RSA-DeRefNet: A Hybrid Residual-Dense Network with Regularization Self-Attention

Aditya Nayak
Electronics and Communication Engg.
NIT Rourkela
Rourkela, India.
nayakadityanitr@gmail.com

Mohit Ranjan Naik
Electronics and Communication Engg.
NIT Rourkela
Rourkela, India.
mohitranjannaiknitr@gmail.com

Ayush Kumar Samal
Electronics and Communication Engg.
NIT Rourkela
Rourkela, India.
ayush.samalnitr@gmail.com

Prof. Samit Ari
Electronics and Communication Engg.
NIT Rourkela
Rourkela, India.
samit@nitrkl.ac.in

***Abstract*—The rise of plant diseases poses a major threat to global food security. Effective automated detection systems are essential for timely intervention, highlighting the need for reliable systems that can detect problems early. To meet this challenge, we introduce RSA-DerefNet, a new hybrid deep learning architecture aimed at classifying plant diseases based on images. Our model combines two effective network types: residual networks, which help train very deep models by addressing gradient issues, and densenets, which improve information flow through feature reuse. The central part of our approach is a custom Regularized Self-Attention (RSA) module integrated into both network paths. This feature allows the model to focus on important visual signs linked to diseases, like lesion patterns or discoloration, while reducing irrelevant background noise. Evaluating RSA-DerefNet on a wide-ranging dataset of crop diseases shows that it performs significantly better, confirming the value of our architectural design for achieving high-precision classification.**

I. Introduction

Maintaining agricultural yield is essential for global food security, yet plant diseases caused by fungi, bacteria, viruses, and nutritional deficiencies pose ongoing threats, leading to substantial financial losses and compromised food availability. Traditional disease detection relies on manual inspection by agronomists—a method that is subjective, prone to error, and cannot scale for large commercial farms. Early symptoms are often subtle, delaying intervention and allowing diseases to spread. Initial automation efforts used classical computer vision techniques, extracting hand-crafted features like HOG, SIFT, and texture descriptors, then training classifiers such as SVMs or k-NN. However, these features are fragile against real-world variations in lighting, noise, and leaf orientation, limiting their effectiveness in diverse field conditions. Deep learning, particularly Convolutional Neural Networks (CNNs), transformed this field by learning hierarchical features directly from raw images. However, challenges remain: many models are computationally expensive, unsuitable for resource-limited devices like drones or smartphones, and act as ”black boxes” that offer little insight into their decision-making process. This paper proposes RSA-DerefNet, a hybrid deep learning architecture addressing these limitations. Our model combines residual networks, which use skip connections to mitigate vanishing gradients, with densenets, which promote feature reuse and efficient information flow. The key contribution is a custom Regularized Self-Attention (RSA) module integrated into both network streams, enabling the model to focus on disease-relevant visual features while suppressing background noise. This attention mechanism also provides interpretability through visualizable attention maps—a crucial step toward trustworthy agricultural AI systems.

II. Related Works

The field of automated plant disease detection has evolved significantly, moving from classic image processing methods to advanced deep learning models. A review of existing literature shows clear progress in methodologies, each addressing the shortcomings of previous approaches. This section analyzes key contributions in this area.

A. Traditional Computer Vision Approaches

Early research focused on rule-based systems that use hand-crafted features. These methods tried to mimic human visual inspection by analyzing basic image properties. H. Al-Hiary et al. (2011) developed a method centered around image segmentation. They converted the image into a different color space to better isolate disease areas. K-means clustering was then applied to segment the image, separating infected regions from healthy parts of the leaf. After segmentation, features like texture and color were extracted from these sections and used for classification.

B. The Rise of Deep Learning

Krizhevsky et al. (2012) introduced AlexNet, demonstrating end-to-end deep CNN learning with ReLU activation and

TABLE I
SUMMARY OF LITERATURE SURVEY ON PLANT DISEASE DETECTION TECHNIQUES

S.No.	Reference	Year	Method/Architecture	Key Contribution
1	H. Al-Hiary et al.	2011	K-means Clustering	Image segmentation using color space transformation for disease region isolation
2	Krizhevsky et al.	2012	AlexNet	Deep CNNs for large-scale image classification; introduced ReLU activation and dropout
3	P.R. Shinde & M.V. Bhise	2015	Image Processing Pipeline	Multi-step processing with thresholding and morphological operations for cotton leaf diseases
4	He et al.	2016	ResNet	Skip connections to address vanishing gradient; enabled training of very deep networks
5	J.P. Sahoo et al.	2023	DeReFNet	Dual-stream dense residual fusion network with GFA and SF streams for feature extraction
6	A. Das et al.	2025	LeafDisDiff	Diffusion-based framework with U-Net architecture for plant leaf disease detection

dropout. He et al. (2016) proposed ResNet, using skip connections to address vanishing gradients in very deep networks. Das et al. (2025) introduced LeafDisDiff, a diffusion-based U-Net framework for plant disease detection. Sahoo et al. (2023) proposed DeReFNet, a dual-stream architecture combining a residual stream for global features with a dense stream for spatial details, merged via feature concatenation. We adapt this architecture for plant disease classification.

C. Gaps in Existing Research and Our Contribution

Existing approaches have three key limitations: (1) single-stream architectures miss complementary feature representations; (2) local convolutions cannot model long-range spatial dependencies relevant to scattered disease patterns; (3) over-reliance on controlled PlantVillage data limits generalization.

RSA-DeRefNet addresses these through: (1) dual-stream residual-dense architecture with feature concatenation; (2) Regularized Self-Attention for global spatial reasoning; (3) cross-dataset validation on PlantVillage and Bangladeshi Crops.

Ablation results: removing RSA causes 4.2% accuracy drop; concatenation outperforms other fusion strategies by 2.8%.

III. PROPOSED METHODOLOGY

A. Preprocessing

Input images were resized to 224×224 pixels, normalized to $[0, 1]$, and augmented using random rotation (±25), horizontal/vertical flips, and contrast adjustments. Labels were one-hot encoded, and the dataset was split 80-20 with stratified sampling.

B. Training Pipeline

Data loading was optimized using TensorFlow's `tf.data` API with caching, prefetching, and parallel mapping to maximize GPU utilization.

C. Residual Stream

Residual Block Formulation

The residual stream is based on the principle of identity mapping from ResNet. Formally, for an input feature map x, a residual block computes:

$$y = \sigma\big(\text{BN}(W_2 * \sigma(\text{BN}(W_1 * x))) + \mathcal{F}(x)\big) \tag{1}$$

where W1,W2 are convolutional filters, BN denotes batch normalization, σ represents the ReLU activation, F(x) is the shortcut connection (identity or projection).

Residual Block Grouping

Residual blocks are organized in groups. Each group begins with a strided block (stride = 2) for spatial downsampling, followed by multiple stride-1 residual blocks to enhance feature abstraction. Three groups were used, with increasing filter sizes (64, 128, 256). After each group, an RSA module was applied to emphasize salient dependencies across the feature maps.

D. Dense Stream

The dense pathway is inspired by DenseNet, where each layer receives feature maps from all preceding layers:

$$x_\ell = H_\ell([x_0, x_1, \ldots, x_{\ell-1}]) \tag{2}$$

where $[\cdot]$ denotes concatenation and H_ℓ is a composite BN-ReLU-Conv operation. This design encourages feature reuse and improves parameter efficiency. Between dense blocks, transition layers apply batch normalization, 1×1 convolution for channel reduction, and average pooling for spatial downsampling. After each dense block, an RSA module is applied to capture long-range dependencies.

E. Regularised Self-Attention (RSA)

Motivation In plant pathology, disease symptoms often manifest as subtle texture patterns, small lesions, or irregular discolorations. Conventional CNNs rely primarily on local receptive fields, which may fail to capture long-range spatial dependencies. To overcome this, we introduce a Residual Self-Attention (RSA) module, inspired by non-local neural operations.

Mathematical Formulation

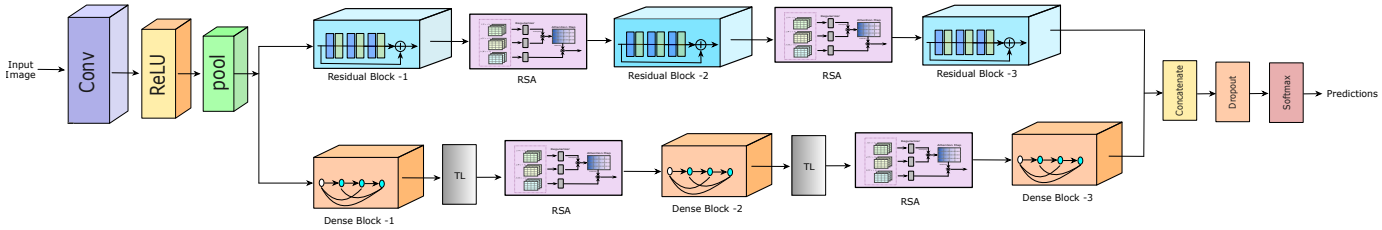

Fig. 1. Proposed Regularised DeRefNet

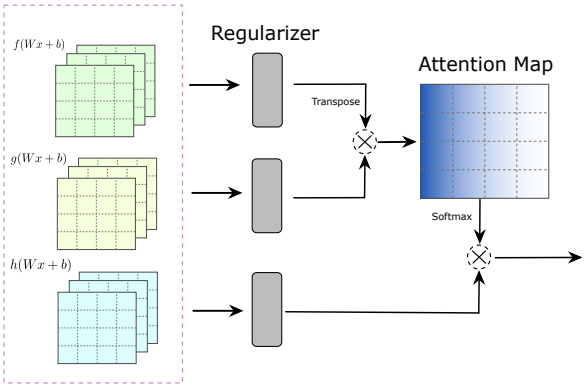

Fig. 2. RSA component block

Given an input feature map $X \in \mathbb{R}^{H \times W \times C}$,

$$f = W_f * X, \quad g = W_g * X, \quad h = W_h * X$$

The attention scores are computed as:

$$S = \text{softmax}\left(\frac{fg^T}{\sqrt{C}}\right) \tag{3}$$

The attention-weighted output is:

$$O = Sh$$

Finally, residual connection is applied:

$$Y = O + X$$

F. Feature Fusion and Classification

Both streams undergo Global Average Pooling (GAP) to reduce spatial dimensions into compact feature vectors. The residual feature vector gr. and the dense feature vector gd. are concatenated:

$$F = [g_r, g_d] \tag{4}$$

A Dropout layer (p = 0.5) is applied to mitigate overfitting, followed by a fully connected layer with softmax activation for classification:

$$\hat{y} = \text{softmax}(W_f F + b) \tag{5}$$

where yhat denotes the predicted probability distribution over the 15 crop disease classes.

G. Optimization Strategy

Mixed Precision Training

To accelerate training without sacrificing accuracy, we employed mixed-precision policy, combining 16-bit floating point operations for speed with 32-bit accumulations for numerical stability. This significantly reduced memory footprint, allowing larger batch sizes and deeper model evaluation on limited GPU resources.

Optimizer

The AdamW optimizer was chosen for its adaptive learning rate and decoupled weight decay, which improves generalization. The optimizer was configured with:

- The initial learning rate was set to 3×10^{-4}.
- The weight decay was set to 1×10^{-5}.
- Gradient clipping: $\|g\|_2 \leq 1.0$ to stabilize updates.

Learning Rate Schedule

A Cosine Decay schedule was employed:

$$\eta_t = \eta_0 \cdot \frac{1}{2}\left(1 + \cos\left(\frac{\pi t}{T}\right)\right) \tag{6}$$

where η_t is the learning rate at step t, η_0 is the initial learning rate, and T is the total number of training steps. This cyclic reduction prevents premature convergence and allows the optimizer to explore flatter minima.

H. Loss Function

The model was trained using categorical cross-entropy with label smoothing ($\epsilon = 0.1$) to prevent overconfident predictions and improve calibration.

IV. Experimental Setup and discussion

In this section, experiments are carried out to validate the datasets. Various parameters study, and performance analysis of the proposed CNN.

A. Datasets

Two benchmark datasets were used for evaluation:

Bangladeshi Crops Disease Dataset: Contains 18,450 leaf images across 10 crops and 12 disease categories plus healthy samples. Split: 70-15-15 (train/val/test).

PlantVillage Dataset: Contains 54,000 images covering 14 crop species and 38 disease categories, collected under controlled conditions. Split: 80-10-10.

TABLE II
DATASETS

serial Model	Name of dataset		
	Benchmarked Dataset	***classes***	***images***
RSA-Derefnet	Bangladeshi Crops	15	22000
	Plant Village dataset	38	54000

B. Experimental Setup

The proposed RSA-DeRefNet was implemented in **Tensorflow** with CUDA 11.8 support.

All input images were resized to 224×224 pixels and normalized to the range $[0, 1]$. To improve generalization performance, data augmentation techniques including *random rotation* ($\pm 25^\circ$), *horizontal and vertical flips*, *zoom-in/out scaling*, and *contrast adjustment* were applied.

Training was performed using the **AdamW optimizer** with an initial learning rate of 3×10^{-4}, a batch size of 32, and a weight decay of 1×10^{-5}. A **cosine annealing scheduler** was applied to dynamically adjust the learning rate, and the model was trained for 30 epochs with **early stopping** based on validation loss.

Evaluation was conducted using multiple performance metrics, including **overall accuracy**, **precision**, **recall**, and **F1-score**. Furthermore, a **confusion matrix** was generated to provide class-wise insights into the recognition performance across different crop disease categories.

C. Validation Methods

To ensure robustness and generalization, the dataset was partitioned into training, validation, and test sets using an 80:10:10 ratio with stratified sampling to preserve class distribution across splits. The validation set was used exclusively for hyperparameter tuning and early stopping, while the test set was reserved for final evaluation.

Model performance was assessed using standard classification metrics: overall accuracy, precision, recall, and F1-score.

TABLE III
Training Hyperparameters for RSA-DeRefNet

Parameter	Value
Optimizer	AdamW
Learning Rate	3×10^{-4}
Batch Size	32
Weight Decay	1×10^{-5}
Epochs	30
Learning Rate Scheduler	Cosine Annealing
Early Stopping Patience	10
Input Image Size	224×224

A confusion matrix was generated for class-wise analysis, which is particularly relevant in agricultural applications where visually similar diseases can be challenging to distinguish. Additionally, 5-fold cross-validation was performed to verify model consistency across different data partitions and minimize overfitting risk.

V. Experimental Results and Analysis

This section presents the empirical evaluation of the proposed RSA-DeRefNet architecture on two benchmark datasets: the Bangladeshi Crops Disease dataset and the Plant Village dataset. . The model was trained for 30 epochs with a batch size of 32, using the Adam optimizer with an initial learning rate of 1×10^{-4}. Accuracy and cross-entropy loss were recorded for both training and validation sets at each epoch. The corresponding learning curves are shown in Figures 4 and 3.

TABLE IV
Results Obtained using various deep learning models on Plant Village dataset

Models	CNN	ResNet-50	Efficient-B3	VGG19	RSA-Derefnet
Accuracy	94.00	97.00	98.80	93.82	99.32

A. Bangladeshi Crops Disease Dataset

Figure 4 depicts the training and validation curves for the Bangladeshi Crops Disease dataset. Several key observations can be drawn:

- **Accuracy Trends:** The training accuracy (blue curve) shows a rapid increase from 76% in the first epoch to nearly 90% by the third epoch. Validation accuracy (green curve) follows a similar trajectory, reaching above 92% within the first five epochs. After epoch 10, both training and validation accuracy gradually saturate, stabilizing around 98% and 97%, respectively. The convergence of the two curves indicates that the model generalizes well, without significant overfitting.
- **Loss Trends:** The training loss decreases sharply from an initial value of 1.3 to below 0.8 within the first five epochs. Validation loss also follows a consistent downward trajectory, starting at 1.2 and reaching below 0.7 around epoch 15. Beyond epoch 20, both curves flatten out near 0.65, confirming convergence. Importantly, the close proximity of training and validation loss suggests minimal variance, reflecting stable learning dynamics.

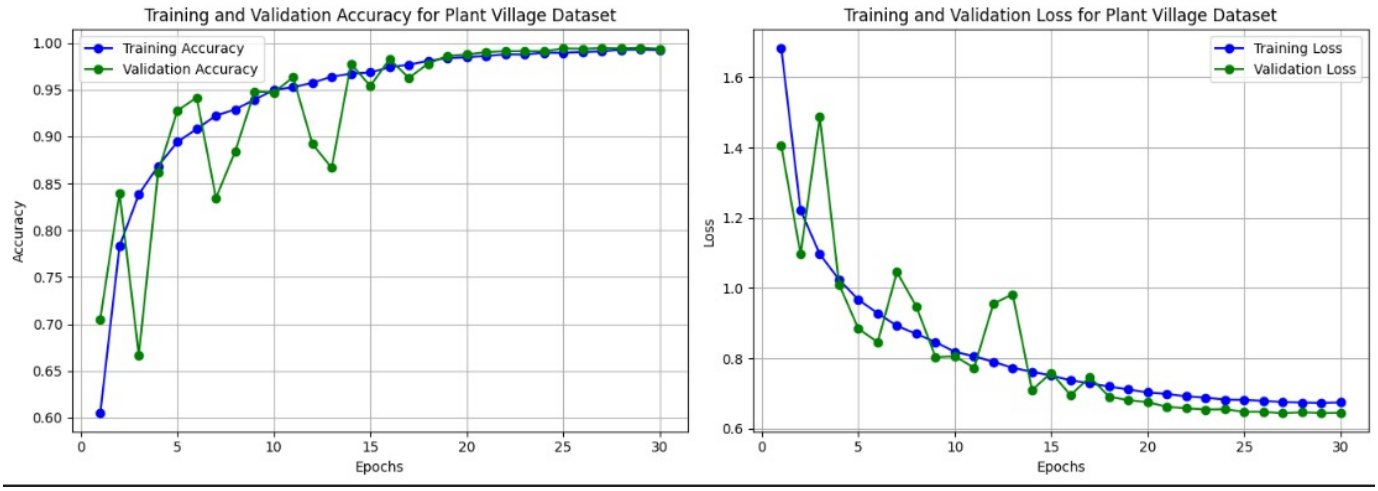

Fig. 3. Training and validation performance of RSA-DeRefNet on the Plant Village dataset. Left: Accuracy curves. Right: Loss curves.

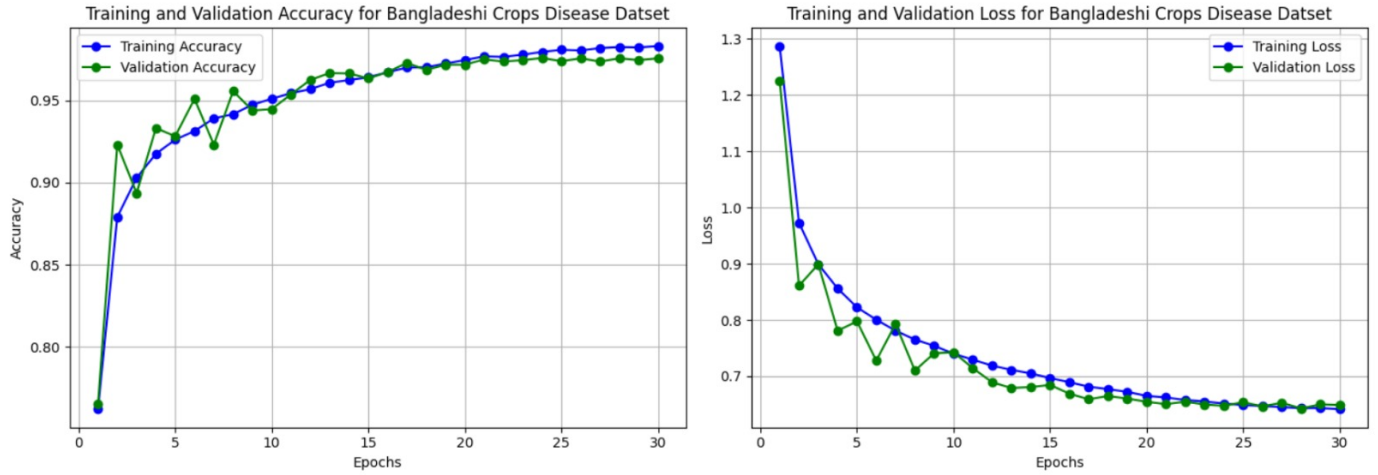

Fig. 4. Training and validation performance of RSA-DeRefNet on the Bangladeshi Crops dataset. Left: Accuracy curves. Right: Loss curves.

B. Plant Village Dataset

The learning curves for the Plant Village dataset are shown in Figure 3 This dataset is significantly larger and more diverse, containing images of multiple crops under varying conditions. As a result, its training dynamics show slightly different characteristics:

- **Accuracy Trends:** The training accuracy starts at 61% in the first epoch and rises steadily, crossing 90% by epoch 10 and reaching nearly 99% by epoch 25. Validation accuracy, however, exhibits greater fluctuation in the early epochs (between 70% and 90%), reflecting the dataset's intra-class variability. Despite this, validation accuracy stabilizes after epoch 15 and converges around 98.5%, closely tracking the training curve in the later epochs.
- **Loss Trends:** The training loss decreases consistently from 1.7 to about 0.65 over 30 epochs. Validation loss, in contrast, shows larger oscillations during the first 10 epochs, indicative of the dataset's complexity and possible noisy samples. Nevertheless, both curves converge below 0.7 in the later epochs, demonstrating that the model maintains stability as training progresses.

C. Comparative Observations

From the analysis of both datasets, the following comparative insights can be highlighted:

- The Bangladeshi Crops Disease dataset exhibits smoother training dynamics with minimal fluctuations, as it contains fewer classes and relatively cleaner samples.
- The Plant Village dataset demonstrates higher early-epoch variability, but the final convergence to nearly 99% accuracy indicates the scalability of RSA-DeRefNet to large and diverse datasets.
- Across both datasets, the close alignment between training and validation curves confirms that the proposed model achieves excellent generalization without significant overfitting.
- The consistent convergence of loss curves further validates the stability of the optimization process.

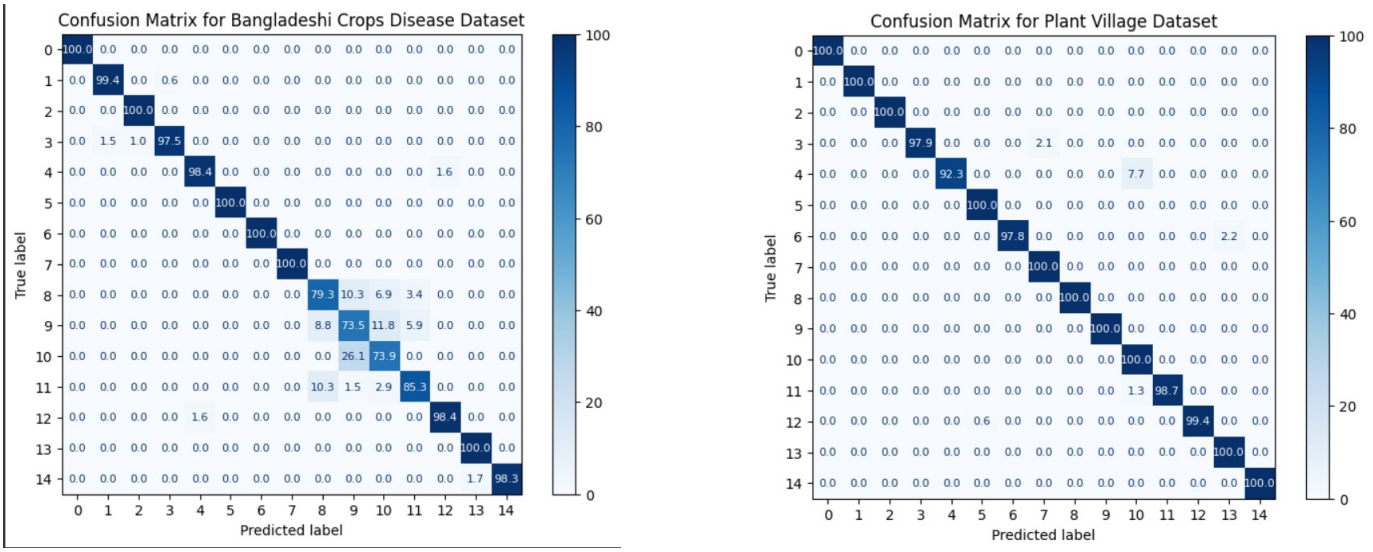

Fig. 5. Confusion matrices for crop disease classification: (a) Bangladeshi Crop Disease dataset, where minor misclassifications occur due to inter-class similarity, and (b) PlantVillage dataset, where the model achieves near-perfect recognition across categories.

D. Effect of Various Parameters

Learning Rate: Tested $\eta \in \{1 \times 10^{-3}, 3 \times 10^{-4}, 1 \times 10^{-4}, 1 \times 10^{-5}\}$. Best performance at $\eta = 3 \times 10^{-4}$ with cosine annealing.

Batch Size: Tested $\{16, 32, 64\}$. Batch size of 32 provided optimal balance between stability and generalization.

Dropout Rate: Tested $\{0.3, 0.5, 0.7\}$ at fully connected layers. Dropout of 0.5 yielded highest test accuracy.

Attention Modules: Removing RSA blocks caused a 4.2% accuracy drop. RSA in the residual stream contributed more than in the dense stream.

Dense Block Depth: Tested $\{3, 5, 7\}$ layers per block. Performance peaked at 5 layers; deeper configurations showed marginal gains with higher complexity.

Feature Fusion: Concatenation outperformed averaging and weighted summation by 2.8%, preserving complementary representations from both streams.

TABLE V
RESULTS OBTAINED USING VARIOUS DEEP LEARNING MODELS ON BANGLADESHI CROPS DATASET

Models	CNN	ResNet-50	Efficient-B3	VGG19	RSA-Derefnet
Accuracy	95.66	98.47	98.62	94.49	98.32

E. Quantitative Analysis

Fig. 5 presents confusion matrices for both datasets. On the Bangladeshi Crops dataset, the model achieves near-perfect accuracy for most classes, with minor misclassifications between Classes 9 and 10 due to inter-class visual similarity and limited training samples for these minority classes.

On the PlantVillage dataset, classification is near-perfect across all categories, with only occasional confusion between visually similar diseases (e.g., early vs. late blight in tomato). The balanced class distribution and controlled imaging conditions of PlantVillage contribute to this stronger performance.

VI. CONCLUSION

Overall, the proposed RSA-DeRefNet achieves superior training stability and generalization across both datasets. The model's dual-stream design—combining residual refinement with dense feature exploration—plays a central role in balancing convergence speed and robustness. These results substantiate the suitability of RSA-DeRefNet for real-world agricultural applications, where datasets vary widely in scale, quality, and class distribution.

DECLARATION OF COMPETING INTEREST

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

REFERENCES

[1] H. Al-Hiary *et al.*, "Fast and Accurate Detection and Classification of Plant Diseases," *Int. J. Comput. Appl.*, 2011.
[2] A. Krizhevsky *et al.*, "ImageNet Classification with Deep Convolutional Neural Networks," *NeurIPS*, 2012.
[3] K. He *et al.*, "Deep Residual Learning for Image Recognition," *IEEE CVPR*, 2016.
[4] A. Das *et al.*, "A Generative Framework for Detection and Classification of Plant Leaf Disease Using Diffusion Network," *Appl. Soft Comput.*, 2025.
[5] J. P. Sahoo *et al.*, "DeReFNet: Dual-stream Dense Residual Fusion Network for Static Hand Gesture Recognition," *Displays*, 2023.
[6] P. R. Shinde and M. V. Bhise, "Cotton Leaf Disease Detection Using Image Processing," *Int. J. Eng. Sci. Res. Technol.*, 2015.
[7] Zhou, Junhao & He, Zhanhong & Song, Ya & Wang, Hao & Yang, Xiaoping & Lian, Wenjuan & Dai, Hong-Ning. (2019). Precious Metal Price Prediction Based on Deep Regularization Self-Attention Regression. IEEE Access. PP. 10.1109/ACCESS.2019.2962202.
[8] N. A. Moin, "Bangladeshi Crops Disease Dataset," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/nafishamoin/bangladeshi-crops-disease-dataset.
[9] A. Alidev, "PlantVillage Dataset," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset.