IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

LeafNet: Dual-Track Feature Fusion with RCF Attention Framework for Pear Leaf Disease Classification

R. Karthik[1], Paarth Jain[2], Aryan Singh[2], Aryan Mahawar[2] and T. Illakiya[3]

[1]Centre for Cyber Physical Systems, Vellore Institute of Technology, Chennai, Tamil Nadu 600127, India
[2]School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu 600127, India
[3]Department of Computational Intelligence, School of Computing, Faculty of Engineering and Technology, SRM Institute of Technology, Chennai, Tamil Nadu 600127, India

Corresponding author: R. Karthik (r.karthik@vit.ac.in).

ABSTRACT Pear plants, a crucial crop in horticulture, are highly vulnerable to various diseases. These diseases significantly affect their yield and quality. Traditional methods for identifying plant diseases rely heavily on expert knowledge and observational skills, making them time-consuming and labor-intensive. Deep learning techniques offer a more efficient and accurate alternative for disease detection through image analysis. This research proposes a novel dual-track architecture for plant disease detection, focusing specifically on Pear plants. The proposed network consists of two tracks: Cross Vision Transformer (Cross ViT) and a custom Convolutional Neural Network (CNN). The Cross ViT path is designed to capture global features by leveraging the transformer's ability to model long-term dependencies and complex feature representations. Simultaneously, the custom Convolutional Neural Network path integrates Residual Channel Shuffled Attention and coordinate attention modules to extract and enhance local features effectively. The original DiaMOS plant dataset consists of 3,505 images across four classes: curl, healthy, slug, and spot. Data preprocessing and augmentation are performed using CycleGAN to address class imbalance, ensuring a diverse and robust training dataset. The CycleGAN method enhances this dataset by generating additional samples, balancing the distribution of classes and improving the model's generalization capabilities. The features from both paths are subsequently concatenated and passed through a condensation attention module. This module refines the feature maps by emphasizing significant patterns and suppressing irrelevant information. The final layers include flattening and fully connected layers, leading to the output layer for disease classification. The proposed network achieves an accuracy level of 88.61%, significantly enhancing detection accuracy.. This novel approach promotes more efficient disease management in horticulture.

INDEX TERMS Pear Leaf Disease, Deep Learning, Cross Vision Transformer, Dual-Track Attention, CycleGAN, Grad-CAM

I. INTRODUCTION

Pear production is an essential agricultural activity, widely practiced in various regions around the world, particularly in countries with temperate climates such as China, Italy, and the United States. According to the Food and Agriculture Organization (FAO), global pear production reached approximately 24 million metric tons in 2022 [1]. China leads as the top producer, contributing more than 70% of the world's total pear output, followed by Italy and the United States. The cultivation of pears significantly supports the economies of these countries, providing employment and sustaining local agricultural industries. Pears are primarily consumed fresh, but they are also used in processed forms such as canned pears, juices, and jams. Despite the economic significance of pear cultivation, pear trees are vulnerable to various diseases that can severely impact yield and quality. Common pear diseases include fire blight, scab, rust, and various fungal infections, which can lead to substantial economic losses for farmers. Fire blight, caused by the bacterium Erwinia Amylovora, is particularly devastating, as it can rapidly destroy young trees and severely damage older ones [2]. Pear scab, resulting from the fungus Venturia Pirina, leads to blemished fruits and defoliated trees which reduces both marketability and photosynthetic efficiency [3]. Pear rust, caused by different species of Gymnosporangium affects leaves and fruits, causing deformations and premature drop [4]. However, traditional disease detection methods, relying on visual inspections, are often time-consuming, labor-intensive, and subject to human error [5].

To address these challenges, automated disease detection systems utilizing advanced machine learning and deep learning techniques have been developed. These systems offer a promising solution for timely and accurate identification of pear leaf diseases, facilitating early intervention and reducing crop losses. The DiaMOS Plant

dataset, specifically designed for the diagnosis and monitoring of plant diseases, provides a robust foundation for development [6]. This dataset includes high-quality images of pear leaves affected by various diseases, enabling the training of deep learning models to recognize and classify disease symptoms with high accuracy [7]. The use of CNNs has shown significant potential in improving the precision and reliability of disease detection in complex agricultural environments. This potential is particularly evident when CNNs are enhanced with attention mechanisms [8].

Developing an automatic disease detection system for pear leaves involves challenges in data quality, model design, real-time processing, user adoption, and the need for well-labeled, diverse datasets for deep learning classification. Additionally, integrating explainable AI is essential for understanding the model's decision-making process. Furthermore, the development of a system that is both user-friendly and scalable, while maintaining reliability, remains essential. In this research, we propose an automatic disease detection system for pear leaves using a customized deep learning network. The proposed network leverages the DiaMOS Plant dataset to train a model capable of accurately identifying and classifying multiple pear leaf diseases [9]. The proposed approach integrates several modules, including Explainable AI (XAI) methods, to provide insights into model decisions. This fosters trust and adoption among users while enhancing the network's ability to capture intricate disease features. The system aims to provide farmers with a practical tool for remote pear leaf disease management, promoting sustainable agriculture through reduced pesticide use and optimized resource allocation. Additionally, it contributes to precision agriculture by improving crop yield and quality.

II. RELATED WORKS

Recent advancements in deep learning have significantly enhanced the detection and classification of plant diseases, including those in pear plants. Methods such as CNNs for feature extraction and classification, Transfer Learning for efficient training with smaller datasets, Generative Adversarial Networks (GANs) for data augmentation and anomaly detection, and Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) units for temporal analysis and environmental integration have all contributed to improved disease management in pear orchards amidst diverse environmental challenges.

The research conducted by Fenu and Malloci on multi-output learning has significantly enhanced the accuracy of plant disease and stress evaluation through the application of advanced machine learning techniques [1]. The integration of attention mechanisms, such as the Convolutional Block Attention Module (CBAM) demonstrated by Alirezazadeh, Schirrmann, and Stolzenburg, has substantially enhanced deep learning models for plant disease classification [3]. In the domain of pear plant diseases, recent studies have also made significant progress. Wu, Luo, and Xu developed DBPNet with a modified MobileNetV2 to address the challenge of recognizing pear leaf diseases in complex backgrounds [4], while Alshammari et al. introduced a Cycle Generative Adversarial Network (CycleGAN) for more accurate classification of various pear diseases [5]. These combined advancements underscore the ongoing progress and refinement in plant disease detection and classification through innovative machine learning approaches.

Ensemble methods and CNNs have also been widely used to enhance the accuracy of plant disease classification. Fenu and Malloci demonstrated the effectiveness of ensemble CNNs for classifying pear leaf diseases, showing the benefits of combining multiple models for higher accuracy [6]. Wang et al. introduced MFBP-UNet for pear leaf disease detection, leveraging advanced techniques such as multi-scale feature extraction and bilinear pooling. These methods enhance the model's ability to capture and integrate detailed information from various scales, improving detection accuracy and robustness in natural agricultural environments.[7]. Li et al. presented a lightweight algorithm based on an improved YOLOv5 model for recognizing pear leaf diseases in natural scenes.[9]. Their model is designed to operate efficiently on low-computing platforms, making it accessible for use in various agricultural settings. Transfer learning and pre-trained models have been employed to mitigate the constraints posed by small datasets thereby leveraging insights gained from larger, related datasets to improve model generalization and performance. Hassan et al. employed CNN and transfer learning approaches for plant-leaf disease identification, demonstrating enhanced model performance with limited training data [9]. Schwarz Schuler et al. presented a robust deep learning classification method based on light-chroma separated branches for enhancing classification accuracy under various lighting conditions [10].

Recent advancements in plant disease detection and classification through deep learning have yielded significant progress across various studies. Ullah et al. introduced DeepPlantNet, a robust CNN model that excels in identifying multiple plant diseases simultaneously, leveraging advanced neural network architectures and extensive training datasets [12]. Yang et al. conducted a comprehensive analysis of key factors influencing pear disease recognition, including environmental variables, image quality, and disease progression stages, contributing valuable insights for more reliable disease identification systems [14]. Gu et al. enhanced multi-plant disease recognition using deep CNNs, optimizing model architectures and training strategies to achieve state-of-the-art performance across diverse plant species [15]. Saleem et al. optimized deep learning models for agricultural applications by employing data augmentation techniques to enhance dataset diversity, transfer learning to leverage pre-trained models, and ensemble learning to improve prediction robustness [17]. They also utilized domain adaptation methods to ensure model performance in real-world settings and systematically tuned hyperparameters for optimal results. These approaches collectively addressed environmental and practical challenges, enhancing model applicability in agricultural contexts.

Although deep learning models have shown promising results in detecting pear leaf diseases, several limitations need to be addressed. These limitations include class imbalance, where certain diseases may be underrepresented in training data. Furthermore, there is a lack of focus on critical leaf features such as texture variations, vein patterns, and subtle discolorations that are indicative of early-stage diseases. The proposed work aims to tackle these research gaps by employing enhanced data augmentation techniques to balance class distributions and developing adaptive learning mechanisms that account for diverse disease manifestations. Additionally, it incorporates advanced feature extraction methods to capture and emphasize critical leaf characteristics essential for accurate disease detection and classification.

A. RESEARCH GAPS AND MOTIVATION

The proposed study effectively addresses the following research gaps in pear leaf disease detection:

1) Previous studies have primarily concentrated on either local feature extraction or global feature extraction independently,

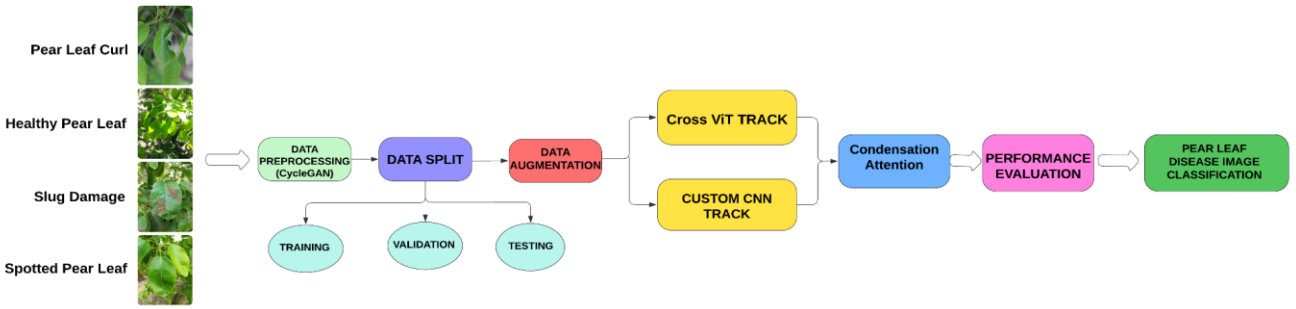


FIGURE 1. ARCHITECTURE OF PROPOSED MODEL

often resulting in suboptimal performance. However, a more effective approach would combine local details with global context, leading to a more comprehensive understanding and improved model performance.

2) A major challenge in employing deep learning for plant disease detection is the problem of data variability. Symptoms of diseases can present a wide range of color changes, spot sizes and distributions with varying levels of wilting. This intra-class variability poses a challenge for conventional deep learning models, potentially leading to erroneous diagnoses.

The insufficient capture of positional relationships and spatial information by CNNs in previous studies limits their ability to fully comprehend global contextual information and effectively manage long-range dependencies, which are crucial for identifying subtle disease symptoms.

B. RESEARCH CONTRIBUTIONS

The following are the contributions made toward addressing the gaps stated above:

1. The proposed network aims to improve the classification of pear leaf diseases by combining features from two different tracks to extract both global and local features. The first track uses a CNN to extract local features from the input image, while the second track utilizes a Cross ViT to extract global features.

2. CvT tackles the issue of diverse presentations of plant diseases within a single class. This research explores the incorporation of Channel Attention Modules (CAMs) into the CvT architecture for plant leaf disease detection. CAMs have the potential to improve CvT's ability to precisely localize disease symptoms within leaf images.

3. To enhance the capability of CNNs in detecting subtle disease symptoms present in plant leaf images, this model integrates Residual Channel Shuffled Attention (RCSA) and Coordinate Attention mechanisms. RCSA enhances inter-channel dependencies, while Coordinate Attention focuses on modeling positional relationships within the images.

III. PROPOSED SYSTEM

The proposed methodology consists of a dual-track architecture. This architecture integrates the Cross ViT and a custom CNN with Residual Channel Shuffled Attention (RCSA) and coordinate attention. The primary objective of the Cross ViT path is to capture intricate patterns and relationships within the image data through transformer mechanisms. Concurrently, the custom CNN path aims to extract predominant features via convolution operations. These operations are embedded with RCSA and coordinate attention modules, which enhance feature representation by focusing on important spatial and channel-wise information. Initially, the data undergoes preprocessing and augmentation using CycleGAN. This process increases the diversity and robustness of the training dataset. The preprocessed images are then directed through the two distinct paths: the Cross ViT path and the custom CNN path. The feature maps generated from both paths are concatenated to combine the strengths of each approach. This combined feature set is further refined using a condensation attention module, which enhances and highlights the most critical information. The refined features are then flattened and passed through fully connected layers to enable the model to learn complex patterns and relationships. Finally, the output layer produces the classification results. This fusion of transformer-based and CNN-based architectures is designed to improve overall performance and generalization capability in detecting pear leaf diseases.

A. PROPOSED MODEL

The proposed network architecture integrates two distinct models for efficient and effective detection and classification of pear leaf diseases: a Cross ViT and a Custom CNN. The Custom CNN is enhanced with Residual Channel Shuffled Attention and a coordinate attention module.

B. CROSS VISION TRANSFORMER BLOCK

A key component of the proposed architecture is the Cross ViT. It is designed to discern intricate patterns and dependencies in pear leaf images. Cross ViT is an advanced variant of the Vision Transformer (ViT) architecture tailored for image processing tasks. It enhances ViT by introducing cross-attention mechanisms alongside traditional self-attention. The architecture initiates the process by dividing the input image into patches. Each patch undergoes token embedding and positional encoding to capture spatial relationships. Cross ViT employs multiple transformer encoder blocks, each with cross-attention layers that facilitate interactions between patches across the image. These layers enable feature extraction by allowing patches to attend to features within

their own spatial domain and to relevant features across the entire image. Within each block, multi-head self-attention mechanisms capture intra-patch dependencies, and feedforward networks apply transformations to refine patch representations. This hierarchical approach is pivotal for extracting subtle features from images, empowering Cross ViT with comprehensive global context understanding, essential for tasks such as object detection and segmentation.

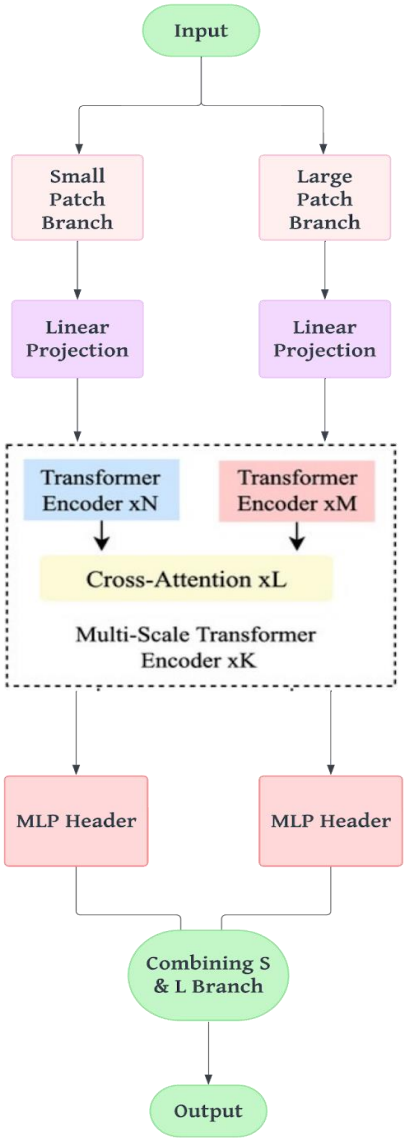


FIGURE 2. ARCHITECTURE OF THE CROSS ViT

The Cross ViT track is designed to integrate and process multi-scale features through several components. The Transformer Encoder xN, consisting of 'N' layers of a standard transformer encoder, processes input from the S-branch (small-scale branch) to focus on extracting features and capturing dependencies within the sequence. In parallel, the Transformer Encoder xM, comprising 'M' layers, handles input from the L-branch (large-scale branch), similarly focusing on feature extraction and dependency capture. The Cross-Attention xL mechanism facilitates interaction between these sequences by allowing one sequence (e.g., from the S-branch) to attend to another (e.g., from the L-branch). This process integrates and aligns information from both sequences to create a unified feature representation. The Multi-Scale Transformer Encoder xK, consisting of K layers, operates on the combined outputs from both branches. It handles multi-scale inputs and further refines the integrated information to ensure accurate final classification

The effectiveness of Cross ViT is highlighted by its final classification head, which uses the processed representations to make accurate predictions. This demonstrates Cross ViT's ability to tackle complex image processing challenges. Figure 2 represents the architecture of Cross ViT, highlighting its distinctive components within the framework.

C. RCF ATTENTION FRAMEWORK

The RCF Attention Framework (Residual Channel Fusion) is a crucial component of the proposed architecture. It is meticulously designed to enhance feature extraction and discrimination in pear leaf images. This architecture integrates techniques aimed at capturing and refining both spatial and channel-wise dependencies. These are critical for precise and reliable disease classification in agricultural settings. The RCF Attention Framework incorporates Residual Channel Shuffled Attention and Coordinate Attention Module (CAM). Together, these techniques enhance feature representation through dynamic recalibration of channel-wise features and selective spatial attention. The RCSA mechanism improves feature extraction by adaptively focusing on salient features while suppressing irrelevant noise. This significantly enhances the network's robustness and discriminative power. This allows the CNN to identify subtle variations in leaf textures and structures, which is essential for accurate detection and classification of diseases such as curl, spot, and slug. Furthermore, the integration of the Coordinate Attention Module within the RCF Attention Framework refines feature extraction capabilities. It selectively attends to significant spatial locations within pear leaf images.

This enhances the CNN's ability to capture fine-grained details and spatial relationships, ensuring precise localization and characterization of disease symptoms across varied environmental conditions. The integration of RCSA and CAT in the RCF Attention Framework empowers the CNN to extract hierarchical features with

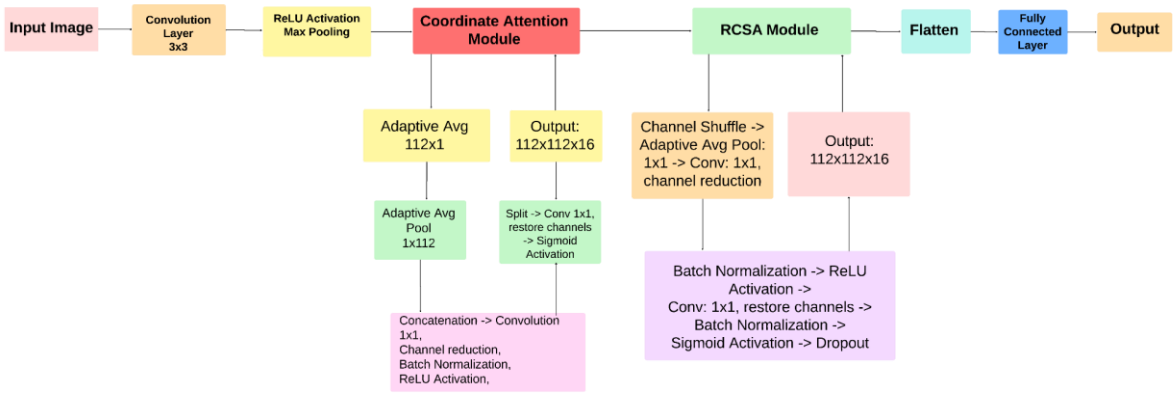


FIGURE 3. ARCHITECTURE OF THE RCF ATTENTION FRAMEWORK

improved accuracy and efficiency. This advancement is crucial for improving disease diagnosis in agricultural contexts. It supports timely interventions and sustainable management practices. Ultimately, it contributes to enhanced crop health and productivity.

The residual connection integrates channel shuffle attention to capture inter-channel dependencies while preserving the original feature map's identity. The attention mechanism is described by Eq.(1):

$$y = RCSA(x) = f(W_1 . Shuffle(W_2 . x)) + x \quad (1)$$

Here, 'x' represents the input feature map or image, 'W_1' and 'W_2' are learnable weights, and 'Shuffle(.)' represents the channel shuffle operation. The residual connection, '$f(.)$', helps retain the original feature map, while the shuffle improves channel-wise feature interaction. The Coordinate Attention Block introduces spatial information into the channel attention mechanism by encoding position-specific information through coordinate encoding. The attention computation is given by Eq.(2):

$$CA(x) = x . \sigma\,(W_h . AvgPool_h(x) + W_w . AvgPool_w(x)) \quad (2)$$

In this equation, 'x' is the input feature map, 'W_h' and 'W_w' are the learnable weights for height and width, respectively, and 'σ' represents the sigmoid activation. '$AvgPool_h(x)$' and '$AvgPool_w(x)$' represent average pooling operations applied along the height and width dimensions, respectively, introducing spatial sensitivity into the attention mechanism. To further enhance generalization, a residual scaling mechanism is added. The operation is described by Eq.(3):

$$y = \alpha . x + \beta . RCSA(x) \quad (3)$$

Here, 'α' and 'β' are learnable parameters that scale the contributions of the original input 'x' and the RCSA output. This improves the model's capacity for learning and adjusting feature representations. To fuse attention-enhanced features with learned spatial information, a convolution layer is applied to the output of the Coordinate Attention mechanism, as defined by Eq.(4):

$$y = Conv(CA(x)) = W_{conv} . CA(x) + b \quad (4)$$

Here, 'W_{conv}' represents the convolution filter weights, and 'b' is the bias term. The convolution operation further processes the attention-weighted features, enhancing local and global context understanding.

D. CONDENSATION BLOCK

The Condensation Block (CB) module is a critical component of the neural network architecture. It is strategically positioned immediately after the concatenation of Track 1 (Cross Vision Transformer - Cross ViT) and Track 2 (RCF Attention Framework). Its primary function is to enhance the performance of convolutional neural networks by refining feature representations extracted from preceding layers. Through the integration of sophisticated attention mechanisms and advanced feature aggregation techniques, the CB module adeptly prioritizes and enhances spatial and channel-wise information crucial for achieving precise and reliable classification results. In the context of our proposed architecture, the CB module optimizes the feature maps derived from both the Cross ViT and RCF Attention Framework tracks. By aggregating and refining these multi-modal features, the CB module ensures the network attains an enhanced ability to identify subtle and complex patterns essential for the accurate identification and classification of various pear leaf diseases. Moreover, its ability to effectively reduce the dimensionality of feature maps significantly enhances computational efficiency during inference. This results in the

FIGURE 4. ARCHITECTURE OF THE CONDENSATION BLOCK

network being highly responsive and ideally suited for real-time applications in
agricultural settings.

The CB module performs average pooling to aggregate features, expressed as Eq.(5):

$$y = avg\_pool(x) = 1 / (W * H) * \sum_{ij}(x_{ij}) \quad (5)$$

where 'W' and 'H' represent the width and height of the spatial dimensions, respectively, and the summation is performed over all spatial dimensions '*i*' and *'j'*. This is followed by a fully connected layer operation, defined by Eq.(6):

$$y = fc(y) = W_2 * ReLU(W_1 * y + b_1) + b_2 \quad (6)$$

Here, 'W_1' and 'W_2' are weight matrices, 'b_1' and 'b_2' are bias vectors, and 'ReLU' is the rectified linear unit activation function. These processes collectively enhance feature representation, ensuring robust and accurate classification while maintaining computational efficiency.

E. CLASSIFICATION

The classification stage plays an important role in identifying and categorizing various types of diseases affecting pear plants. Following the feature extraction from both the Transformer Track (Cross ViT) and the RCF Attention Framework, the condensed feature maps are processed through the classification pipeline. To manage the complexity and enhance model generalization, a strategy similar to RCSANet and the SA block approach is employed. First, global average pooling is applied to reduce the dimensionality of the feature maps, effectively mitigating overfitting and enhancing the model's robustness against spatial variations in leaf images. Next, the reduced feature representations are fed into fully connected layers tailored for multi-class classification of pear leaf diseases. Specifically, the network is designed to classify pear leaf images into distinct categories such as healthy leaves, those affected by curl, spot, slug, and other identifiable diseases prevalent in pear plants. The network parameters are learned by maximizing the focal loss of the predicted class probabilities with respect to the target class, as shown in Eq. (7):

$$FL\ (p_t) = -\alpha_t((1-p_t)^{\gamma} log(p_t)) \quad (7)$$

p_t is the model's estimated probability for the true class label, α_t is the class weights to balance the loss and γ is the focusing parameter that modulates the effect of the loss. The Focal Loss function effectively addresses class imbalance by down weighting easier samples and emphasizing harder examples, thereby enhancing the model's ability to classify pear leaf diseases accurately. The learning process involves optimizing the Focal Loss function. This aims to minimize the loss between predicted and actual class probabilities. This ensures that the network achieves high accuracy and reliability in pear leaf disease classification.

IV. RESULT

This section presents the dataset description, data augmentation, environmental setup, ablation studies, and performance analysis.

A. DATASET DESCRIPTION

The proposed model uses the DiaMOS Plant Dataset, specifically focusing on pear leaves. The dataset comprises a total of 3505 images, categorized into four classes: curl, healthy, slug, and spot. Before any augmentation, the dataset has the following distribution: 54 images of curl, 43 images of healthy leaves, 2025 images of slug, and 884 images of spot. These classes represent different conditions of pear leaves, as detailed in Table 1.

TABLE 1. DIAMOS DATASET DESCRIPTION

Leaf Symptoms	Class Size
Healthy	43
Spot	884
Curl	54
Slug	2025

B. DATA AUGMENTATION

This subsection discusses the various techniques used for data augmentation to enhance the robustness of the model. The DiaMOS dataset contains images with varying dimensions. For consistency and effective feature extraction, all images were resized to 224x224 pixels before training. Given the class imbalance within the dataset, with a predominant number of slug images, data augmentation was essential. This contributed to preventing biased learning and enhancing generalization. The CycleGAN approach was utilized for data augmentation to generate realistic variations of the images. This method allows the model to learn from a more diverse set of examples, enhancing its ability to generalize. The following augmentations were performed using CycleGAN:

1. Generation of synthetic images by translating healthy leaf images to diseased leaf images.
2. Enhancement of existing images by altering leaf conditions to simulate different disease stages.
3. Improvement of dataset diversity by creating new, realistic images representing underrepresented classes.

C. ENVIRONMENTAL SETUP

All experiments involving the proposed network were conducted on a 24GB Nvidia A10G Tensor Core GPU utilizing PyTorch on an AWS EC2 instance. The computing environment consisted of an

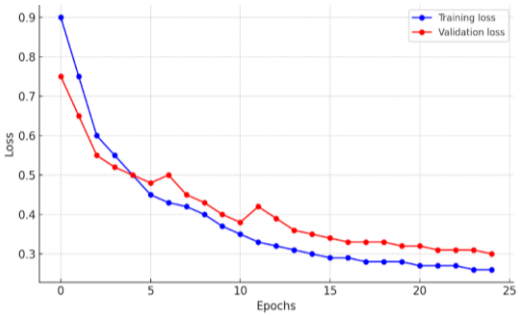
Ubuntu 20.04 operating system, 4 AMD vCPUs, and 16GB RAM. Hyperparameter tuning during the training phase was performed using the Adam gradient descent optimization algorithm, with a learning rate of 0.001. To address the class imbalance in the training dataset, focal loss was employed as the loss function. Training parameters included the use of the Adam optimizer, a batch size of 32, and training over 25 epochs. The learning rate was adjusted using the StepLR scheduler with a step size of 7 and a gamma of 0.1. The dataset was split into 80% for training, 10% for validation and 10% for testing.

V. ABLATION STUDY

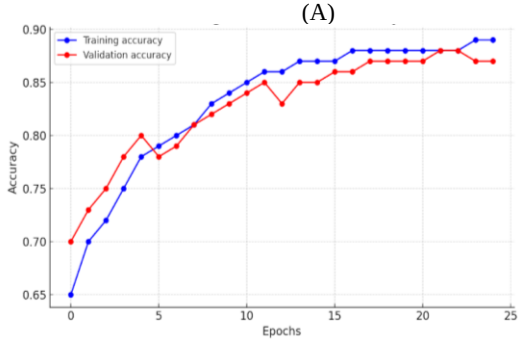
This section presents the ablation studies conducted to evaluate the impact of different components and configurations on the performance of the proposed model. The aim was to identify the contributions of each part of the architecture to the overall performance.

A. ANALYSIS OF CROSS VISION TRANSFORMER

To evaluate the effectiveness of the Cross ViT block, experiments were conducted using Cross ViT as a standalone component. The training was run for 25 epochs, resulting in a testing accuracy of 85.31%. The training and validation loss, along with the training and validation accuracy, are presented in the graphs below. The results demonstrate that the Cross ViT block enhances the model's ability to extract and learn features effectively, contributing to the overall performance improvement. The training and validation loss graph shows a steady decrease in both losses over the epochs, while the training and validation accuracy graph indicates a significant improvement.



(A)

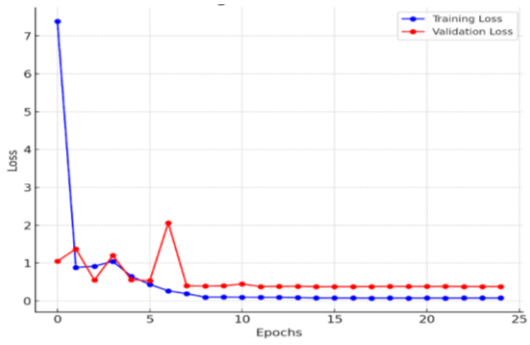


(B)

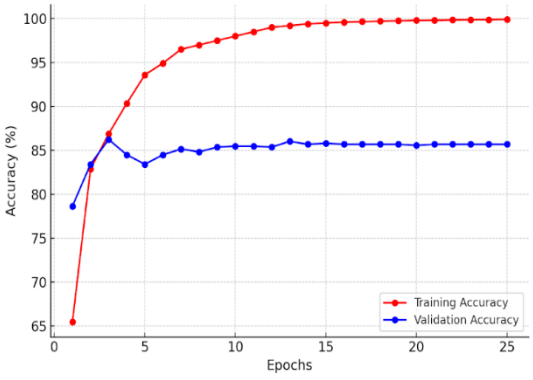
FIGURE 5. ANALYSIS OF CROSS ViT (A) LOSS (B) ACCURACY

B. ANALYSIS OF RCF ATTENTION FRAMEWORK

To evaluate the role of the RCF Attention Framework in refining the features extracted from the images, we conducted experiments isolating this component. The training was run for 25 epochs, resulting in a testing accuracy of 84.61%. The training and validation loss, along with the training and validation accuracy, are presented in the graphs below. The results indicate that the RCF Attention Framework effectively enhances the model's performance by emphasizing important features and suppressing irrelevant information. The training loss graph demonstrates a clear reduction in loss over the epochs, while the validation loss shows some fluctuation before stabilizing. The training accuracy graph illustrates a steady improvement, with the validation accuracy showing an initial rise followed by a plateau.



(A)



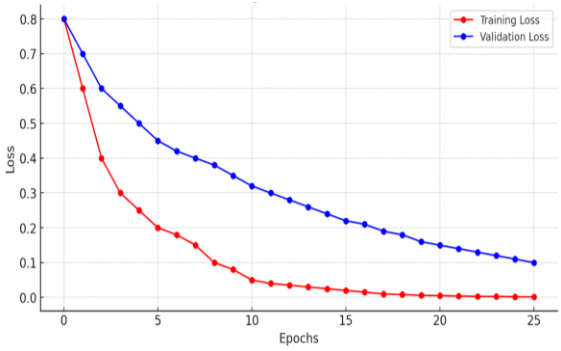
(B)

FIGURE 6. ANALYSIS OF THE RCF ATTENTION FRAMEWORK (A) LOSS (B) ACCURACY

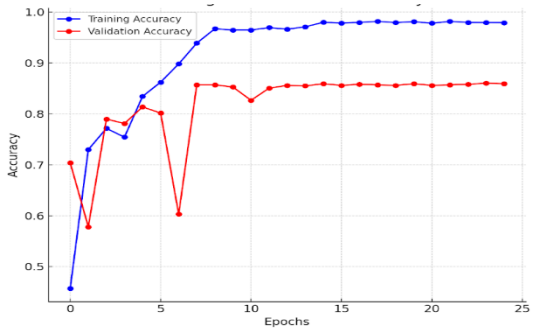
C. ANALYSIS OF CUSTOM CNN AND TRANSFORMER

To evaluate the combined effect of the Custom CNN and Transformer components on the model's performance, experiments were conducted isolating these components together. The training, run for 25 epochs, resulted in a testing accuracy of 85.22%. The training and validation loss, along with the training and validation accuracy, are presented in the graphs below. These results

demonstrate that combining the Custom CNN with Transformer components substantially enhances the model's learning capacity. However, the validation accuracy suggests potential overfitting, as evidenced by the discrepancy between the training and validation accuracies.



(A)



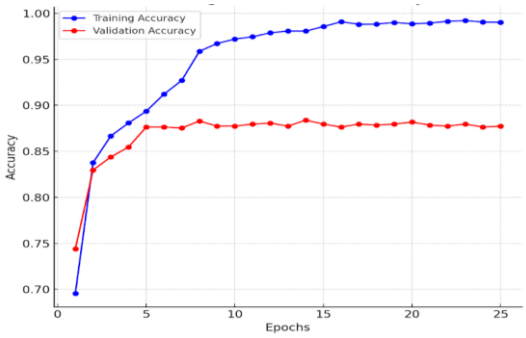
(B)

FIGURE 7. ANALYSIS OF THE PROPOSED NETWORK WITHOUT CONDENSATION ATTENTION (A) LOSS (B) ACCURACY

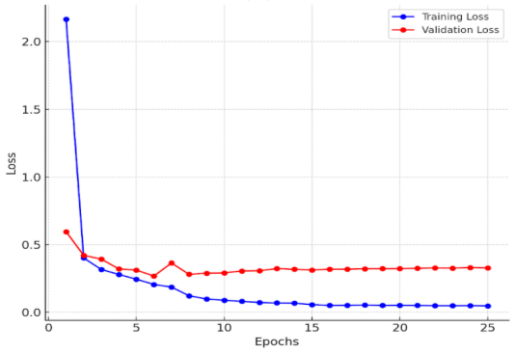
D. ANALYSIS OF PROPOSED WORK

To evaluate the overall performance of the proposed system, which integrates the Cross ViT, RCF Attention Framework, Custom CNN and Transformer components, extensive experiments were conducted. The training was executed over 25 epochs, resulting in a testing accuracy of 88.61%, with a testing loss of 30.34%. The precision, recall, and F1 score for the testing set were 88.62%, 88.61%, and 88.61%, respectively. The graphs below illustrate the training and validation loss, along with the training and validation accuracy.

The results indicate that the proposed system effectively combines the strengths of each component, achieving a high training accuracy and a solid validation accuracy, suggesting good generalization. The training and validation loss graph shows a consistent reduction in losses over the epochs, and the training and validation accuracy graph highlights the performance trends.



(A)

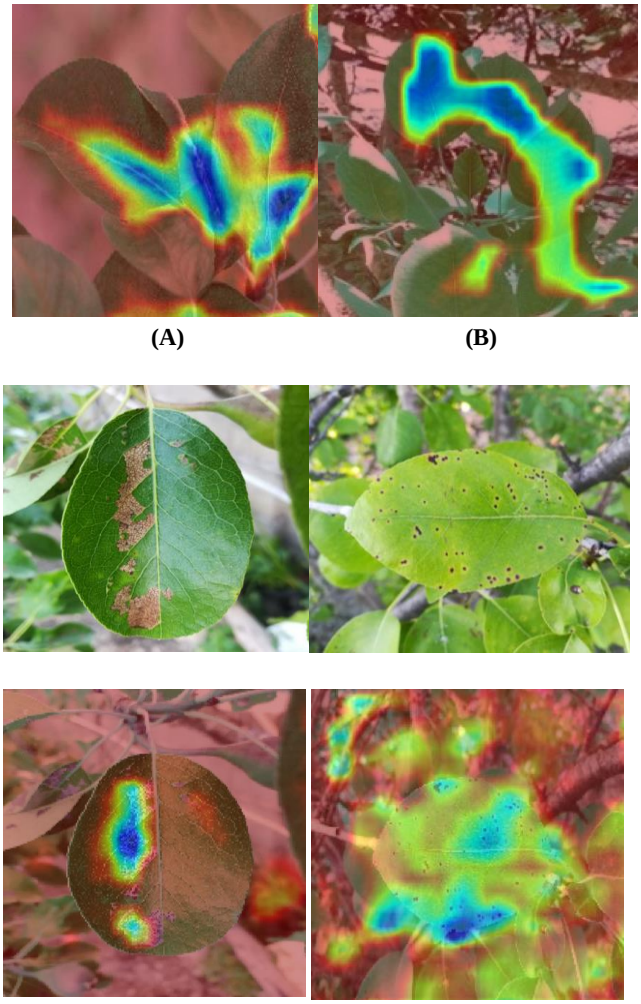


(B)

FIGURE 8. ANALYSIS OF THE PROPOSED NETWORK (A) ACCURACY (B) LOSS

E. GRAD CAM VIASULISATION

The Grad-CAM visualization results highlight the important regions influencing the trained model's performance. By applying Grad-CAM, we were able to identify the specific areas within the images from the DiaMOS plant dataset that significantly impacted the process of the proposed network. This enabled validation of the model's ability to accurately identify and focus on critical features of the plant leaves, thus providing a more interpretable and reliable framework for plant disease detection.



(A) (B)

(C) (D)

FIGURE 9. GRAD-CAM VISUALIZATION: (A) CURL; (B) HEALTHY; (C) SLUG; (D) SPOT

VI. DISCUSSION

A. PERFORMANCE ANALYSIS WITH STATE-OF-THE-ART NETWORKS

In the experiments, the performance of several pre-trained models for leaf disease detection was evaluated using the Diamos dataset. The models included EfficientNetB0, InceptionV3, MobileNetV2, ResNet50, VGG19, DenseNet, Xception, and AlexNet. Although these models are highly effective for general-purpose image classification tasks, they may not capture the specific features necessary to distinguish between subtle differences in leaf diseases. EfficientNetB0 demonstrated the highest accuracy among the pre-trained models, achieving a score of 86.33%. Xception followed with an accuracy of 84.77%, and MobileNetV2 attained an accuracy of 83.83%. InceptionV3 and DenseNet reached accuracies of 80.29% and 78.71%, respectively. VGG19 obtained an accuracy of 76.53%, while ResNet50 and AlexNet had lower accuracies of 68.47% and 68.55%. In contrast, the proposed dual-track system achieved an overall accuracy of 88.61%. This demonstrates the effectiveness of our approach in capturing the specific features necessary for accurate leaf disease detection.

TABLE 2. COMPARATIVE ANALYSIS WITH STATE-OF-THE-ART NETWORKS

Sl.No.	Neural Networks	Accuracy (in %)
1	ResNet50	68.47%
2	AlexNet	68.55%
3	VGG19	76.53%
4	DenseNet	78.71%
5	InceptionV3	80.29%
6	MobileNetV2	83.83%
7	Xception	84.77%
8	EfficientNetB0	86.33%
9	**Proposed Network**	**88.61%**

B. PERFORMANCE ANALYSIS WITH THE EXISTING STUDIES

The performance of the proposed approach is compared against existing studies for leaf disease detection using the Diamos dataset. For a fair comparison, only the studies that employed the same Diamos dataset were included. These studies performed classification in four classes: Healthy, Curl, Slug, and Spot. The evaluation metric employed in this study for comparison and performance analysis with other studies is accuracy. The existing studies applied pre-trained models (EfficientNetB0) for the detection of plant diseases. Although it is an effective strategy, it is designed for general-purpose image classification tasks, which means it does not capture the specific features needed for distinguishing between subtle differences in leaf diseases.

It is important to note that the proposed dual-track system, which combines CycleGAN with CrossViT and a custom CNN incorporating RSCA and condensation attention layers, is the first such custom architecture designed and implemented specifically for this dataset. The classification accuracies obtained by the existing studies range from 82% to 86%. In contrast, the proposed network has shown improved results. Overall, the proposed study outperformed the existing state-of-the-art approaches with an overall accuracy of 88.61%. Furthermore, the fusion of two tracks results in a network with greater performance and precise feature learning.

TABLE 3. COMPARATIVE ANALYSIS OF THE PERFORMANCE OF THE PROPOSED WORK WITH THAT OF OTHER EXISTING WORKS

Sl. no	Source	Methodology	Accuracy (in%)
1	Schwarz Schuler, J. P., Romani, S., Abdel-Nasser, M., Rashwan, H., & Puig, D. [11]	Modified Inception V3	80.5
2	Fenu, G., & Malloci, F. M. [1]	EfficientNetB0	82.82
3	Wongchai, A., Jenjeti, D. rao, Priyadarsini, A. I., Deb, N., Bhardwaj, A., & Tomar, P. [6]	Recursive Neural Network(RNN)	83.56
4	Alshammari, K., Alshammari, R., Alshammari, A., & Alkhudaydi, T. [4]	ResNet50 VGG19	84.28
5	Alirezazadeh, P., Schirrmann, M., & Stolzenburg, F. [3]	EfficientNetB0 + CBAM	84.89
6	Wu, X., Luo, Z., & Xu, H. [4]	MobileNetv2	85.65
7	**Proposed Network**	**Dual Track CNN and CrossViT**	**88.61**

C. LIMITATIONS AND FUTURE WORKS:

The current study is limited by its validation primarily in controlled experimental settings, which may not fully reflect the variability and complexity of real-world agricultural environments. Further research is required to assess the model's performance on diverse datasets, incorporating various crop types, disease conditions, and environmental factors. Future work will focus on training the model with large-scale landscape images captured via drone technology to facilitate real-time deployment in broader agricultural contexts. Additionally, refining segmentation algorithms to improve the precision of disease marker delineation and expanding the model's applicability to other crops will be essential. Integrating this system into mobile or web applications will enable real-time disease detection, offering farmers a practical tool to mitigate crop losses and enhance yield quality.

VII. CONCLUSION

Effective management of plant diseases is essential for sustaining agricultural productivity and ensuring food security. Traditional methods for detecting plant diseases are often time-consuming, labor-intensive, and susceptible to human error. This research introduces a novel two-track system for plant disease detection utilizing computer vision techniques, designed to address these challenges. The proposed architecture integrates a Cross-ViT path and a convolutional path, incorporating RCSA and coordinate attention mechanisms to enhance feature extraction. The Cross-ViT path captures long-range dependencies and global contextual information, while the convolutional path focuses on relevant spatial and channel-wise features. The features extracted from both paths are combined and further refined using a condensation attention module, thereby enhancing the network's representational power. The final output layer, derived from fully connected layers, provides accurate disease classification. The proposed system achieved an accuracy of 88.61%, outperforming existing methods. This two-track system, with its advanced feature extraction and attention mechanisms, represents a significant contribution to the field of automated plant disease detection.

REFERENCES

1. Fenu, G., & Malloci, F. M. (2021). DiaMOS Plant: A Dataset for Diagnosis and Monitoring Plant Disease. In Agronomy (Vol. 11, Issue 11, p. 2107). MDPI AG. https://doi.org/10.3390/agronomy11112107

2. Fenu, G., & Malloci, F. M. (2021). Using Multi Output Learning to Diagnose Plant Disease and Stress Severity. In A. Khan (Ed.), *Complexity* (Vol. 2021, pp. 1–11). Hindawi Limited. https://doi.org/10.1155/2021/6663442

3. Alirezazadeh, P., Schirrmann, M., & Stolzenburg, F. (2022). Improving Deep Learning-based Plant Disease Classification with Attention Mechanism. In *Gesunde Pflanzen*

(Vol. 75, Issue 1, pp. 49–59). Springer Science and Business Media LLC. https://doi.org/10.1007/s10343-022-00796-y

4. Wu, X., Luo, Z., & Xu, H. (2023). Recognition of Pear Leaf Disease under Complex Background Based on DBPNet and Modified MobileNetV2. In *IET Image Processing* (Vol. 17, Issue 10, pp. 3055–3067). Institution of Engineering and Technology (IET). https://doi.org/10.1049/ipr2.12855

5. Alshammari, K., Alshammari, R., Alshammari, A., & Alkhudaydi, T. (2024). An Improved Pear Disease Classification Approach Using Cycle Generative Adversarial Network. In *Scientific Reports* (Vol. 14, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41598-024-57143-6

6. Wongchai, A., Jenjeti, D. R., Priyadarsini, A. I., Deb, N., Bhardwaj, A., & Tomar, P. (2022). Farm Monitoring and Disease Prediction by Classification Based on Deep Learning Architectures in Sustainable Agriculture. In *Ecological Modelling* (Vol. 474, p. 110167). Elsevier BV. https://doi.org/10.1016/j.ecolmodel.2022.110167

7. Fenu, G., & Malloci, F. M. (2023). Classification of Pear Leaf Diseases Based on Ensemble Convolutional Neural Networks. In *AgriEngineering* (Vol. 5, Issue 1, pp. 141–152). MDPI AG. https://doi.org/10.3390/agriengineering5010009

8. Wang, H., Ding, J., He, S., Feng, C., Zhang, C., Fan, G., Wu, Y., & Zhang, Y. (2023). MFBP-UNet: A Network for Pear Leaf Disease Segmentation in Natural Agricultural Environments. In *Plants* (Vol. 12, Issue 18, p. 3209). MDPI AG. https://doi.org/10.3390/plants12183209

9. Li, J., Liu, Z., & Wang, D. (2024). A Lightweight Algorithm for Recognizing Pear Leaf Diseases in Natural Scenes Based on an Improved YOLOv5 Deep Learning Model. In *Agriculture* (Vol. 14, Issue 2, p. 273). MDPI AG. https://doi.org/10.3390/agriculture14020273

10. Hassan, S. M., Maji, A. K., Jasiński, M., Leonowicz, Z., & Jasińska, E. (2021). Identification of Plant-Leaf Diseases Using CNN and Transfer-Learning Approach. In *Electronics* (Vol. 10, Issue 12, p. 1388). MDPI AG. https://doi.org/10.3390/electronics10121388

11. Schwarz Schuler, J. P., Romani, S., Abdel-Nasser, M., Rashwan, H., & Puig, D. (2021). Reliable Deep Learning Plant Leaf Disease Classification Based on Light-Chroma Separated Branches. In *Frontiers in Artificial Intelligence and Applications*. IOS Press. https://doi.org/10.3233/faia210157

12. Ullah, N., Khan, J. A., Almakdi, S., Alshehri, M. S., Al Qathrady, M., El-Rashidy, N., El-Sappagh, S., & Ali, F. (2023). An Effective Approach for Plant Leaf Diseases Classification Based on a Novel DeepPlantNet Deep Learning Model. In *Frontiers in Plant Science* (Vol. 14). Frontiers Media SA. https://doi.org/10.3389/fpls.2023.1212747

13. Liu, Y., Liu, J., Cheng, W., Chen, Z., Zhou, J., Cheng, H., & Lv, C. (2023). A High-Precision Plant Disease Detection Method Based on a Dynamic Pruning Gate Friendly to Low-Computing Platforms. In *Plants* (Vol. 12, Issue 11, p. 2073). MDPI AG. https://doi.org/10.3390/plants12112073

14. Yang, F., Li, F., Zhang, K., Zhang, W., & Li, S. (2020). Influencing Factors Analysis in Pear Disease Recognition Using Deep Learning. In *Peer-to-Peer Networking and Applications* (Vol. 14, Issue 3, pp. 1816–1828). Springer Science and Business Media LLC. https://doi.org/10.1007/s12083-020-01041-x

15. Gu, Y. H., Yin, H., Jin, D., Zheng, R., & Yoo, S. J. (2022). Improved Multi-Plant Disease Recognition Method Using Deep Convolutional Neural Networks in Six Diseases of Apples and Pears. In *Agriculture* (Vol. 12, Issue 2, p. 300). MDPI AG. https://doi.org/10.3390/agriculture12020300

16. Wang, B. (2022). Identification of Crop Diseases and Insect Pests Based on Deep Learning. In A. Farouk (Ed.), *Scientific Programming* (Vol. 2022, pp. 1–10). Hindawi Limited. https://doi.org/10.1155/2022/9179998

17. Saleem, M. H., Potgieter, J., & Arif, K. M. (2022). A Performance-Optimized Deep Learning-Based Plant Disease Detection Approach for Horticultural Crops of New Zealand. In *IEEE Access* (Vol. 10, pp. 89798–89822). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/access.2022.3201104

18. Fu, J., Li, X., Chen, F., & Wu, G. (2024). Pear Leaf Disease Segmentation Method Based on Improved DeepLabv3+. In *Cogent Food & Agriculture* (Vol. 10, Issue 1). Informa UK Limited. https://doi.org/10.1080/23311932.2024.2310805

19. Moupojou, E., Tagne, A., Retraint, F., Tadonkemwa, A., Wilfried, D., Tapamo, H., & Nkenlifack, M. (2023). FieldPlant: A Dataset of Field Plant Images for Plant Disease Detection and Classification with Deep Learning. In *IEEE Access* (Vol. 11, pp. 35398–35410). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/access.2023.3263042

20. Rashid, J., Khan, I., Ali, G., Almotiri, S. H., AlGhamdi, M. A., & Masood, K. (2021). Multi-Level Deep Learning Model for Potato Leaf Disease Recognition. In *Electronics* (Vol. 10, Issue 17, p. 2064). MDPI AG. https://doi.org/10.3390/electronics10172064

21. Chowdhury, M. E. H., Rahman, T., Khandakar, A., Ayari, M. A., Khan, A. U., Khan, M. S., Al-Emadi, N., Reaz, M. B. I., Islam, M. T., & Ali, S. H. M. (2021). Automatic and Reliable Leaf Disease Detection Using Deep Learning Techniques. In

AgriEngineering (Vol. 3, Issue 2, pp. 294–312). MDPI AG. https://doi.org/10.3390/agriengineering3020020

22. J., A., Eunice, J., Popescu, D. E., Chowdary, M. K., & Hemanth, J. (2022). Deep Learning-Based Leaf Disease Detection in Crops Using Images for Agricultural Applications. In *Agronomy* (Vol. 12, Issue 10, p. 2395). MDPI AG. https://doi.org/10.3390/agronomy12102395

23. Amin, H., Darwish, A., Hassanien, A. E., & Soliman, M. (2022). End-to-End Deep Learning Model for Plant Disease Detection and Classification in Smart Agriculture. In *Sustainable Computing: Informatics and Systems* (Vol. 36, p. 100713). Elsevier BV. https://doi.org/10.1016/j.suscom.2022.100713

24. Jiang, P., Chen, Y., Liu, B., He, D., & Liang, C. (2019). Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks. In *IEEE Access* (Vol. 7, pp. 59069–59080). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/access.2019.2914929

25. Barburiceanu, S., Meza, S., Orza, B., Malutan, R., & Terebes, R. (2021). Convolutional Neural Networks for Texture Feature Extraction. Applications to Leaf Disease Classification in Precision Agriculture. In *IEEE Access* (Vol. 9, pp. 160085–160103). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/access.2021.3131002

26. Enkvetchakul, P., & Surinta, O. (2021). Effective Data Augmentation and Training Techniques for Improving Deep Learning in Plant Leaf Disease Recognition. In *Applied Science and Engineering Progress*. King Mongkut's University of Technology North Bangkok. https://doi.org/10.14416/j.asep.2021.01.003

27. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using Deep Learning for Image-Based Plant Disease Detection. In *Frontiers in Plant Science* (Vol. 7). Frontiers Media SA. https://doi.org/10.3389/fpls.2016.01419

28. Fenu, G., & Malloci, F. M. (2019). An Application of Machine Learning Technique in Forecasting Crop Disease. In *Proceedings of the 2019 3rd International Conference on Big Data Research* (ICBDR 2019: The 3rd International Conference on Big Data Research). ACM. https://doi.org/10.1145/3372454.3372474

29. Saleem, M. H., Potgieter, J., & Arif, K. M. (2019). Plant Disease Detection and Classification by Deep Learning. In *Plants* (Vol. 8, Issue 11, p. 468). MDPI AG. https://doi.org/10.3390/plants8110468

30. Sakkarvarthi, G., Sathianesan, G. W., Murugan, V. S., Reddy, A. J., Jayagopal, P., & Elsisi, M. (2022). Detection and Classification of Tomato Crop Disease Using Convolutional Neural Network. In *Electronics* (Vol. 11, Issue 21, p. 3618). MDPI AG. https://doi.org/10.3390/electronics11213618

31. Singh, A., & Kaur, H. (2021). Potato Plant Leaves Disease Detection and Classification using Machine Learning Methodologies. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1022, Issue 1, p. 012121). IOP Publishing. https://doi.org/10.1088/1757-899X/1022/1/012121

32. J., A., Eunice, J., Popescu, D. E., Chowdary, M. K., & Hemanth, J. (2022). Deep Learning-Based Leaf Disease Detection in Crops Using Images for Agricultural Applications. In *Agronomy* (Vol. 12, Issue 10, p. 2395). MDPI AG. https://doi.org/10.3390/agronomy12102395

33. Jepkoech, J., Mugo, D. M., Kenduiywo, B. K., & Too, E. C. (2021). Arabica Coffee Leaf Images Dataset for Coffee Leaf Disease Detection and Classification. In *Data in Brief* (Vol. 36, p. 107142). Elsevier BV. https://doi.org/10.1016/j.dib.2021.107142

34. Harakannanavar, S. S., Rudagi, J. M., Puranikmath, V. I., Siddiqua, A., & Pramodhini, R. (2022). Plant Leaf Disease Detection Using Computer Vision and Machine Learning Algorithms. In *Global Transitions Proceedings* (Vol. 3, Issue 1, pp. 305–310). Elsevier BV. https://doi.org/10.1016/j.gltp.2022.03.016

35. Sambasivam, G., & Opiyo, G. D. (2021). A Predictive Machine Learning Application in Agriculture: Cassava Disease Detection and Classification with Imbalanced Dataset Using Convolutional Neural Networks. In *Egyptian Informatics Journal* (Vol. 22, Issue 1, pp. 27–34). Elsevier BV. https://doi.org/10.1016/j.eij.2020.02.007

36. Chohan, M., Khan, A., Chohan, R., Katpar, S. H., & Mahar, M. S. (2020). Plant Disease Detection Using Deep Learning. In *International Journal of Recent Technology and Engineering (IJRTE)* (Vol. 9, Issue 1, pp. 909–914). Blue Eyes Intelligence Engineering and Sciences Publication - BEIESP. https://doi.org/10.35940/ijrte.a2139.059120

37. Sreya, J., & Arul, L. R. (2021). Machine Learning Techniques in Plant Disease Detection and Classification – A State of the Art. In *INMATEH Agricultural Engineering* (pp. 362–372). INMA Bucharest-Romania. https://doi.org/10.35633/inmateh-65-38

38. Sarkar, C., Gupta, D., Gupta, U., & Hazarika, B. B. (2023). Leaf Disease Detection Using Machine Learning and Deep Learning: Review and Challenges. In *Applied Soft Computing* (Vol. 145, p. 110534). Elsevier BV. https://doi.org/10.1016/j.asoc.2023.110534

R. KARTHIK received the master's degree from Anna University, India, and the Ph.D. degree from the Vellore Institute of Technology, Chennai, India. Currently, he is an Associate Professor with the Research Centre for Cyber Physical Systems, Vellore Institute of Technology. He has published around 75 papers in peer reviewed journals and conferences. His research interests include deep learning, computer vision, digital image processing, and medical image analysis. He is an active reviewer of journals published by Elsevier, IEEE, Springer, and Nature.

PAARTH JAIN is currently pursuing a Bachelor of Technology in Computer Science and Engineering at the Vellore Institute of Technology, Chennai, India. His research interests include machine learning, deep learning, computer vision, and natural language processing.

ARYAN M. is currently pursuing a Bachelor of Technology in Computer Science and Engineering with a specialization in Data Science at the Vellore Institute of Technology, Chennai, India. His research interests include machine learning, deep learning and computer vision.

ARYAN SINGH is currently in the final year of a B.Tech degree in Computer Science with a specialization in Robotics and AI from Vellore Institute of Technology, Chennai, India. His research interests focus on deep learning and machine learning-related topics. He actively participates in academic research and projects in these fields.

T. ILLAKIYA is working as an Assistant professor in the Department of Computational Intelligence, SRM Institute of Science and Technology, Chennai, India. She has completed a Ph.D. degree from the School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India. She received the M.Tech. degree in Information Technology from Anna University, India. She has published more than 20 research papers in journals and conferences. Her research interests include deep learning, computer vision, and medical image analysis.