A CNN-BiLSTM Hybrid for Plant Leaf Disease Classification: Comparative Performance Evaluation of Deep Learning Algorithms

Aekkarat Suksukont and Ekachai Naowanich*

Department of Digital Media Technology, Faculty of Science and Technology,
Rajamangala University of Technology Suvarnabhumi (RMUTSB), Nonthaburi, Thailand
Email: 166490431006-st@rmutsb.ac.th (A.S.); ekachai.n@rmutsb.ac.th (E.N.)
*Corresponding author

***Abstract*—The existence of plant leaf diseases is a big problem for farmers all over the world because they make crops less healthy and less plentiful, which puts global food security at risk. The most common problems with diagnosing plant leaf diseases are a lack of experience, different ways of undertaking visual assessments, and image overlaps, all of which can lead to wrong diagnoses. This study introduces a hybrid deep learning architecture that integrates squeeze-and-excitation residual blocks, capsule networks, bidirectional long short-term memory, and attention mechanisms. The design utilizes convolutional operators for effective feature extraction, Squeeze-and-Excitation (SE) block for channel reweighting, capsule networks for spatial relationship capture Bidirectional Long Short-Term Memory (BiLSTM) for sequential dependencies, and attention mechanisms for emphasizing prominent features. Experiments were performed on 2 empirical datasets: the Corn Leaf Disease Dataset (CLDD) and the Rice Leaf Disease Dataset (RLDD). The data were divided into 60% for training, 20% for validation, and 20% for testing. The proposed method attained 99.88% training accuracy on CLDD and 100% on RLDD. During testing, the class-wise accuracies were 99.29% for blight and 100% for the other CLDD categories. In the case of RLDD, the accuracies attained were 78.95% for bacterial leaf blight, 85.53% for brown spot, 89.77% for healthy samples, 77.27% for leaf blast, 100% for leaf scald, and 97.73% for narrow brown spot. This work highlights practical potential for deployment in terms of mobile applications, enabling farmers to obtain rapid, reliable, and cost-effective field diagnoses, thereby improving agricultural productivity and sustainability.**

***Keywords*—plant leaf disease classification, deep learning algorithms, convolutional neural network**

I. Introduction

Plant diseases are a big problem for farmers all over the world because they lower crop yield, reduce product quality, can have serious cost implications, and put global food security at risk [1–3]. To manage and reduce these diseases effectively, it is important to diagnose them accurately and quickly. Traditional methods, which often rely on expert evaluation or visual assessment, are often ineffective, take a long time, and depend heavily on specialized knowledge. In addition, when resources are limited, relying on manual monitoring leads to increased delays in both diagnosis and prevention. These limitations show how important it is to have advanced methods that make it easier to quickly, reliably, and accurately identify plant diseases. Deep Learning (DL) has recently emerged as a promising methodology, enhancing diagnostic precision and reducing dependence on manual expertise [4, 5].

In recent years, DL, especially Convolutional Neural Networks (CNN), has garnered significant interest due to its capacity to learn and extract spatial features, rendering it exceptionally effective with regard to the analysis of plant leaf diseases [6, 7]. This has led to significant progress in developing techniques for leaf disease detection and classification, for example, Thaseentaj and Ilango [8] proposed a customized deep CNN framework for the automatic detection and classification of mango leaf diseases. This approach utilizes image preprocessing and data augmentation techniques to improve classification performance across multiple disease categories. Experimental results demonstrate that the deep CNN model achieves a high degree of accuracy and outperforms CNN architectures, indicating its effectiveness as a reliable tool for mango leaf disease diagnosis and precision agriculture applications. Lokhande *et al.* [9] proposed a comparative analysis of different CNN architectures for plant leaf disease classification and detection. The study evaluates the performance of models such as AlexNet and Residual Network (ResNet-50) using image datasets of crop leaves, demonstrating the effectiveness of DL approaches for accurate plant disease recognition. Kaviarasu *et al.* [10] proposed a hybrid CNN- Extreme Gradient Boosting (XG-Boost) framework, designed for the identification of root rot disease from coconut leaf images. This framework assigns the task of extracting high-level features to the CNN, while XG-Boost takes on the classification role,

Manuscript received December 10, 2025; revised February 9, 2026; accepted March 20, 2026; published July 10, 2026.

achieving results that exceed those of traditional Machine Learning (ML) and independent DL models. This investigation further emphasizes the potential for scalability in applying this method to real-time smart agriculture and sustainable farming practices.

Shanmugam *et al.* [11] proposed a real-time plant disease detection system, utilizing fine-tuned YOLOv11 on PlantVillage/PlantDoc data and integrating Explainable AI (XAI) to provide farmer-friendly, text-based explanations. Paul *et al.* [12] proposed a web and android application to assist farmers with the real-time classification of tomato leaf diseases. Hussain and Srikaanth [13] suggested that Visual Geometry Group 16-layer Network (VGG16) and Visual Geometry Group 19-layer Network (VGG19) are at the core of this state-of-the-art system, applied with transfer learning while underscoring the role that data augmentation can play in terms of the accuracy of models. Kaur *et al.* [14] proposed CNN based on VGG19 to identify paddy leaf diseases such as rice blast, bacterial blight, brown spot, and healthy leaves using image datasets from Kaggle and PlantVillage/PlantDoc for fine-tuning purposes. The model is tuned by training only the last layers after preprocessing and segmentation on Hue, Saturation, Value (HSV). The results found that VGG19 outperformed VGG16 in this respect with computational feasibility maintained. Hessane *et al.* [15] proposed an image analysis methodology integrating ML emphasized feature extraction with 80 gray level co-occurrence matrix and HSV features. The approach has been evaluated with Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), and Light Gradient Boosting Machine (Light-GBM) in order to identify and categorize disease epidemic intensities.

Taji *et al.* [16] proposed an ensemble DL strategy that integrates CNN and Local Binary Patterns (LBP) with binary dragonfly, ant colony, and moth flame optimization algorithms for the categorization of plant leaf diseases. Muthusamy and Ramu [17] proposed an experiment with three kinds of networks to achieve enhanced classification efficiency, where ensemble learning with CNN by parameter optimization in the dense layers, and amalgamating many networks through an average approach, was proposed. The results so obtained would therefore give insights into the use of CNN in studying and classifying plant diseases. However, adequate architecture tuning and optimization with regard to solving real-world problems in agricultural applications are mandatory. Zhang *et al.* [18] proposed the hybridization of Residual Network (ResNet) and Capsule Network (CapsNet) to better extract structural information and the location of lesions. Thus, leading to more accurate models dealing with variations in images. In addition, ResNet has been tuned with the use of an attention mechanism to further enhance its ability to pick salient features of diseases. Prommakhot *et al.* [19] proposed a hybrid DL architecture combining 2-stream CNN, Bidirectional Long Short-Term Memory (BiLSTM), Sequence-to-Sequence (Seq2Seq), and Gated Recurrent Unit (GRU). This architecture was proposed for plant leaf disease classification. The two Convolutional Layers–Bidirectional Long Short-Term Memory (TwoConV-BiLSTM) model outperformed other investigated models when trained on the PlantVillage dataset with data augmentation, demonstrating strong generalization and performance, while suggesting room for improvement through knowledge distillation, quantization, and pruning. However, its performance was slightly worse than that of some state-of-the-art models. Recently, Suksukont and Naowanich [20] have explored hybrid DL architectures for plant leaf disease classification by integrating SE-block, Attention Mechanisms (AM), Capsule Networks (CN), and recurrent learning modules to enhance feature representation and spatial awareness.

Even though CNN-based models have made some progress, many current methods still have problems with regard to visual inconsistencies, which often lead to misclassification. In this study we use CN, SE-block, BiLSTM, and AM to make learning more efficient and more accurate. The CN helps to identify small patterns and spatial relationships in data, thus eliminating the need for large datasets or massive data augmentation. During training, the SE-block retains important information, enabling deep networks to perform better. The AM also helps to draw attention to the most important features, thereby reducing information loss. Meanwhile, BiLSTM improves feature extraction by modeling the relationship between data points in both directions. These approaches enable the model to handle complex datasets more efficiently, making it ideal for use in agricultural environments. This research presents improvement in terms of an integrated CNN and BiLSTM for plant leaf disease in order to increase classification. It also engages in performance analyses with DL models in order to evaluate relative performance and show improvements.

II. Experimental Method

A. Experimental Method

This study included 2 disease datasets, with the Corn Leaf Disease Dataset (CLDD) [21] utilized for testing, as illustrated in Fig. 1. The dataset comprises 4188 authentic photographs of corn leaves, categorized into 4 disease characteristics: Blight (BL) with 1145 images, Common Rust (CR) with 1306 images, Gray Leaf (GL) spot with 575 images, and Healthy Leaves (He) with 1162 images. Additionally, the Rice Leaf Diseases Dataset (RLDD) [22] was utilized for experimentation. The dataset comprises 2628 rice leaf photos classified into 6 disease categories: Bacterial Leaf Blight (BLB), Brown Spot (BS), Leaf Blast (LB), Leaf Scald (LS), Narrow Brown Spot (NBS), and Healthy (HE), with each category containing 428 images, as illustrated in Fig. 2. The photos were systematically arranged and classified for the training and evaluation of the DL model. The dataset was partitioned into a training set consisting of 80% of the data (60% for training, 20% for validation) and a testing set comprising 20%.

The datasets exhibit minor class imbalance, particularly in the CLDD; however, all categories contain sufficient samples to support supervised learning. The images were collected under diverse environmental conditions, including variations in illumination, background

complexity, leaf orientation, and partial occlusion, resulting in a heterogeneous image quality that reflects realistic field scenarios. This diversity enhances the robustness and credibility of the experimental evaluation. All experiments were conducted using the TensorFlow/Keras deep learning framework.

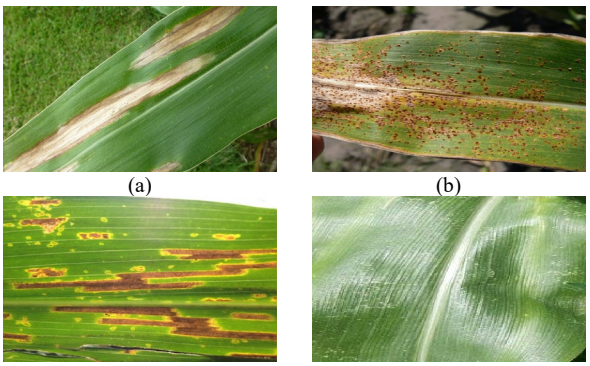

Fig. 1. Sample image of Corn Leaf Disease Dataset (CLDD). (a) Blight (BL). (b) Common Rust (CR). (c) Gray Leaf (GL). (d) H Healthy Leaves (He) e.

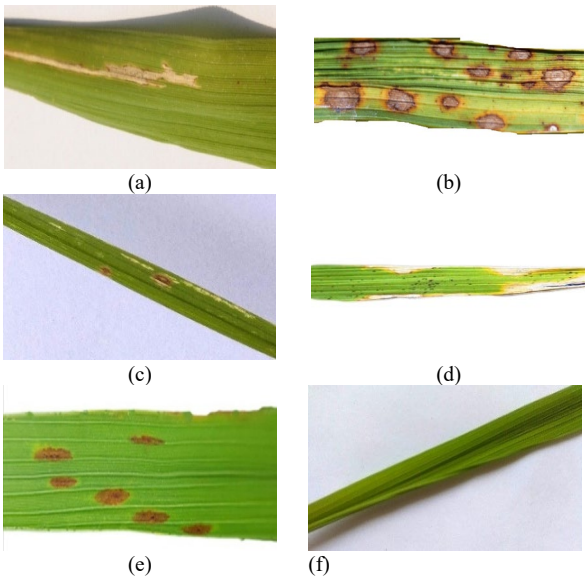

Fig. 2. Sample image of Rice Leaf Diseases Dataset (RLDD). (a) Bacterial Leaf Blight (BLB). (b) Brown Spot (BS). (c) Leaf Blast (LB). (d) Leaf Scald (LS). (e) Narrow Brown Spot (NBS). (f) Healthy Leaves (He).

B. Deep Learning Network Design

All input images were resized to 224×224×3 pixels. Pixel intensity values were normalized from [0, 255] to [0, 1] using a uniform rescaling operation. The same preprocessing pipeline was consistently applied to all DL models to ensure a fair and unbiased comparison. During training, data augmentation was applied to improve generalization, while overlapping leaf regions were handled implicitly by the proposed architecture without explicit segmentation.

To improve the robustness and generalization ability of the model, several data augmentation techniques were applied during the training process. These augmentation strategies include random horizontal flipping, small-angle rotation within ±10°, scaling within the range of 0.8–1.2, and Gaussian noise addition. These transformations were applied randomly to the training images to simulate variations in leaf orientation, size, and environmental conditions. The augmentation process helps the model learn more representative features and reduces the risk of overfitting, thereby improving the overall classification performance. All augmentation operations were applied only to the training dataset, while the validation and test datasets remained unchanged to ensure fair performance evaluation.

Fig. 3 illustrates the overall pipeline of the proposed DL architecture for plant leaf disease classification. After preprocessing, which aims to reduce noise and enhance image quality, the input images are first processed by convolutional layers equipped with SE-block to extract hierarchical feature representations and perform channel-wise feature recalibration. The extracted feature maps are then forwarded to the CN to preserve spatial relationships among local features. Subsequently, the capsule outputs are passed to a BiLSTM layer to capture contextual dependencies, followed by the use of an attention mechanism that emphasizes the most discriminative features. Finally, Global Average Pooling (GAP) and a fully connected layer are employed to generate the final classification output.

In the next step, the images are input into the designed network. In this experiment, we utilize an existing convolutional network, VGG19 [23], as well as a modified VGG19 structure. The modified VGG19 integrates the functionality of a CNN, SE-block, and an LSTM, collectively referred to as the Modified-VGG19 [24].

In a subsequent step, the study further enhanced the network by incorporating 2 additional SE-block, bringing the total to 3 blocks, and integrating BiLSTM, AM and GAP to improve network performance CN and LSTM. This design aimed to enhance the model's performance across all dimensions.

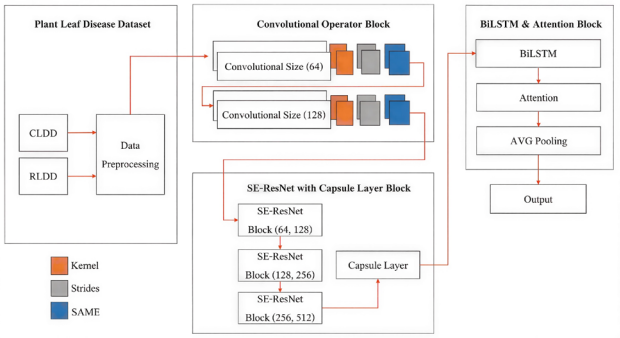

Fig. 3. Proposed method.

Fig. 3 shows the components and parameters of the network. This consists of convolutional layers as delineated in Eq. (1), with the number of convolutional filters set to 64 and 128 channels. Each layer consists of two stacked convolutional operators, with an increasing number in each block, designed to identify data features using a 3×3 filter with SAME (a padding strategy that preserves the input spatial dimensions) padding to

maintain the intended image size during processing. To reduce the spatial size while emphasizing the original core features, we use a GAP size of 2×2 with a 2×2 stride.

$$O(i,j) = \sum_{m=0}^{f-1} \sum_{n=0}^{f-1} I(i+m, j+n) \times K(m,n) + b \quad (1)$$

where $O(i,j)$ denotes the output feature map at position (i,j), $I(i+m, j+n)$ represents the input feature map value at spatial location $(i+m, j+n)$, and $K(m,n)$ denotes the convolution kernel (filter) at position (m,n). The indices m and n iterate over the spatial dimensions of the kernel of size $f \times f$. The term b denotes the bias applied to the convolution operation.

The residual block, which is integrated with the SE-block [25] to improve the model's focus on particular features, receives the convolution results. This is accomplished by layering filters (block 1 to block 3) with channel dimensions of 64, 128, and 256 filters, with a stride of 2 to enhance feature aggregation across various channels as Eq. (2):

$$\mathrm{y} = F(x\{W_i\}) \times \mathrm{s} + x \quad (2)$$

When $F(,\{W_i\})$ represents a linear and non-linear transformation function applied to the input x using the weights $\{W_i\}$, s from the SE-block serves to enhance the importance of specific features.

The output from the SE-block is transmitted to a CN [26], designed with 32 capsules and a capsule diameter of 3×3 filter, while a stride of 2 is employed to capture intricate spatial correlations as Eq. (3):

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

When C denotes the output of the capsule network, X_N is a feature vector or tensor of size $N \times d$, with N representing the number of input CN and d the dimension of each vector. W_{cap} is the weight matrix used to compute the relationship between the input and output capsules, and b_{cab} is the CN bias or stretching value. The experimental results are then passed to the BiLSTM with 128 units in each direction (forward and backward).

The BiLSTM leverages the forward and backward LSTM [27] to compute the forward hidden state $\vec{h}_t$ and the backward hidden state $\overleftarrow{h}_t$. These 2 hidden states are concatenated to form the final hidden state $h_t \in \mathbb{R}^{256}$ when h_t is the concatenated hidden state. This operation is expressed as Eq. (4):

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

When the input data $X = \{x_1, x_2, \ldots, x_T\}$ is processed by the BiLSTM, it computes the hidden states in both forward and backward directions. The combined result H was calculated as Eq. (5):

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

when $H \in \mathbb{R}^{T \times 256}$ denotes the output tensor of the BiLSTM, with T representing the sequence length and 256 denoting the output dimension obtained by concatenating the forward and backward hidden states. The result of H is determined as Eq. (6):

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

When $X \in \mathbb{R}^{T \times d}$ represents the input sequence with length T and feature dimension d. The parameters W, U, and b denote the trainable weight matrices and bias vectors of the BiLSTM. The output H is the hidden representation at all time steps, where each hidden vector is obtained by concatenating the forward and backward LSTM hidden states.

An AM [28] takes the output of the BiLSTM and picks out the most significant parts of the sequence so that the model can learn, or make decisions based on those parts. In addition, GAP is utilized to lower the number of dimensions in the data, which makes the network's training more effective.

The proposed network was trained using the parameters summarized in Table I.

TABLE I. PARAMETERS FOR MODEL TRAINING

Parameter	Value
Image size	224 × 224 × 3 (pixels)
Learning rate	10^{-4}
Epoch	50 (epochs)
Batch size	64 (samples per batch)
Loss function	Categorical cross-entropy
Optimization	Adam

C. Evaluation

Accuracy is employed to assess training efficiency by determining the ratio of accurately predicted samples to the total number of samples. Which ca be calculated as Eq. (7):

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (7)$$

True Positive (TP) denotes the count of correct samples accurately predicted by the model, True Negative (TN) signifies the count of incorrect samples correctly identified as negative, False Positive (FP) reflects the count of incorrect samples erroneously predicted as positive, and False Negative (FN) indicates the count of correct samples that the model inaccurately predicts as negative.

The confusion matrix was employed to evaluate and show the model's classification findings, with the outcomes formatted in a table for enhanced interpretation [29].

III. RESULT

A. Training Performance

The training performance of the models is shown in Fig. 4. Fig. 4(a), shows how well the models that were trained on the CLDD dataset did when compared to each other. The recommended strategy has an accuracy at 99.88%, which is the best. During the first 10 epochs of training, it keeps getting better, and it stays that way for the rest of the time. VGG19, on the other hand, has an

accuracy at 99.70%, which put it in second place. However, its training showed changes that happened every 10 to 30 epochs. When trying to combine certain data patterns, problems arise that show overfitting is possible. The Modified-VGG19 model had the lowest accuracy rate at 97.65%. The results showed that growth slowed down during the first training phase and stability dropped off in later epochs. The results show that the suggested strategy works well because it uses BiLSTM, SE-block, and AM to help it learn spatial features and work with different types of data.

Fig. 4. Training performance for plant leaf disease. (a) Training performance of CLDD. (b) Training performance of RLDD.

Fig. 4(b) shows how well the models trained on the RLDD dataset did when compared to one another. The suggested method worked perfectly, with a 100% success rate. It learned quickly during the first 10 epochs and kept up the same level of performance for all 50 epochs. VGG19, on the other hand, came in second with an accuracy at 98.53%, but its training showed differences of 20 to 30 epochs. The Modified-VGG19 model had the lowest success rate at 96.11%. This means that it got better at a slower rate than the original VGG19 model and the new method.

B. *Classification Performance*

Fig. 5 offers a confusion matrix which shows how well these models were able to sort things. Fig. 5 shows classification results for the CLDD dataset with VGG19, Modified-VGG19, and the proposed method. The proposed method works best, with a score of 100% in all categories with the exception of BL, with 99.29%. This shows that it works well with data that is hard to understand. VGG19, on the other hand, doesn't do as well. It only gets 98.58% of the TP rate in the BL group, even though it gets perfect scores in other groups. The Modified-VGG19 does the worst, getting 35.90% with regard to the He group and 51.33% with regard to the BL group. The suggested method is the best way to sort through complicated data because it uses CNN, BiLSTM, SE-block, and AM all at once.

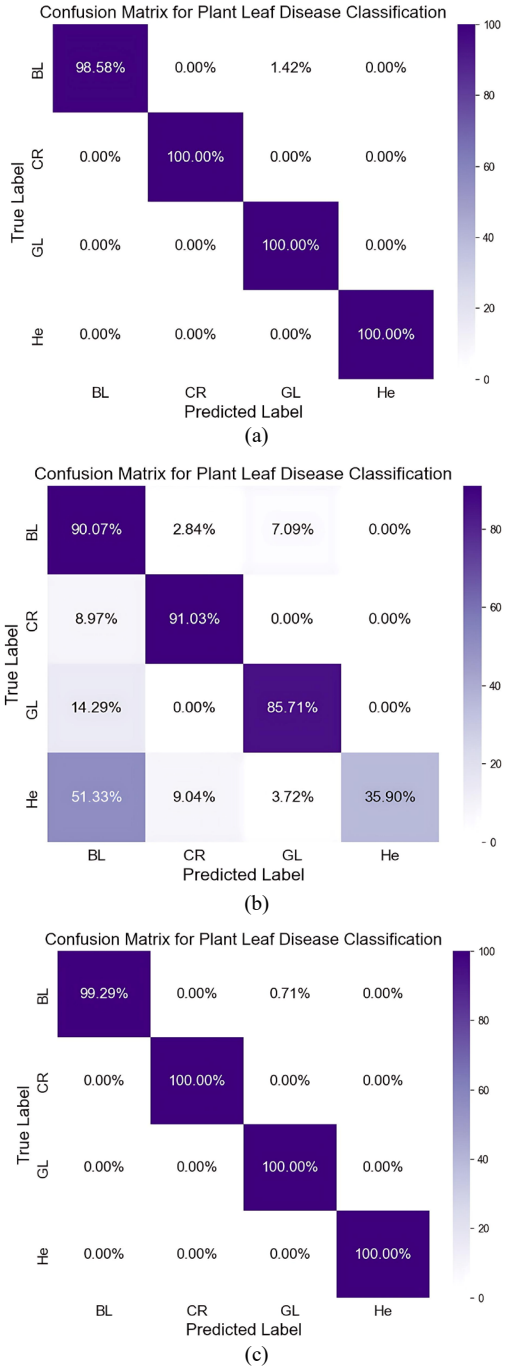

Fig. 5. Classification results for the CLDD. (a) CNN (VGG19). (b) Modified-VGG19. (c) Proposed method.

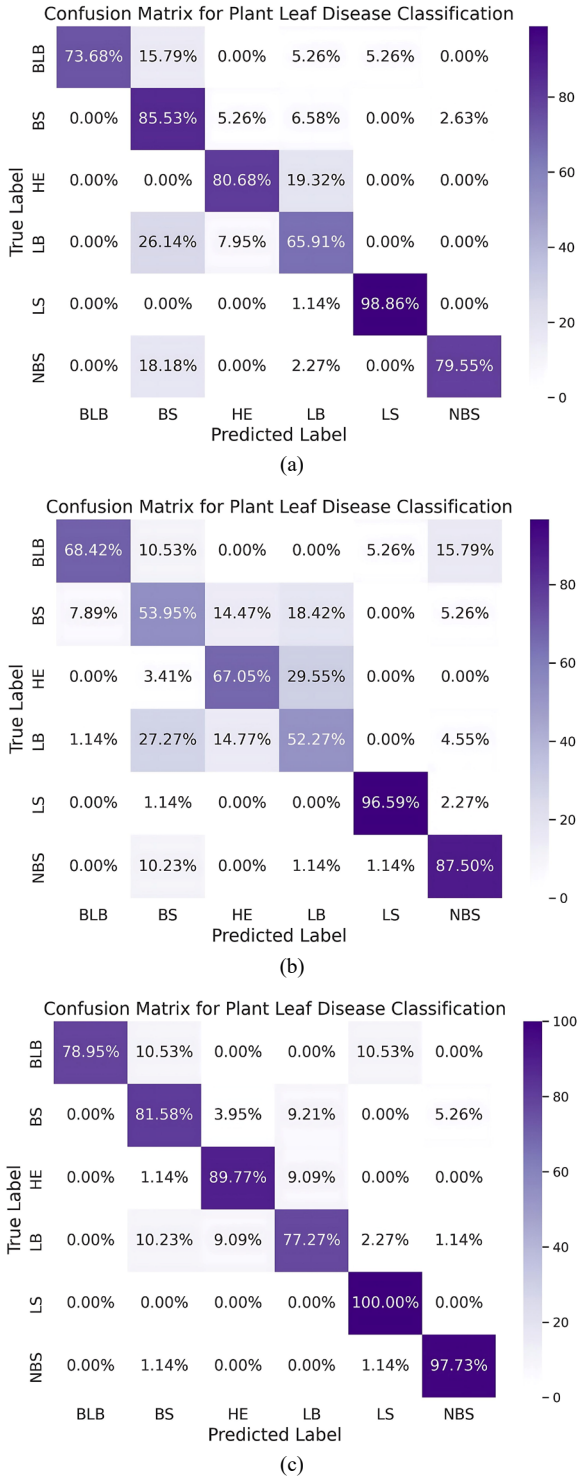

Fig. 6. Classification results for the RLDD. (a) CNN (VGG19). (b) Modified-VGG19. (c) Proposed method.

In Fig. 6, confusion matrix shows how well these models sorted things. Fig. 6 shows the classification results for the RLDD dataset using VGG19, Modified-VGG19, and the proposed method. The proposed method achieves the highest performance, obtaining a perfect True Positive (TP) rate of 100% in the LS class and maintaining minimal misclassification across the remaining classes, demonstrating its strong capability in handling complex leaf disease image patterns. VGG19 achieves the second-highest accuracy, showing robust performance in the LS class with a TP rate of 98.86%. However, its performance decreases when classifying visually similar disease symptoms, as reflected by the reduced TP rates of 79.55% in the NBS class and 73.68% in the BLB class. These results suggest limitations in distinguishing diseases with overlapping visual characteristics.

The Modified-VGG19 model shows moderate improvements in certain categories, achieving 96.59% accuracy for LS and 87.50% for NBS. Nevertheless, substantial confusion remains among several disease classes, particularly BLB, BS, HE, and LB, which reduces the overall classification reliability.

A more detailed inspection of the confusion matrix indicates that misclassification mainly occurs between BLB and LB, which exhibit visually similar lesion patterns. Both diseases often present elongated or irregular brown lesions and diffuse discoloration areas along the leaf surface, making it difficult for the model to clearly separate discriminative features. Additionally, variations in lighting conditions, lesion size, and background noise in field images further increase the difficulty of distinguishing these disease categories. These overlapping visual characteristics contribute to the relatively lower precision values observed for BLB and LB in the RLDD dataset.

Overall, the results demonstrate that the proposed method provides superior robustness and classification capability when dealing with complex plant disease datasets.

Fig. 7 shows the representative sample prediction outputs of the proposed method on both the Corn Leaf Disease Dataset (CLDD) and the Rice Leaf Disease Dataset (RLDD). The visual results include the original input images along with their corresponding ground truth labels, predicted classes, and confidence scores, providing intuitive insight into the model's decision-making behavior. Fig. 7(a), the proposed method demonstrates robust classification performance on CLDD, successfully distinguishing visually similar disease categories such as BL, CR, GL, and He groups. The high confidence values associated with the predicted labels indicate that the model effectively captures discriminative spatial features and disease-specific patterns, even in the presence of complex backgrounds and overlapping leaf regions. Similarly, Fig. 7(b), shows sample predictions on the RLDD dataset, where the model accurately identifies multiple rice leaf disease classes, including BLB, BS, LB, LS, NBS, and HE groups. Despite the visual similarity among certain disease symptoms, the proposed architecture consistently produces reliable predictions with strong confidence, reflecting its ability to learn subtle inter-class differences.

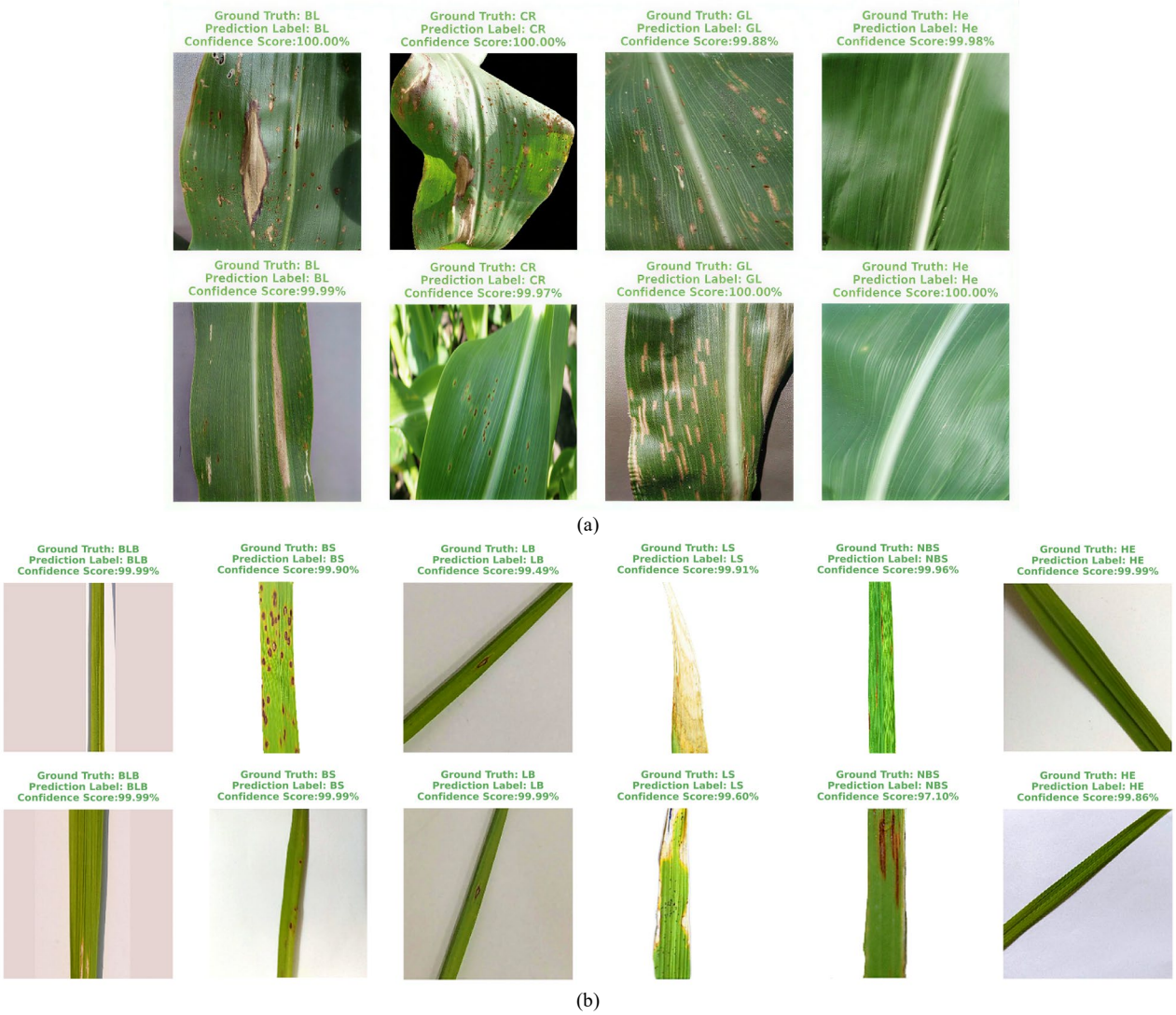

Fig. 7. Sample prediction outputs of the proposed method. (a) Sample predictions on the CLDD. (b) Sample predictions on the RLDD.

C. Comparison Performance

The training results on the CLDD dataset are shows in Fig. 8(a). According to the updated experimental evaluation, ResNet50 and ResNet101 achieved the highest training accuracies, reaching 91.07% and 91.12%, respectively. These results indicate that, when appropriately trained under a consistent preprocessing and optimization scheme, deeper residual architectures are capable of effectively capturing discriminative features for corn leaf disease classification. MobileNetV2 achieved a peak accuracy of 88.94%, demonstrating competitive performance while maintaining a lightweight architecture suitable for resource-constrained environments. Xception followed closely with an accuracy of 88.42%, reflecting its effectiveness in learning spatially separable convolutional features. InceptionV3 attained 87.15%, while DenseNet121 reached 84.43%, showing comparatively lower but stable learning behavior.

Fig. 8(b), shows the training performance of various deep learning models on the RLDD. Based on the updated experimental results, ResNet101 achieved the highest training accuracy at 86.94%, followed closely by ResNet50 with 85.04%. These results demonstrate that deeper residual architectures are capable of learning representative features for rice leaf disease classification when trained under consistent preprocessing and optimization settings. Xception attained an accuracy of 82.28%, reflecting its effectiveness in modeling spatial feature representations through depthwise separable convolutions. MobileNetV2 achieved 80.38%, highlighting its competitive performance despite its lightweight design, which is particularly suitable for deployment on resource-constrained platforms. InceptionV3 showed comparable behavior with an accuracy of 80.13%, indicating stable learning across multiple scales. DenseNet121 recorded a peak accuracy of 74.32%, which, while lower than that of the other architectures, exhibited stable convergence behavior during training. Overall, the revised results confirm that deeper architectures such as ResNet50 and ResNet101 provide improved feature representation capability for RLDD compared to earlier observations, while conventional CNN-based models remain limited in handling visually overlapping disease patterns when compared with the proposed hybrid architecture.

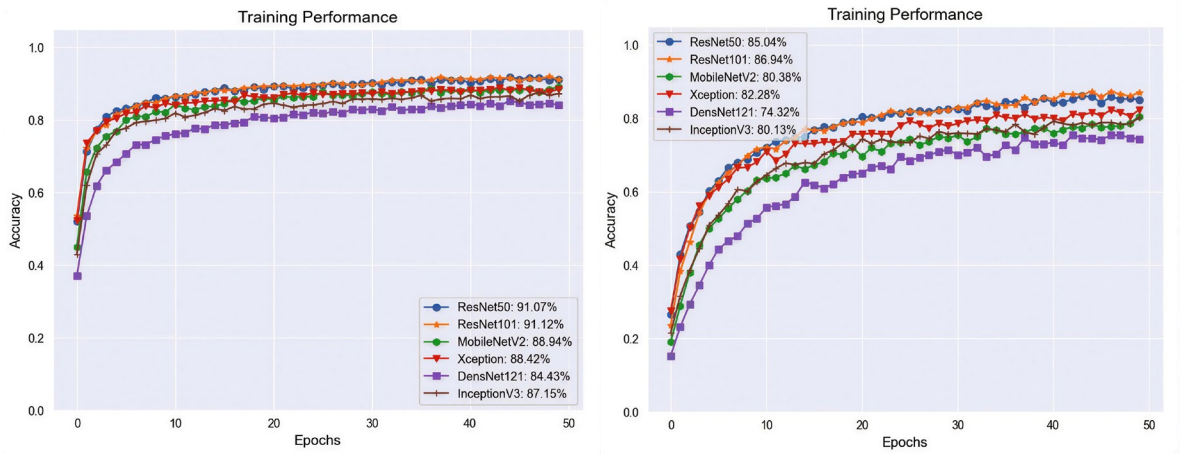

Fig. 8. Training performance of DL with the plant leaf disease. (a) Training performance of DL with the CLDD. (b) Training performance of DL with the RLDD.

In addition to accuracy, precision, recall, and F1-score were also calculated to provide a more comprehensive evaluation of the classification performance and diagnostic reliability of the proposed model and the baseline deep learning algorithms on both CLDD and RLDD datasets. The detailed results are presented in Table II.

Table III reports the classification accuracy obtained from repeated experiments with different random initialization seeds on the CLDD and RLDD datasets to evaluate the stability and statistical reliability of the proposed model. Each experiment was repeated 5 times using different random initialization seeds, and the mean and standard deviation are reported to demonstrate the robustness and stability of the proposed model. Based on these repeated results, a paired t-test was conducted to assess the statistical significance of the performance improvement, and the corresponding results are presented in Table IV. The classification results for leaf diseases are illustrated in Fig. 9 and Fig. 10.

A paired *t*-test was conducted to evaluate whether the performance improvement of the proposed CNN-BiLSTM model over the baseline models is statistically significant. The significance level was set to $\alpha = 0.05$. As shown in Table IV, all *p*-values are significantly lower than 0.05 for both CLDD and RLDD datasets, indicating that the proposed model significantly outperforms the baseline models.

TABLE II. STATE-OF-THE-ART

Data	DL Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
CLDD	VGG19 [22]	99.70	99.55	99.40	99.47
	Modified VGG19 [23]	97.65	97.30	97.10	97.20
	ResNet50 [30]	91.07	90.60	90.30	90.45
	ResNet101 [31]	91.12	90.70	90.40	90.55
	MobileNetV2 [32]	88.94	88.30	88.10	88.20
	Xception [33]	88.42	87.90	87.60	87.75
	DenseNet121 [34]	84.43	83.90	83.50	83.70
	InceptionV3 [35]	87.15	86.70	86.30	86.50
	Proposed method	**99.88**	**99.82**	**99.75**	**99.78**
RLDD	VGG19 [22]	98.53	98.20	98.00	98.10
	Modified VGG19 [23]	96.11	95.70	95.40	95.55
	ResNet50 [30]	85.04	84.60	84.10	84.35
	ResNet101 [31]	86.94	86.40	86.00	86.20
	MobileNetV2 [32]	80.38	79.90	79.50	79.70
	Xception [33]	82.28	81.70	81.30	81.50
	DenseNet121 [34]	74.32	73.80	73.20	73.50
	InceptionV3 [35]	80.13	79.70	79.30	79.50
	Proposed method	**100.00**	**99.92**	**99.85**	**99.88**

TABLE III. ACCURACY OBTAINED FROM REPEATED EXPERIMENTS WITH DIFFERENT RANDOM SEEDS ON CLDD AND RLDD DATASETS (UNITE: %)

Run	Proposed (CLDD)	VGG19 (CLDD)	ResNet101 (CLDD)	Xception (CLDD)	Proposed (RLDD)	VGG19 (RLDD)	ResNet101 (RLDD)	Xception (RLDD)
1	**99.70**	97.65	91.12	88.94	**100.00**	96.11	86.94	80.38
2	**99.68**	97.50	91.05	88.70	**99.98**	96.05	86.80	80.20
3	**99.72**	97.55	90.98	88.60	**100.00**	96.00	86.75	80.10
4	**99.69**	97.60	91.10	88.75	**99.99**	96.08	86.88	80.25
5	**99.71**	97.58	91.08	88.80	**100.00**	96.10	86.92	80.30
Mean ± SD	**99.70 ± 0.015**	97.58 ± 0.06	91.07 ± 0.05	88.76 ± 0.13	**99.99 ± 0.01**	96.07 ± 0.04	86.86 ± 0.07	80.25 ± 0.10

TABLE IV. STATISTICAL SIGNIFICANCE TEST RESULTS USING PAIRED T-TEST (A = 0.05)

Dataset	Model Comparison	*t*-value	*p*-Value	Statistical Significance
CLDD	Proposed vs VGG19	87.00	1.05×10^{-7}	Significant
	Proposed vs ResNet101	304.12	7.01×10^{-10}	Significant
	Proposed vs Xception	188.66	4.74×10^{-9}	Significant
RLDD	Proposed vs VGG19	199.83	3.76×10^{-9}	Significant
	Proposed vs ResNet101	375.16	3.03×10^{-10}	Significant
	Proposed vs Xception	426.49	1.81×10^{-10}	Significant

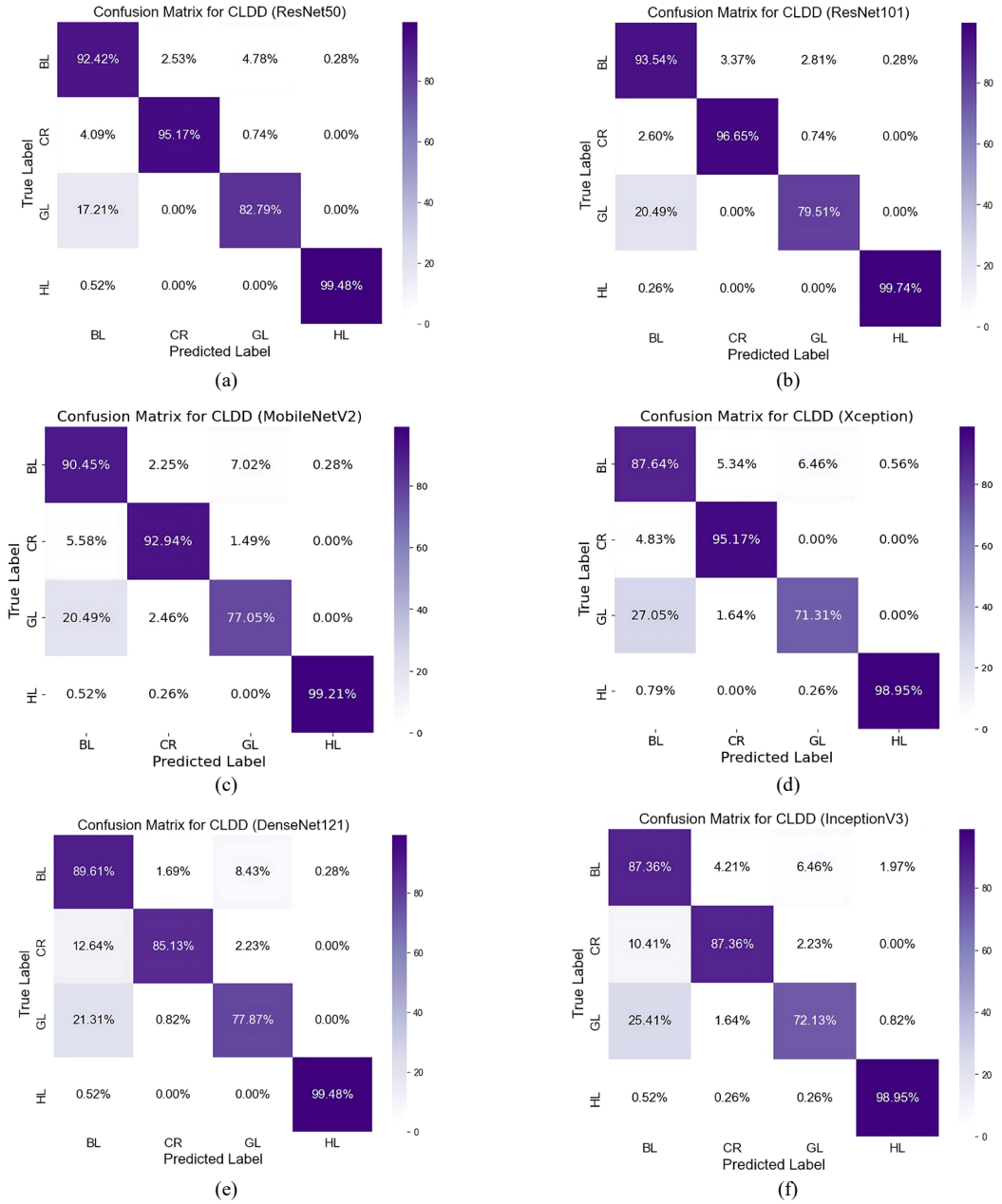

Fig. 9. Classification results of DL with the CLDD. (a) ResNet50. (b) ResNet101. (c) MobileNetV2. (d) Xception. (e) DenseNet121. (f) InceptionV3.

Fig. 9 shows the confusion matrices of different deep learning models evaluated on the CLDD. Based on the revised experimental results, all models demonstrate improved classification behavior compared to earlier observations, indicating more stable and effective learning across disease categories. ResNet50 and ResNet101 exhibit substantially improved performance, showing consistent classification accuracy across most classes without severe misclassification in previously challenging categories. In particular, the residual learning mechanism enables these models to better preserve discriminative feature representations for visually similar disease patterns, resulting in more balanced class-wise predictions. MobileNetV2, Xception, and InceptionV3 maintain strong classification capability for healthy leaves and major disease classes, demonstrating reliable recognition of dominant visual characteristics. Their confusion matrices indicate reduced ambiguity between disease categories,

although minor confusion remains in classes with overlapping symptom appearance. DenseNet121 shows stable classification performance with effective recognition of both healthy and diseased samples, benefiting from dense feature reuse and efficient gradient propagation.

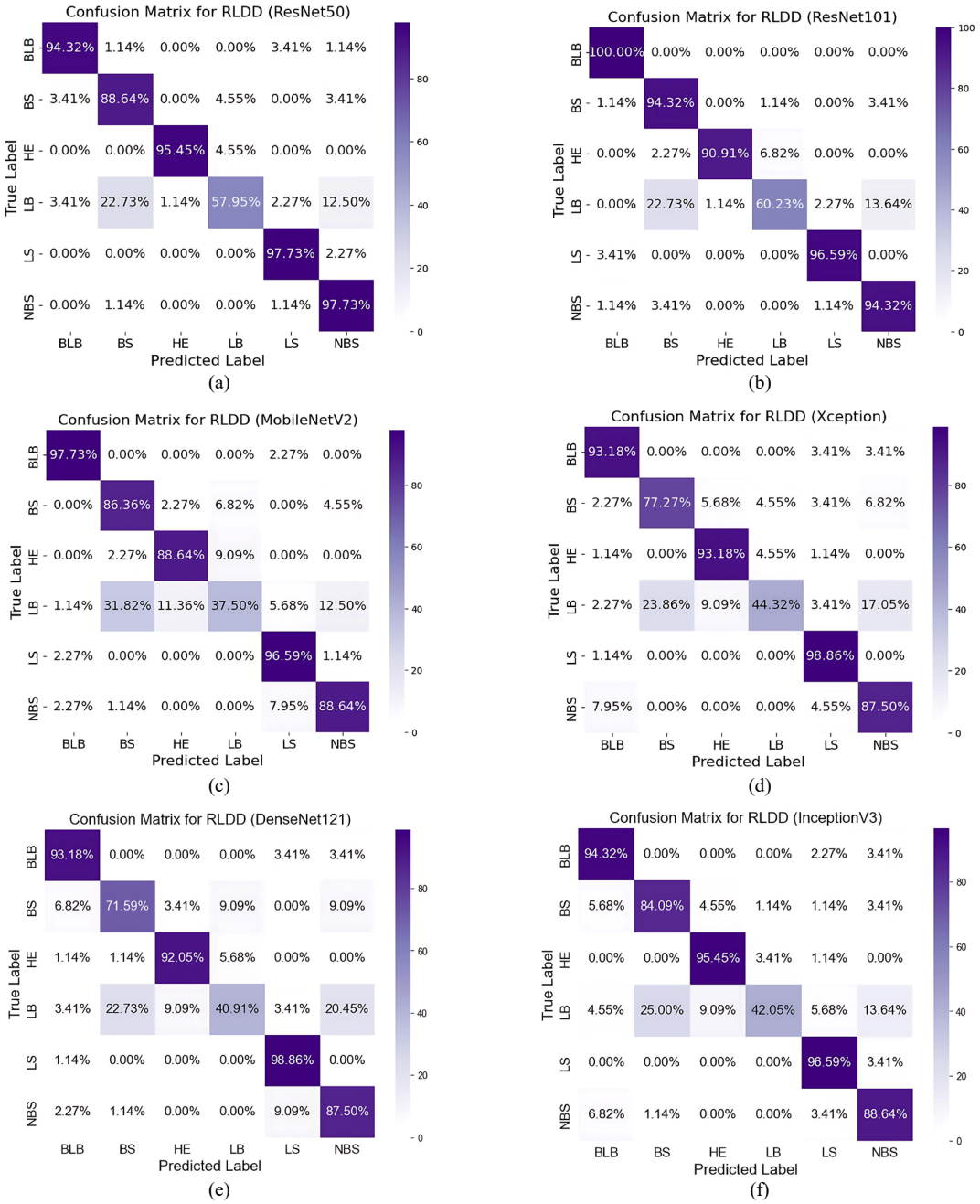

Fig. 10. Classification results of DL with the RLDD. (a) ResNet50. (b) ResNet101. (c) MobileNetV2. (d) Xception. (e) DenseNet121. (f) InceptionV3.

Fig. 10 shows the confusion matrices obtained from different deep learning models evaluated on the RLDD. Based on the revised experimental results, all models exhibit noticeably improved class-wise classification behavior compared to earlier observations, indicating more stable learning and reduced misclassification across rice leaf disease categories. ResNet50 and ResNet101 demonstrate significant performance improvement, showing balanced true positive predictions across most disease classes without severe failure cases in previously challenging categories. The residual learning mechanism allows these models to better preserve discriminative features, enabling more reliable differentiation among visually similar rice leaf diseases. MobileNetV2, Xception, and InceptionV3 maintain strong recognition capability for dominant disease classes, with their confusion matrices indicating high true positive rates and reduced inter-class confusion. These architectures effectively capture salient spatial patterns while maintaining stable generalization performance on RLDD. DenseNet121 exhibits consistent classification performance across multiple classes, benefiting from

dense feature reuse and efficient gradient propagation. Although minor confusion remains among disease categories with overlapping visual symptoms.

To further understand the model behavior when dealing with overlapping disease symptoms, we analyzed the classification patterns observed in the confusion matrix. The results suggest that the proposed model primarily relies on distinctive lesion characteristics such as lesion shape, boundary irregularity, and color intensity when distinguishing different disease categories. In many correctly classified cases with overlapping symptoms, the lesions still contain subtle discriminative patterns that allow the model to identify the disease class accurately.

However, in some misclassified samples, particularly between BLB and LB, the lesion regions exhibit highly similar visual characteristics, including elongated brown spots and diffuse discoloration patterns. These similarities may reduce the discriminative power of extracted features, leading to feature ambiguity during classification. In addition, variations in lighting conditions, lesion size, and background noise in field images may further increase the difficulty of distinguishing these disease categories.

This analysis indicates that while the proposed model is capable of capturing relevant lesion features, classification errors may still occur when different diseases present highly overlapping visual patterns.

IV. Discussion

Plant leaf diseases are challenging to classify due to substantial visual variability among symptoms, which can lead to misdiagnosis. This study introduces a hybrid DL model integrating SE-block, CN, BiLSTM, and an AM to emphasize critical features, preserve spatial relationships, and enhance feature representation.

Each component of the proposed architecture is designed to address a specific limitation in conventional CNN-based disease classification. SE-block enhance channel-wise feature recalibration, enabling the network to focus on disease-relevant visual patterns while suppressing redundant information. CN explicitly preserve spatial relationships among local features, which is crucial for capturing lesion structures and spatial configurations of plant diseases. BiLSTM is incorporated to model contextual dependencies among extracted features, allowing the network to learn complex interactions beyond local representations. The attention mechanism further refines this process by highlighting the most discriminative feature sequences and reducing the influence of irrelevant or noisy information.

The experimental results demonstrate that the proposed architecture outperforms existing approaches, achieving improved classification accuracy and exhibiting a high degree of robustness. Building on the work of Suksukont *et al.* [24] and Krstinić *et al.* [29], this design extends their functionalities by incorporating multi-level feature integration.

The proposed method surpasses conventional CNN-based models, achieving accuracies of 99.88% on the CLDD dataset and 100% on the RLDD dataset with stable learning behavior. In comparison, VGG19 attains 99.70%, while Modified-VGG19 achieves 97.65%, both exhibiting reduced stability during mid-training. These findings indicate that SE-Block contributes to effective feature extraction, while CN facilitates the preservation of spatial relationships. Class-wise evaluation further shows that the proposed method achieves superior accuracy in the case of CLDD, reaching 99.29% for blight and 100% for the remaining classes, whereas VGG19 and Modified-VGG19 exhibit limitations, particularly with regard to blight and healthy categories. Similar trends are observed on the RLDD, where the proposed method achieves near-perfect performance, while baseline models encounter difficulties in visually overlapping classes.

These results further highlight that CN enhances spatial structure representation, while BiLSTM improves sequential feature learning, rendering the proposed hybrid architecture particularly effective for diagnosing complex and diverse plant leaf diseases, with the additional computational cost primarily incurred during training rather than inference.

Despite the promising performance, several limitations should be acknowledged. The proposed model is trained and evaluated on specific datasets, and its performance may be influenced by dataset dependency and environmental variability. Differences in illumination, background complexity, and disease appearance across crops and geographic regions may pose challenges to generalization. Addressing these limitations through larger multi-crop datasets and evaluations under diverse field conditions will be considered in future work.

V. Conclusion

This study presents an enhanced hybrid DL architecture for plant leaf disease classification by integrating CN, SE-block, BiLSTM, and an AM. Each component is designed to address specific challenges in disease recognition, including spatial structure preservation, channel-wise feature emphasis, contextual dependency modeling, and discriminative feature selection. The combination of AM and GAP, further improves spatial representation and increases the robustness of the learned features.

In this study, classification accuracy is used as the primary evaluation metric to ensure consistency with related works using the same datasets. In addition, class-wise performance analysis is provided to examine the classification behavior across different disease categories. The proposed method achieved training accuracies of 99.88% on the CLDD dataset and 100% on the RLDD dataset. During testing, class-wise accuracies on CLDD reached 99.29% for blight and 100% for the remaining categories. On the RLDD dataset, the model attained accuracies of 78.95% for bacterial leaf blight, 85.53% for brown spot, 89.77% for healthy samples, 77.27% for leaf blast, 100% for leaf scald, and 97.73% for narrow brown spot. To mitigate potential overfitting despite the high training accuracy, data augmentation, regularization strategies, and validation monitoring were employed during training. Overall, the experimental results demonstrate that the proposed hybrid architecture provides superior and stable performance compared to existing

approaches, highlighting its effectiveness for accurate plant leaf disease diagnosis under diverse visual and environmental conditions.

Future work will focus on developing neural network architectures capable of handling diverse and complex datasets while remaining suitable for real-time deployment. Particular attention will be given to reducing model complexity, improving inference efficiency, and evaluating the computational feasibility of deploying the proposed model on mobile and edge devices. In addition, model compression techniques, such as pruning and quantization, will be explored to further reduce computational requirements and enable efficient deployment. Future studies will also investigate the robustness of the model under real-field conditions, including variations in illumination and complex backgrounds. These advancements aim to support practical mobile-based disease diagnosis systems that assist farmers in making timely and informed crop management decisions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, Methodology: AS; Investigation, Resources: AS and EN; Visualization: AS; Writing and editing-original draft: AS and EN; Validation: AS; Review: AS and EN; Supervision and Lead: EN. All authors had approved the final version.

ACKNOWLEDGMENT

The authors would like to sincerely thank the reviewers for their valuable comments and constructive suggestions, which have significantly improved the quality of this manuscript.

REFERENCES

[1] J. B. Ristaino, P. K. Anderson, D. P. Bebber *et al.*, "The persistent threat of emerging plant disease pandemics to global food security," in *Proc. the National Academy of Sciences*, 2021.

[2] R. Karthickmanoj and T. Sasilatha, "Development of plant disease detection for smart agriculture," *Multimedia Tools and Applications*, vol. 83, pp. 54391–54410, 2024.

[3] Q. Heng, S. Yu, and Y. Zhang, "A new AI-based approach for automatic identification of tea leaf disease using deep neural network based on hybrid pooling," *Heliyon*, vol. 10, no. 5, e26465, 2024.

[4] C. Sarkar, D. Gupta, U. Gupta *et al.*, "Leaf disease detection using machine learning and deep learning: Review and challenges," *Applied Soft Computing*, vol. 145, 110534, 2023.

[5] V. Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques," *Information Processing in Agriculture*, vol. 4, no. 1, pp. 41–49, 2017.

[6] X. Liu, W. Min, S. Mei *et al.*, "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.

[7] K. Mahadevan, A. Punitha, and J. Suresh, "Automatic recognition of rice plant leaf diseases detection using deep neural network with improved threshold neural network," *e-Prime-Advances in Electrical Engineering Electronics and Energy*, vol. 8, 100534, 2024.

[8] S. Thaseentaj and S. S. Ilango, "Deep convolutional neural networks for South Indian mango leaf disease detection and classification," *Computers Materials and Continua*, vol. 77, no. 3, pp. 3593–3618, 2023.

[9] N. Lokhande, V. Thool, and P. Vikhe, "Comparative analysis of different plant leaf disease classification and detection using CNN," in *Proc. Int. Conf. on Recent Innovation in Smart and Sustainable Technology (ICRISST)*, 2024, pp. 1–6.

[10] K. Kaviarasu, J. P. Immanuvel, R. Sanjay *et al.*, "Revolutionizing crop health: CNN and XGBoost for accurate disease detection in smart agriculture," in *Proc. Int. Conf. Advanced Computing Technologies (ICoACT)*, 2025, pp. 1–6.

[11] G. Shanmugam, D. Balusamy, K. R. KavinSubash *et al.*, "Sustainable agriculture with advanced plant disease detection using YOLOv11 and XAI," in *Proc. IEEE 14th International Conf. on Communication Systems and Network Technologies (CSNT)*, 2025, pp. 598–603.

[12] S. G. Paul, A. A. Biswas, A. Saha *et al.*, "A real-time application-based convolutional neural network approach for tomato leaf disease classification," *Array*, vol. 19, 100313, 2023.

[13] A. Hussain and P. B. Srikaanth, "VGG19 enhanced convolutional neural network for paddy leaf disease detection," in *Proc. 3rd International Conf. on Pervasive Computing and Social Networking (ICPCSN)*, 2023, pp. 840–844.

[14] A. Kaur, V. Kukreja, P. Tiwari *et al.*, "An efficient deep learning-based VGG19 approach for rice leaf disease classification," in *Proc. IEEE 9th International Conf. for Convergence in Technology (I2CT)*, 2024, pp. 1–6.

[15] A. Hessane, A. E. Youssefi, Y. Farhaoui *et al.*, "A machine learning-based framework for a stage-wise classification of date palm white scale disease," *Big Data Min. Anal.*, vol. 6, no. 3, pp. 263–272, 2023.

[16] K. Taji, A. Sohail, T. Shahzad *et al.*, "An ensemble hybrid framework: A comparative analysis of metaheuristic algorithms for ensemble hybrid CNN features for plants disease classification," *IEEE Access*, vol. 12, pp. 61886–61906, 2024.

[17] S. Muthusamy and S. P. Ramu, "IncepV3Dense: Deep ensemble-based average learning strategy for identification of micro-nutrient deficiency in banana crop," *IEEE Access*, vol. 12, pp. 73779–73792, 2024.

[18] X. Zhang, Y. Mao, Q. Yang *et al.*, "A plant leaf disease image classification method integrating capsule network and Residual network," *IEEE Access*, vol. 12, pp. 44573–44585, 2024.

[19] A. Prommakhot, J. Onshaunjit, W. Ooppakaew *et al.*, "Hybrid CNN and transformer-based sequential learning techniques for plant disease classification," *IEEE Access*, vol. 13, pp. 122876–122887, 2025.

[20] A. Suksukont and E. Naowanich, "Designing a squeeze-and-excitation-capsule BiLSTM-transformer for plant leaf disease recognition" *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 14, no. 6, pp. 5069–5080, 2025.

[21] G. Geetharamani and J. A. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," *Computers & Electrical Engineering*, vol. 76, pp. 323–338, 2019.

[22] H. B. Prajapati, J. P. Shah, and V. K. Dabhi, "Detection and classification of rice plant diseases," *Intelligent Decision Technologies*, vol. 11, no. 3, pp. 357–373, 2017.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2015.

[24] A. Suksukont and E. Naowanich, "Integration of capsule network with CNN for plant leaf disease classification," *APHEIT International Journal of Interdisciplinary Social Sciences and Technology*, vol. 14, no. 1, pp. 48–58, 2025.

[25] K. He, X. Zhang, S. Ren *et al.*, "Deep residual learning for image recognition," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[26] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," arXiv preprint arXiv:1710.09829, 2017. d

[27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[28] Y. Tan and X. Ding, "Split-attention CNN and self-attention with RoPE and GCN for voice activity detection," *IEEE Access*, vol. 12, pp. 156673–156682, 2024.

[29] D. Krstinić, A. K. Skelin, I. Slapničar *et al.*, "Multi-label confusion tensor," *IEEE Access*, vol. 12, pp. 9860–9870, 2024.

[30] M. M. Adnan, M. S. M. Rahim, A. R. Khan *et al.*, "Automated image annotation with novel features based on deep ResNet50-SLT," *IEEE Access*, vol. 11, pp. 40258–40277, 2023.
[31] P. K. Sethy, L. Korada, S. K. Behera *et al.*, "Maximizing steel slice defect detection: Integrating ResNet101 deep features with SVM via bayesian optimization," *Systems and Soft Computing*, vol. 6, 200170, 2024.
[32] M. Akay, Y. Du, C. L. Sershen *et al.*, "Deep learning classification of systemic sclerosis skin using the MobileNetV2 model," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 2, pp. 104–110, 2021.
[33] S. Ramya, R. I. Minu, and K. T. Magesh, "Xception spiking fractional neural network for oral squamous cell carcinoma classification based on histopathological images," *IEEE Access*, vol. 13, pp. 98574–98590, 2025.
[34] G. Huang, Z. Liu, L. Maaten *et al.*, "Densely connected convolutional networks," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269.
[35] C. Szegedy, V. Vanhoucke, S. Ioffe *et al.*, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.