND RESULTS

3.1. Dataset

This study utilized two plant disease datasets: the CLDD [25] as shown in Figure 2 and the RLDD [26] as shown in Figure 3. The CLDD consists of 4,188 real corn leaf images divided into four categories: blight (BL) with 1,145 images, common rust (CR) with 1,306 images, gray leaf spot (GL) with 575 images, and healthy (He) with 1,162 images. Meanwhile, the RLDD contains 2,628 rice leaf images classified into six types: bacterial leaf blight (BLB), brown spot (BS), leaf blast (LB), leaf scald (LS), narrow brown spot (NBS), and healthy (HE); each with 428 images, and both datasets were split into 80% for training (60% training and 20% validation) and 20% for testing.

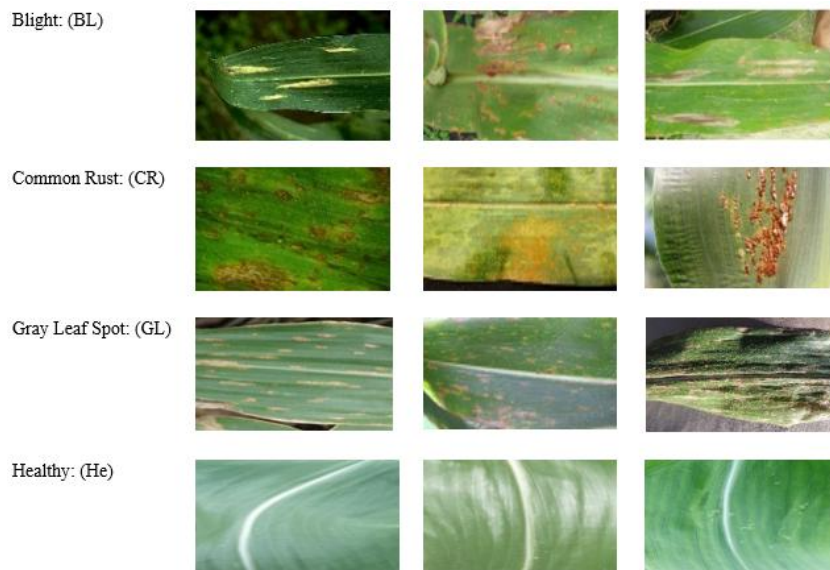


Figure 2. Sample representation of corn leaf disease dataset



Figure 3. Sample representation of rice leaf diseases dataset

3.2. Experiment setup

The experimental system is configured on a Windows-based platform designed to provide high computational efficiency and stability. It features an Intel Core i5-12400F LGA 1700 CPU operating at 2.5 GHz, 32 GB of RAM clocked at 5600 MHz, and an RTX 4070 GPU equipped with 12 GB of VRAM and 5888 CUDA cores, ensuring powerful parallel processing for DL tasks. Furthermore, the experiment employs NumPy and TensorFlow libraries, with the NN trained under the parameters summarized in Table 1 to ensure reproducibility and performance consistency.

Table 1. Parameter for training model

Parameter	Value
Image size	224×224×3
Learning rate	10^{-4}
Epoch	30
Batch size	64
Function	Categorical crossentropy
Optimization	Adam

3.3. Evaluation

The assessment model parameters, including accuracy, precision, recall, area under the curve (AUC), and mean squared error (MSE), are defined as essential metrics for evaluating the performance of the proposed method.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (8)$$

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

$$AUC = \int_0^1 TP(FP)d(FP) \quad (11)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$

When true positive (TP) is instances where a leaf is diseased, and the model accurately identifies the disease. True negative (TN) is a leaf is healthy, and the model correctly predicts it as healthy. False positive (FP) is a healthy leaf is incorrectly recognized as diseased by the model, and false negative (FN) is cases where a diseased leaf is incorrectly identified as healthy. Additionally, when y_i is the actual value of the leaf is diseased, \hat{y}_i is the predicted value by the model, and n is the total number of data points.

3.4. Training performance

Figure 4 presents the overall training results of the proposed model, illustrating its performance across both datasets. In Figure 4(a), the learning curves after 10 epochs demonstrate efficient learning, with CLDD achieving rapid improvement in the early stages and stabilizing at 99.88%, while RLDD exhibits slight fluctuations before reaching a maximum performance of 99.10%. In Figure 4(b), the MSE decreases rapidly during the initial stages and stabilizes at a minimum of 8.22×10^{-4} and 8.027×10^{-4} for RLDD. These results highlight the model's ability to handle complex data effectively, ensuring high performance in both learning and prediction.

Figure 5 shown the precision and recall values of the RLDD and CLDD datasets for 30 epochs. The training precision for epochs 0-10 is higher: it often climbs sharply early on and is more or less equal to 100%. The precision and recall for the RLDD dataset are perfect at 100%: this means that the model learns quickly and continues to perform well consistently. For the CLDD dataset, the precision is stable at 98.85% and recall at 98.61%, less than the RLDD one but still satisfactory. The model performance on both datasets improves rapidly within the first 5 epochs before stabilizing.

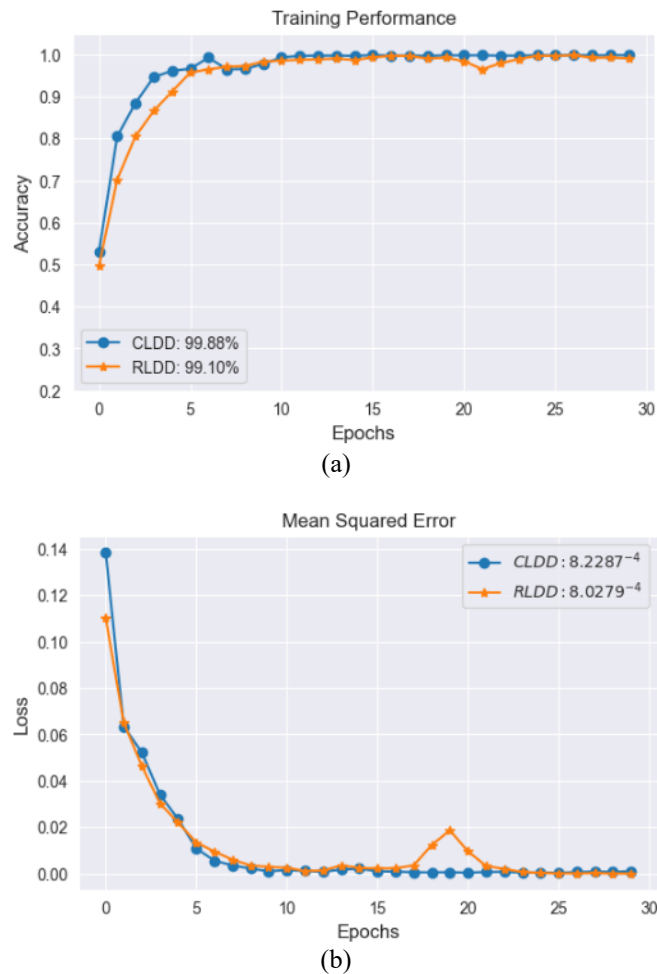


Figure 4. Performance of training of (a) accuracy performance for CLDD and RLDD, and (b) MSE results for CLDD and RLDD

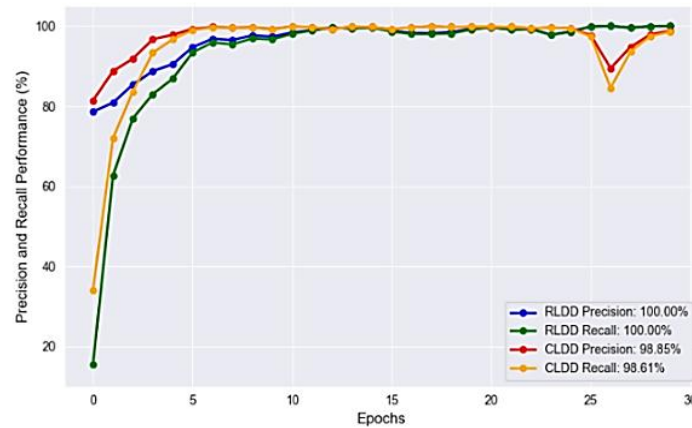


Figure 5. Precision, recall, and AUC performance of training model

3.5. Result of recognition

Figure 6 presents the overall AUC performance of the proposed model across both datasets, illustrating its capability to distinguish between multiple plant disease classes with high reliability. In Figure 6(a) the per-class AUC on CLDD is near ceiling across all categories, with BL 99.73%, CR 99.96%, GL 99.96%, HE 99.98%; micro-average 99.84%. The AUC curves cluster near the top-left with a sharp knee at low false positive rate (FPR) indicating strong separability at practically relevant operating points. HE attains the highest AUC whereas BL is lower by ≈ 0.25 percentage points suggesting occasional confusions for BL under challenging imaging conditions or overlapping visual traits. These results indicate that the model reliably captures disease characteristics and separates diseased from healthy leaves despite substantial sample variability. For completeness the macro-averaged AUC is also reported and class-specific operating thresholds calibrated on validation data are recommended to further mitigate BL-related errors. Furthermore, in Figure 6(b) the RLDD results are high overall with a micro-average AUC of $\approx 96.07\%$. HE, NBS, and LS exhibit the strongest separability $\approx 97\text{--}99\%$, whereas BS and BLB are $\approx 93\text{--}95\%$ particularly at low FPS where its AUC rises more slowly from the origin. This spread suggests residual overlap in possible class imbalance which slightly depress the micro-average relative to the best classes. Overall, the model effectively recognizes a wide range of plant diseases.

In Figure 7, CLDD recognition results for BL, CR, GL, and HE, achieving accuracies ranging from 99.92-100% across all categories. These results highlight the model's ability to accurately identify disease characteristics and differentiate between diseased and healthy leaves. Despite the complexity and variability of the image samples, the model demonstrates stable, and consistent performance, showcasing its effectiveness in learning and distinguishing the features of various leaf diseases.

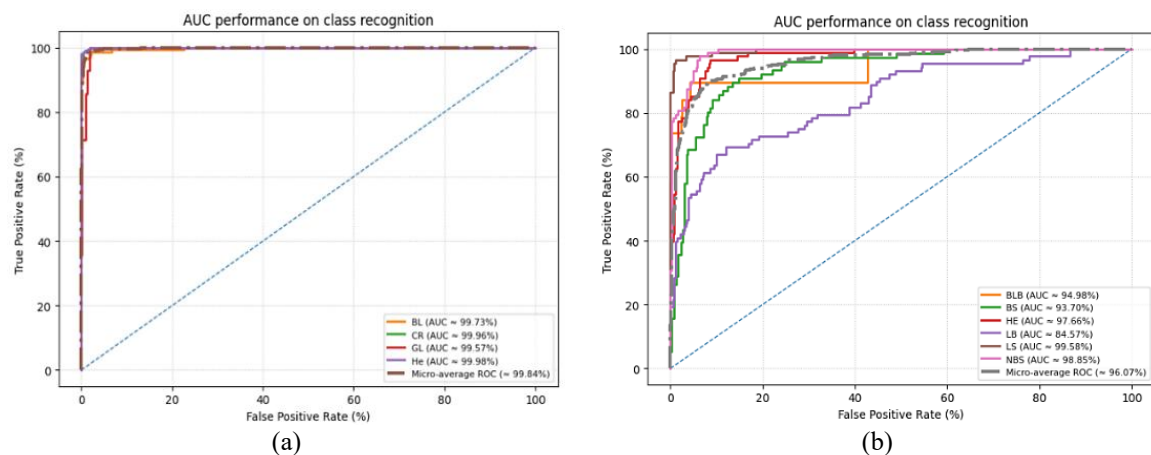


Figure 6. AUC Performance for plant leaf disease of (a) AUC performance of CLDD and (b) AUC performance of RLDD



Figure 7. Sample representation of CLDD recognition performance.

In Figure 8, RLDD recognition results for BLB, LB, BS, HE, LS, and NBS. BLB shows high recognition accuracy of 100% in some samples and an average accuracy of 98-100%, though one sample drops to 90.07%, demonstrating overall reliability in disease recognition. BS achieves high accuracy of 99.99% in some cases, with most samples ranging between 93-99%, reflecting the model's ability to recognize diseases accurately. HE shows a wide range of accuracy, from 67.93-99.67%, reflecting some variability in the model's performance. LB exhibits accuracy ranging from 71.63-100%, with minor errors observed in certain cases. LS consistently achieves high accuracy of 99.99% across many samples, highlighting the model's reliability in disease recognition. NBS achieves the highest accuracy of 99.98%, with averages between 88-99%, indicating the model's effectiveness in recognizing. However, lower accuracies in some cases, such as 67.93% for HE or 71.63% for LB, may stem from ambiguous or complex input data, presenting opportunities for improvement in future studies.



Figure 8. Sample representation of RLDD recognition performance

3.6. Comparison performance

This study compares the experimental results of the proposed method with those of previous research on various plant disease datasets [10]–[14], [27]–[32]. It also compares the method with the original network [6], [20]–[23]. The comparison takes into account key metrics such as accuracy, precision, recall, giga floating-point operations per second (GFLOPs), Params, and *t-test*. The results of the experiment are shown in Table 2.

In Table 2, the results clearly showcase that this model is capable of achieving higher performance with training accuracies of 99.88 and 99.10% for CLDD and RLDD. Precision and recall are at 100% for RLDD, whereas precision and recall are at 98.85 and 98.61% for CLDD, respectively. Moreover, the GFLOPs statistics, parameter counts, and t-test results show comparable performance with other studies. The results are not only testament that it is feasible to implement such designs for accurate performance by capturing very complex spatial and sequential features but also that it could effectively be utilized to diagnose plant diseases in data scarce agricultural settings.

Table 2. Performance comparison of start-of-the-art models

Year/ Paper	Dataset	Accuracy (%)	Precision (%)	Recall (%)	GFLOPs	Params	<i>t-test</i>
2020/ [10]	AI Challenger	91.41	-	-	-	-	-
2024/ [11]	PlantVillage	93	-	-	-	-	-
2022/ [12]	AI challenger 2018	91.94	89.23	88.83	-	2.82 M	-
2022/ [13]	Rice leaf disease	99.58	-	-	-	9.18 M	-
2020/ [14]	PlantVillage	99.40	96.51	-	-	21.8 M	-
2023/ [15]	VillagePlant	99.94	99.89	99.88	-	25 M	-
2022/ [27]	Citrus fruits disease	99.4	99.36	-	-	-	-
	Citrus leaf disease	99.4	99.38	-	-	-	-
2024/ [28]	Plant Village	98.58	98.58	98.58	-	33.26 M	-
	AI Challenger 2018	95.08	95.26	95.08	-	33.26 M	-
2022/ [29]	Rice leaf disease	99.94	-	-	-	0.42 M	-
	Plant Village	99.73	-	-	-	0.42 M	-
2025/ [30]	Soursop leaf disease	99.6	99.6	99.6	-	59.3 M	0.0001
2025/ [31]	Benchmark	-	0.939 (93.9)	0.848 (84.8)	≈ 6.3	2.32 M	-
2024/ [32]	Pepper leaf disease	100.0	-	-	-	0.973 M	-
CNN [6]	CLDD	86.99	88.41	85.79	≈ 0.408	61,496	1.9417×10^{-11}
	RLDD	77.89	70.87	75.45	≈ 0.408	61,666	4.5886×10^{-5}
SE-CNN [20]	CLDD	84.62	86.36	83.13	≈ 0.409	65,437	6.0623×10^{-9}
	RLDD	67.92	76.95	51.15	≈ 0.409	65,607	1.4908×10^{-3}
CapsNet [21]	CLDD	99.76	98.89	99.98	≈ 0.442	1.088 M	3.6723×10^{-5}
	RLDD	97.65	95.62	99.50	≈ 0.442	1.487 M	5.1337×10^{-11}
BiLSTM [22]	CLDD	87.77	88.89	86.56	≈ 0.296	0.756 M	4.4283×10^{-14}
	RLDD	75.63	80.07	69.77	≈ 0.296	0.756 M	8.3007×10^{-1}
TransNet [23]	CLDD	96.23	96.63	95.85	≈ 1.311	16.02 M	3.5613×10^{-6}
	RLDD	67.07	74.43	59.31	≈ 1.311	16.02 M	4.9924×10^{-4}
Proposed	CLDD	99.88	98.85	98.61	≈ 11.69	15.32 M	1.7252×10^{-2}
	RLDD	99.10	100.0	100.0	≈ 11.69	15.32 M	4.0081×10^{-12}

Note: “-” indicates values not reported or not applicable in the cited study.

4. DISCUSSION

This study proposes a two-stream convolutional operator combining SE-residual blocks, CapsNet, BiLSTM-TransNet, and an attention mechanism, which are used to improve plant disease recognition accuracy in this study. The proposed technique uses DL methods to combine the functionality of each module, enabling the model to capture spatial relationships in plant disease images while preserving sequential and contextual information. Based on SE-residual blocks which emphasize critical features, CapsNet maintain part-whole structural relationships through vector-based representations and dynamic routing, BiLSTM with TransNet for modeling sequential dependencies and global dependencies to highlight the most relevant regions. These features together help the network learn features which give richness and discriminative information in a better performance with lower information loss as seen in regular CNN models for more accurate detection of disease patterns.

CLDD and RLDD experiments, the training accuracies of 99.88 and 99.10% are clearly illustrated in Figures 4 and 5 with convergence speed at 5 epochs, where RLDD reaches 100% precision and recall, whereas the CLDD comes across 98.85% precision and 98.61% recall after 10 epochs. Meanwhile, an AUC analysis on CLDD (micro-average ≈ 99.84% in Figure 6) and high performance on RLDD (micro-average ≈ 96.07%) is also presented. Overall, the majority of disease classes (in CLDD) performed well, while BS and BLB (in RLDD) were slightly lower (≈ 93-95%) and LB had the least performance (84.57%),

indicating that the classes have overlapping and imbalanced performance. These findings emphasize that the proposed network architecture should have been able to make an efficient learning and learning of rich, diverse plant disease information, and hence would be applicable to real field agriculture systems.

In Table 2, synthesizes accuracy, precision, recall, GFLOPs, parameter, and *t-test* outcomes across prior work [6], [10]–[16], [20]–[23], [27]–[31], evidencing the ability of the proposed approach to learn disease patterns. Computationally, the proposed architecture operates at about 11.69 GFLOPs with 15.32 MB parameters, markedly heavier than compact CNN and SE-CNN baselines at roughly 0.408-0.409 GFLOPs with 0.061-0.065 MB parameters [6], [20], heavier than CapsNet at about 0.442 GFLOPs with 1.09-1.49 MB parameters [21], and higher in FLOPs yet comparable in parameter count to the TransNet baseline at about 1.311 GFLOPs and 16.02 MB parameters [23]. Furthermore, compared with the lightweight models developed in [30], the proposed model is noticeably smaller, demonstrating its suitability for deployment on portable mobile devices. The paired two-sided *t-test* is performed across model comparisons, and the results were statistically significant ($p < 0.05$). This indicates that the performance improvements of the proposed model over CNN, SE-CNN, CapsNet, BiLSTM, TransNet, and other baseline models are genuine and not due to random variation.

The work has proved the great performance for the diagnosis of plant disease in resource-limited setups and it is also promising for further research with lightweight models for real-time processing [29]–[32]. These results are useful for quick disease identification, minimizing crop losses, and optimizing resource use, thereby contributing to crop protection strategies and safeguarding yield. Because of its adaptability, the model can also be applied to mobile devices and drones [33], [34] for large-scale field monitoring. Experiments involving diverse datasets in the future will increase the reliability of the network and make it useful for sustainable agriculture; while demonstrating the different applications the network can undertake in medical imaging, pest detection, and crop quality grading.

5. CONCLUSION

This study proposed a hybrid DL framework that combines SE-residual blocks, CapsNet, BiLSTM layers, and TransNet-based attention is proposed to enhance recognition accuracy. This structure increases the performance of the network towards improving discriminative feature extraction, therefore enhancing recognition performance. Combined, these methods improve feature extraction and representation learning, leading to more accurate recognition of complex plant disease. The model obtained the highest training accuracies of 99.88 and 99.10% in CLDD and RLDD. So, when considering RLDD dataset, precision is both very high at 100% and recall is at 100% as well, which clearly shows that model learns fast and performs well. In the case of CLDD the precision and recall values were also high at 98.85 and 98.61%, respectively. The AUC achieved almost ceiling recognition for CLDD with BL 99.73%, CR 99.96%, GL 99.96%, HE 99.98%, and this was similarly high for RLDD with BLB 94.98%, BS 93.70%, HE 97.66%, LB 84.57%, LS 99.58%, and NBS 98.85% respectively. These results indicate the model has the capacity of learning complex plant disease properties for different spatial datasets. Its performances are higher than those of previous models, especially on restricted or unfeasible data, promising it most to be used in early disease detection. This results in crop yield protection and resource optimization by supporting sustainable agricultural decision making. In the future, this study highlights not only the high accuracy of the proposed method but also its ability to learn quickly compared to other models. These findings underscore the potential of this approach for leaf disease recognition and pave the way for developing lightweight models with transfer learning. Such advancements could lead to the development of real-time for leaf disease challenges in the agricultural industry.

FUNDING INFORMATION

The authors state no funding is involved.

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
Aekkarat Suksukont	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ekachai Naowanich	✓		✓	✓		✓		✓		✓	✓	✓	✓	✓

C : Conceptualization	I : Investigation	Vi : Visualization
M : Methodology	R : Resources	Su : Supervision
So : Software	D : Data Curation	P : Project administration
Va : Validation	O : Writing - Original Draft	Fu : Funding acquisition
Fo : Formal analysis	E : Writing - Review & Editing	

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in the Kaggle database at <https://www.kaggle.com/datasets/smaranjitghose/corn-or-maize-leaf-disease-dataset>, reference number [25] and <https://www.kaggle.com/datasets/dedeikhsandwisaputra/rice-leafs-disease-dataset>.





REFERENCES

- [1] X. Liu, W. Min, S. Mei, L. Wang, and S. Jiang, "Plant disease recognition: a large-scale benchmark dataset and a visual region and loss reweighting approach," *IEEE Transactions on Image Processing*, vol. 30, pp. 2003-2015, 2021, doi: 10.1109/TIP.2021.3049334.
- [2] Y. Alqahtani, M. Nawaz, T. Nazir, A. Javed, F. Jeribi, and A. Tahir, "An improved deep learning approach for localization and recognition of plant leaf diseases," *Expert Systems with Applications*, vol. 230, 2023, doi: 10.1016/j.eswa.2023.120717.
- [3] P. Hari and M. P. Singh, "Adaptive knowledge transfer using federated deep learning for plant disease detection," *Computers and Electronics in Agriculture*, vol. 229, 2025, doi: 10.1016/j.compag.2024.109720.
- [4] R. Polly and E. A. Devi, "Semantic segmentation for plant leaf disease classification and damage detection: a deep learning approach," *Smart Agricultural Technology*, vol. 9, 2024, doi: 10.1016/j.atech.2024.100526.
- [5] I. Bouacida, B. Farou, L. Djakhedjakha, H. Seridi, and M. Kulay, "Innovative deep learning approach for cross-crop plant disease detection: a generalized method for identifying unhealthy leaves," *Information Processing in Agriculture*, vol. 12, no. 1, 2024, doi: 10.1016/j.inpa.2024.03.002.
- [6] A. Prommakhot, J. Onshaunjit, W. Ooppakaew, G. Samseemoung, and J. Srinonchat, "Hybrid CNN and transformer-based sequential learning techniques for plant disease classification," *IEEE Access*, vol. 13, pp. 122876-122887, 2025, doi: 10.1109/ACCESS.2025.3586285.
- [7] T. Chaturya, Y. Swathi, V. Kumar, P. Karthik, A. Nayan, and A. Yadav, "Detection of plant disease using convolutional neural networks (CNN)," in *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, Erode, India, 2023, pp. 1-6, doi: 10.1109/ICSCDS56580.2023.10105099.
- [8] J. Wongbongkotpaisan and S. Phumeechanya, "Plant leaf disease classification using local-based image augmentation and convolutional neural network," in *2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2021, doi: 10.1109/ECTI-CON51831.2021.9454672.
- [9] A. Salam, M. Naznine, N. Jahan, E. Nahid, M. Nahiduzzaman, and M. E. H. Chowdhury, "Mulberry leaf disease detection using CNN-based smart android application," *IEEE Access*, vol. 12, pp. 83575-83588, 2024, doi: 10.1109/ACCESS.2024.3407153.
- [10] C. Zhang, D. Wu, J. Chen, and J. Yang, "Efficient plant diseases recognition based on modified residual neural network and transfer learning," in *2020 IEEE 18th International Conference on Industrial Informatics (INDIN)*, Warwick, United Kingdom, 2020, pp. 1-6, doi: 10.1109/INDIN45582.2020.9442151.
- [11] V. Jayanthi and M. Kanchana, "Modified BIRCH and hybrid deep framework for tomato leaf disease segmentation and classification," *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, Tirunelveli, India*, 2024, pp. 149-160, doi: 10.1109/ICICV62344.2024.00030.
- [12] Y. Liu, G. Gao, and Z. Zhang, "Crop disease recognition based on modified light-weight CNN with attention mechanism," *IEEE Access*, vol. 10, pp. 112066-112075, 2022, doi: 10.1109/ACCESS.2022.3216285.
- [13] L. Yang et al., "GoogLeNet based on residual network and attention mechanism identification of rice leaf diseases," *Computers and Electronics in Agriculture*, vol. 204, 2022, doi: 10.1016/j.compag.2022.107543.
- [14] V. Kumar, H. Arora, Harsh, and J. Sisodia, "ResNet-based approach for detection and classification of plant leaf diseases," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, 2020, pp. 1-6, doi: 10.1109/ICESC48915.2020.9155585.
- [15] S. Yu, L. Xie, and Q. Huang, "Inception convolutional vision transformers for plant disease identification," *Internet of Things*, vol. 21, 2023, doi: 10.1016/j.iot.2022.100650.
- [16] F. Rajeeva P. P., A. S. U., M. A. Moustafa, and M. A. S. Ali, "Detecting plant disease in corn leaf using efficientNet architecture—an analytical approach," *Electronics*, vol. 12, no. 8, 2023, doi: 10.3390/electronics12081938.
- [17] 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.
- [18] J. Liu and X. Wang, "Plant diseases and pests detection based on deep learning: a review," *Plant Methods*, vol. 17, no. 22, 2021, doi: 10.1186/s13007-021-00722-9.
- [19] I. Attri, L. K. Awasthi, T. P. Sharma, and P. Rathee, "A review of deep learning techniques used in agriculture," *Ecological Informatics*, vol. 77, 2023, doi: 10.1016/j.ecoinf.2023.102217.
- [20] S. Naqvi and M. El-Sharkawy, "SE-RCN: an economical capsule network," in *2023 6th International Conference on Information and Computer Technologies (ICICT)*, Raleigh, NC, USA, 2023, pp. 60-65, doi: 10.1109/ICICT58900.2023.00017.
- [21] J. Pan, X. Cai, D. Mo, Y. Yu, and Y. Li, "Residual attention capsule network for multimodal EEG- and EOG-based driver vigilance estimation," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-12, 2023, doi: 10.1109/TIM.2023.3307756.
- [22] S. Kim and S.-P. Lee, "A BiLSTM-transformer and 2D CNN architecture for emotion recognition from speech," *Electronics*, vol. 12, no. 19, 2023, doi: 10.3390/electronics12194034.





- [23] J. Xie, B. Chen, X. Gu, F. Liang, and X. Xu, "Self-attention-based BiLSTM model for short text fine-grained sentiment classification," *IEEE Access*, vol. 7, pp. 180558–180570, 2019, doi: 10.1109/ACCESS.2019.2957510.
- [24] N. Kim, I.-S. Choi, S.-S. Han, and C. S. Jeong, "DA-Net: dual attention network for haze removal in remote sensing image," *IEEE Access*, vol. 12, pp. 136297–136312, 2024, doi: 10.1109/ACCESS.2024.3459588.
- [25] S. Ghose, "Corn or maize leaf disease dataset," *Kaggle*. Accessed: Jan. 16, 2025, [Online]. Available: <https://www.kaggle.com/datasets/smaranjithhose/corn-or-maize-leaf-disease-dataset>
- [26] VBookshelf, "Rice leaf diseases dataset," *Kaggle*. Accessed: Jan. 14, 2025, [Online]. Available: <https://www.kaggle.com/datasets/vbookshelf/rice-leaf-diseases>
- [27] M. Hassam, M. A. Khan, A. Armghan, S. A. Althubiti, M. Alhaisoni, and A. Alqahtani, "A single stream modified MobileNet V2 and whale controlled entropy based optimization framework for citrus fruit diseases recognition," *IEEE Access*, vol. 10, pp. 91828–91839, 2022, doi: 10.1109/ACCESS.2022.3201338.
- [28] X. Zhang, Y. Mao, Q. Yang, and X. Zhang, "A plant leaf disease image classification method integrating capsule network and residual network," *IEEE Access*, vol. 12, pp. 44573–44585, 2024, doi: 10.1109/ACCESS.2024.3377230.
- [29] S. M. Hassan and A. K. Maji, "Plant disease identification using a novel convolutional neural network," *IEEE Access*, vol. 10, pp. 5390–5401, 2022, doi: 10.1109/ACCESS.2022.3141371.
- [30] S. Mustofa, S. Khan, S. A. Shovo, Y. R. Emon, and M. S. Rahman, "Optimizing soursop leaf disease classification with a lightweight ensemble model and explainable AI," *Current Plant Biology*, vol. 43, 2025, doi: 10.1016/j.cpb.2025.100526.
- [31] J. Yao, Y. Li, Z. Xia, P. Nie, X. Li, and Z. Li, "WTAD-YOLO: a lightweight tomato leaf disease detection model based on YOLO11," *Smart Agricultural Technology*, vol. 12, 2025, doi: 10.1016/j.atech.2025.101349.
- [32] Y. Fu, L. Guo, and F. Huang, "A lightweight CNN model for pepper leaf disease recognition in a human palm background," *Heliyon*, vol. 10, no. 12, 2024, doi: 10.1016/j.heliyon.2024.e33447.
- [33] R. S. Dutta, S. Mukherjee, S. Chowdhuri, S. Mandal, A. K. Dalai, and P. Das, "CNN and edge detection based novel leaf disease identification from bush with irrespective of plant types and drone footage-based real-time application," in *2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN)*, India, 2024, pp. 1–6, doi: 10.1109/CICN63059.2024.10847455.
- [34] G. Fernandez I, Y. G. M, G. S. R. Joshi, S. Lingshetty, R. P. Gowdra and M. Arif, "Smart management of crop monitoring using drone technology," *2025 International Conference on Frontier Technologies and Solutions*, Chennai, India, 2025, pp. 1–7, doi: 10.1109/ICFTS62006.2025.11031877.

BIOGRAPHIES OF AUTHORS



Aekkarat Suksukont     received a B.Eng. degree in Electronics and Telecommunication Engineering from Rajamangala University of Technology Thanyaburi (RMUTT), in 2011, and the M.Eng. degrees in Electronics and Telecommunication Engineering from Rajamangala University of Technology Thanyaburi, in 2017. Currently studying for a Ph.D. degree in Digital Media Technology, Faculty of Science and Technology, Rajamangala University of Technology Suvannabhumi (RMUTSB). His research interests include speech and image processing, computer vision, and applied AI. He can be contacted at email: 166490431006-st@rmutsb.ac.th.



Ekachai Naowanich     received a B.S.Tech.Ed. degree in Computer Engineering from Rajamangala Institute of Technology North Bangkok Campus in 1999, an M.Sc. degree in Computer Science from King Mongkut's University of Technology North Bangkok (KMUTNB) in 2008, and a Ph.D. degree in Information and Communication Technology for Education from King Mongkut's University of Technology North Bangkok (KMUTNB) in 2017. His research interests include machine learning and applied AI for education. He can be contacted at email: ekachai.n@rmutsb.ac.th.