# Illakiya T

## Team5_PlagCheckDoc.docx

Paper 5

Agile Practices I

SRM Institute of Science & Technology

## Document Details

**Submission ID**

**trn:oid:::1:3178897735**

**Submission Date**

**Mar 10, 2025, 8:00 PM GMT+5:30**

**Download Date**

**Mar 10, 2025, 8:03 PM GMT+5:30**

**File Name**

**Team5_PlagCheckDoc.docx**

**File Size**

**89.0 KB**

**8 Pages**

**6,189 Words**

**38,622 Characters**

# 9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

## Filtered from the Report

- ▸ Bibliography
- ▸ Quoted Text

## Match Groups

**43** Not Cited or Quoted 9%
Matches with neither in-text citation nor quotation marks

**0** Missing Quotations 0%
Matches that are still very similar to source material

**0** Missing Citation 0%
Matches that have quotation marks, but no in-text citation

**0** Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

## Top Sources

2% 🌐 Internet sources

8% 📖 Publications

1% 👤 Submitted works (Student Papers)

## Match Groups

**43** Not Cited or Quoted 9%
Matches with neither in-text citation nor quotation marks

**0** Missing Quotations 0%
Matches that are still very similar to source material

**0** Missing Citation 0%
Matches that have quotation marks, but no in-text citation

**0** Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

## Top Sources

2% 🌐 Internet sources

8% 📖 Publications

1% 👤 Submitted works (Student Papers)

## Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

| 1 | Publication |
|---|---|

R. Karthik, R. Menaka, M. V. Siddharth, Sameeha Hussain, Bala Murugan, Daehan ...  **3%**

| 2 | Publication |
|---|---|

R. Karthik, Sameeha Hussain, Timothy Thomas George, Rashmi Mishra. "A dual tr...  **2%**

| 3 | Publication |
|---|---|

R. Karthik, R. Menaka, S. Ompirakash, P. Bala Murugan, M. Meenakashi, Sindhia L...  **2%**

| 4 | Publication |
|---|---|

H L Gururaj, Francesco Flammini, V Ravi Kumar, N S Prema. "Recent Trends in He...  **<1%**

| 5 | Internet |
|---|---|

isteonline.in  **<1%**

| 6 | Publication |
|---|---|

R. Karthik, R. Menaka, M. V. Siddharth, Sameeha Hussain, P. Siddharth, Daehan W...  **<1%**

| 7 | Internet |
|---|---|

export.arxiv.org  **<1%**

| 8 | Internet |
|---|---|

getsmallbusinessideas.blogspot.com  **<1%**

| 9 | Internet |
|---|---|

fjs.fudutsinma.edu.ng  **<1%**

| 10 | Internet |
|---|---|

scipost.org  **<1%**

**11**  **Student papers**

**Queen's University of Belfast**                                              **<1%**

**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 (DL) 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 CNN path integrates Residual Channel Shuffled Attention (RCSA) and Coordinate Attention Modules (CAM) 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 Cycle Generative Adversarial Network (CycleGAN) to address class imbalance, ensuring a diverse 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 CAM. 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, Cycle Generative Adversarial Network, 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. Moreover, the rising demand for organic and high-quality fruits has increased the necessity for effective disease management strategies in pear orchards to ensure both quantity and quality in production.

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, reducing both marketability and photosynthetic efficiency [3]. Pear rust, caused by different species of Gymnosporangium, affects leaves and fruits, causing deformations and premature drop [4]. Additionally, fungal infections such as powdery mildew and black spot disease further contribute to declining pear production. These diseases not only reduce market value but also increase the cost of pest management, making early and accurate disease detection is important for sustainable agriculture.

Traditional disease diagnosis in pear orchards primarily relies on visual inspection by farmers and agricultural experts. This method, while widely practiced, has several limitations, including subjectivity, time consumption, and susceptibility to human error [5]. Moreover, by the time visible symptoms appear, the disease might have already spread extensively, leading to substantial crop losses. Various diagnostic techniques, such as microscopy, biochemical assays, and polymerase chain reaction (PCR), have been developed to detect plant pathogens more accurately. However, these methods are often expensive, require specialized knowledge, and may not be feasible for large-scale field applications. Computer-aided diagnosis (CAD) systems, integrating image processing and traditional statistical models, have been introduced to assist in disease detection, yet they often lack the strength required for complex agricultural environments.

To address these challenges, automated disease detection systems utilizing advanced machine learning (ML) and DL 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 strong 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 better accuracy [7]. Among ML techniques, CNNs have demonstrated significant potential in improving precision and reliability in plant disease classification. Furthermore, CNN architectures enhanced with attention mechanisms have shown remarkable improvements in performance by focusing on important areas of diseased leaf images [8].

Another aspect of implementing ML-based disease detection systems is their adaptability to real-world conditions. Environmental factors such as lighting variations, different growth stages, and overlapping symptoms of multiple diseases can introduce challenges in classification accuracy. Therefore, developing strong pre-processing techniques and data augmentation strategies is essential to improve model generalization. By employing techniques such as generative adversarial networks (GANs) for synthetic data creation and transfer learning for feature extraction, the system can enhance its resilience to variations in input images. Additionally, the integration of edge computing solutions can facilitate real-time disease detection

directly in the field, reducing the dependency on centralized computing resources and enabling rapid decision-making for farmers.

Future advancements in this domain can also benefit from multi-modal approaches that combine image-based analysis with other sensing technologies such as hyperspectral imaging and Internet of Things (IoT) sensors. These technologies can provide complementary information, such as leaf moisture levels, temperature variations, and disease progression rates, which can further refine detection accuracy. By fusing data from multiple sources, an intelligent system can make more informed predictions and assist farmers in implementing precise disease management strategies. Ultimately, the convergence of AI, remote sensing, and agricultural expertise has the potential to revolutionize plant disease management, ensuring higher productivity, sustainability, and food security in the pear industry.

Developing an automatic disease detection system for pear leaves involves multiple challenges, including data quality, model design, real-time processing, user adoption, and the need for well-labelled, diverse datasets for DL classification. Additionally, integrating explainable AI (XAI) is essential for understanding the model's decision-making process and fostering trust among users. A user-friendly and scalable system that maintains high reliability is important for practical deployment in agricultural settings. In this research, we propose an automatic disease detection system for pear leaves using a customized DL 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 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, ensuring that pear production remains a viable and profitable agricultural activity in the long term.

## II. RELATED WORKS

Recent advancements in DL have significantly improved the detection and classification of plant diseases, including those affecting pear plants. In the early stages of these developments, investigated multi-output learning approaches that enabled simultaneous prediction of different disease and stress parameters. Their work demonstrated that leveraging multiple outputs from a single model can lead to improved evaluation of plant health, as various indicators are considered together. Complementing this, explored the integration of attention mechanisms into CNN architectures. By incorporating the Convolutional Block Attention Module (CBAM), they were able to guide the network to focus on the most informative regions of an image, thereby refining the feature extraction process and contributing to better classification performance. These early studies laid the groundwork for more specialized models in plant disease detection. Addressing the challenges presented by natural, complex backgrounds in field images has been another focus of recent research. developed DBPNet, a model that modified MobileNetV2 to recognize pear leaf diseases even when images contain distracting background elements. Their approach was designed to overcome the variability and clutter inherent in natural settings, which often degrade model performance. At the same time, introduced a CycleGAN to generate synthetic images that mimic the diverse conditions seen in actual orchards. By incorporating several pre-trained models—such as VGG19, ResNet50, Inception-V3, MobileNet-V2, and EfficientNetB0—their method enhanced the model's ability to adapt to different lighting and environmental conditions, thereby improving the overall reliability of disease classification.

Feature extraction has been further enhanced by approaches that combine attention-based mechanisms and ensemble learning. proposed a deep attention layer-based convolutional learning (DAL_CL) model that extracts salient features from pear leaf images; these features are then classified using a recursive neural network (RNN) architecture. In parallel, demonstrated the benefits of using ensemble CNNs. By combining the outputs of multiple CNN models, they were able to achieve better accuracy in classifying pear leaf diseases compared to single-model approaches. This ensemble strategy mitigates individual model weaknesses and leverages complementary strengths, thereby contributing to improved disease detection performance. In addition to these complex methods, researchers have also focused on developing models that are practical for field applications. presented a lightweight algorithm based on an improved YOLOv5 model for recognizing pear leaf diseases in natural scenes. This model was specifically designed to operate on low-computing platforms, making it accessible for deployment in various agricultural settings where high-performance computing resources may not be available. The emphasis on lightweight design ensures that advanced DL techniques can be utilized in real-world environments without prohibitive computational costs. Further improvements have been achieved through the strategic use of transfer learning. employed CNNs in conjunction with transfer learning, enabling their model to benefit from pre-trained weights on large-scale datasets. This approach significantly reduced the need for extensive labeled data, which is often a limitation in agricultural research. Similarly, introduced a DL classification method that separates light and chroma components of an image. This innovative approach helped to mitigate the effects of varying lighting conditions, thereby enhancing the consistency of the classification outcomes across different environmental settings.

Building on these methods, proposed DeepPlantNet, a CNN-based model specifically designed for multi-disease recognition. DeepPlantNet can identify several plant diseases simultaneously by leveraging advanced network architectures and extensive training datasets. This multi-disease approach reflects the complex reality of agricultural environments, where plants may be affected by more than one type of disease at a time, and it underscores the importance of models that can handle such complexity. Analyzing the factors that influence disease recognition has also been an area of active research. conducted a detailed analysis of environmental variables, image quality, and disease progression stages that affect the performance of plant disease detection systems. Their work provided valuable insights into how external conditions and image capture quality can influence the effectiveness of DL models. Complementary to this, focused on optimizing network architectures and training strategies to better handle the diversity of plant species and disease manifestations. By refining the balance between model complexity and computational efficiency, their work contributed to better performance in multi-plant disease recognition tasks.

Data augmentation and domain adaptation have emerged as important techniques to address challenges such as limited and imbalanced datasets. employed various data augmentation techniques along with transfer learning and ensemble learning strategies. Their approach included the use of domain adaptation methods to align the training data more closely with real-world conditions, ensuring that the models perform consistently when deployed in different agricultural regions. By systematically tuning hyperparameters and integrating synthetic data generation methods, their work has advanced the field toward models that generalize better to unseen environments. Another critical aspect of plant disease detection is the accurate segmentation of disease-affected regions. proposed an innovative approach based on an improved DeepLabv3+ network model. By leveraging MobileNetV2 as the backbone, this method enhanced both computational efficiency and segmentation speed, which is essential for real-time applications in the field. In parallel, presented a comprehensive DL-based framework that utilizes advanced CNN architectures for both detection and classification of plant diseases. Their models were trained on the FieldPlant dataset, which consists of high-quality images captured under diverse field conditions. This framework has proven effective in performing accurate disease classification in realistic agricultural settings, thereby validating its potential for practical use.

Recent advancements have also explored models that integrate multiple DL architectures to leverage their complementary strengths. Several studies have combined CNNs with Vision Transformers (ViTs) to enhance feature extraction by capturing both local and global dependencies in leaf images. The incorporation of ViTs has enabled models to understand spatial relationships more effectively, which is particularly useful when disease symptoms are distributed across an image. Additionally, hybrid approaches that integrate CNNs with Long Short-Term Memory (LSTM) networks have been utilized to analyze sequential dependencies in disease progression over time. These integrative models have shown promise in achieving better accuracy and improved adaptability to both environmental variations and real-world agricultural conditions. Despite these promising developments, challenges persist.

One notable limitation is class imbalance, where certain diseases are underrepresented in the available training data. This imbalance can hinder the model's ability to detect less common disease types effectively. Furthermore, there is often insufficient focus on detailed leaf features—such as subtle texture variations, vein patterns, and slight discolorations—that are indicative of early-stage diseases. Addressing these gaps is essential for early detection, which in turn facilitates timely management interventions. To this end, the proposed work aims to tackle these research gaps by employing enhanced data augmentation techniques to balance class distributions and by developing adaptive learning mechanisms that account for diverse disease manifestations. Additionally, the proposed approach incorporates advanced feature extraction methods designed to capture and emphasize the intricate leaf characteristics essential for accurate disease detection and classification. Overall, the evolving body of research in DL for plant disease detection illustrates a clear trend toward integrating multiple methodologies and refining network architectures to address practical challenges in agriculture. The studies discussed above provide a strong foundation for developing more integrated, adaptable, and accurate disease detection systems in pear orchards. The proposed work, by addressing the identified limitations and incorporating advanced techniques, aspires to make a meaningful contribution to this growing literature while supporting sustainable practices in pear cultivation.

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

2. A major challenge in employing DL 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 DL models, potentially leading to erroneous diagnoses.

3. 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 important for identifying subtle disease symptoms.

### B. RESEARCH CONTRIBUTIONS

The following are the contributions made to address 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. Cross ViT tackles the issue of diverse presentations of plant diseases within a single class. This research explores the incorporation of CAMs into the Cross ViT architecture for plant leaf disease detection. CAMs have the potential to improve Cross ViT'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 RCSA and CAM. RCSA enhances inter-channel dependencies, while CAM focuses on modeling positional relationships within the images.

### III. PROPOSED SYSTEM

The proposed methodology consists of dual-track architecture. This architecture integrates the Cross ViT and a custom CNN with RCSA and CAM. 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 CAM, 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 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 CAM, which enhances and highlights the 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 RCSA and a CAM.

### 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 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 global context understanding, essential for tasks such as object detection and segmentation.

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

### RESIDUAL CROSS CHANNEL ATTENTION FRAMEWORK

The RC Attention Framework (Residual Cross Channel Attention) is an important 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 important for precise and reliable disease classification in agricultural settings. The RC Attention Framework incorporates RCSA and 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 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 CAM within the RC Attention Framework refines feature 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 CAM in the RC Attention Framework empowers CNN to extract hierarchical features with improved accuracy and efficiency. This advancement is important 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):

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

*(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 CAM introduces spatial information into the channel attention mechanism by encoding position-specific information through coordinate encoding. The attention computation is given by Eq. (2):

$$CA(x) = x \cdot \sigma (W_h \cdot AvgPool_h(x) + W_w \cdot AvgPool_w(x))$$

*(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 '$\sigma$' 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):

$$y = \alpha \cdot x + \beta \cdot RCSA(x)$$

*(3)*

Here, '$\alpha$' and '$\beta$' 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 CAM, as defined by Eq. (4):

$$y = Conv(CA(x)) = W_{conv} \cdot CA(x) + b$$

*(4)*

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

### CONDENSATION BLOCK

The Condensation Block (CB) module is an important component of the neural network architecture. It is strategically positioned immediately after the concatenation of Track 1 (Cross ViT) and Track 2 (RC Attention Framework). Its primary function is to enhance the performance of CNN 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 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 RC 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 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):

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

*(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):

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

*(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 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 RC Attention Framework, the condensed feature maps are processed through the classification pipeline. To manage the complexity and enhance model generalization, a strategy like 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 adaptability 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):

$$FL \quad (p_t) \quad = \quad - \quad \square_t \quad ((1-p_t)^{\square} \quad log(p_t))$$

(7)

$p_t$ is the model's estimated probability for the true class label, $\square_t$ is the class weights to balance the loss and $\square$ 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 better 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.

### DATA AUGMENTATION

This subsection discusses the various techniques used for data augmentation to enhance the adaptability 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 Figure 7. 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 epoch, while the training and validation accuracy graph indicates a significant improvement.

### B. ANALYSIS OF RC ATTENTION FRAMEWORK

To evaluate the role of the RC 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 training and validation accuracy are presented in Figure 8. The results indicate that the RC 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 epoch, 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.

### 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 Figure 9. These results demonstrate that combining 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.

### D. ANALYSIS OF PROPOSED WORK

To evaluate the overall performance of the proposed system, which integrates the Cross ViT, RC 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. Figure 10 illustrates 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 epoch, and the training and validation accuracy graph highlights the performance trends.

### 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 important features of the plant leaves, thus providing a more interpretable and reliable framework for plant disease detection.

## VI. DISCUSSION

This section presents an overview of the performance analysis, contrasting state-of-the-art networks with existing studies, while also examining the limitations of the current research framework and proposing directions for future work.

### 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 accuracy of 80.29% and 78.71%, respectively. VGG19 obtained an accuracy of 76.53%, while ResNet50 and AlexNet had lower accuracy 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.

### B. PERFORMANCE ANALYSIS WITH THE EXISTING STUDIES

The performance of the proposed approach is compared against existing studies on 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 CAM, 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.

### 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, labour-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 CAM 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 CAM, 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. The contributions of this research extend beyond accuracy improvements, addressing key challenges in plant disease classification. By integrating both convolutional and transformer-based feature extraction techniques, the proposed framework effectively captures diverse symptom patterns and intra-class variations. The use of CAMs enhances the model's ability to localize disease symptoms accurately, refining spatial attention to focus on the most informative regions of an image. This improvement is important for distinguishing between diseases with subtle visual differences, thereby reducing misclassification errors. Furthermore, the incorporation of RCSA strengthens inter-channel dependencies within the convolutional path, enabling better feature selection and improving the detection of fine-grained disease symptoms. The combination of these attention mechanisms with a dual-track structure enhances the model's generalization capabilities, making it more adaptive against variations in plant leaf appearances and environmental factors. This research provides a scalable and efficient solution for automated plant disease detection, offering practical applications in precision agriculture. By improving classification accuracy, disease localization, and feature extraction, this study contributes to advancing intelligent agricultural technologies, promoting early disease detection, and supporting more effective crop management strategies.