

Team 5_final_ - Copy.docx

by SENSE SECRETARY

Submission date: 27-Sep-2024 02:36PM (UTC+0530)

Submission ID: 2467115525

File name: Team_5_final_-_Copy.docx (129.65K)

Word count: 5206

Character count: 32129

2
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 Residual Cross Channel Attention Framework for Pear Leaf Disease Classification

ABSTRACT Pear plants are an essential crop in horticulture and 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 Convolutional Neural Network (CNN) based Residual Cross Channel Attention Framework (RCCAF). The Cross ViT path is designed to capture global features by leveraging the ability of the transformer to model long-term dependencies and complex feature representations. Simultaneously, RCCAF path integrates Residual Channel Shuffled Attention (RCSA) and coordinate attention module to extract and enhance local features effectively. The features from both paths are subsequently concatenated and passed through a condensation attention module. This module refines the features 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 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 proposed network achieves an accuracy level of 88.61%, significantly enhancing the performance of the proposed model. This novel strategy contributes to the management of diseases in horticulture in an effective way.

2
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 Mallocci 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 Mallocci 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 replicability 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 or global feature extraction independently, often resulting in suboptimal performance. However, a more effective approach would combine local details with global context, leading to improved model performance.
2. One of the primary challenges in utilizing deep learning for plant disease detection is addressing the issue 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.

FIGURE 1. ARCHITECTURE OF PROPOSED MODEL

3. In prior research, CNNs have shown limited ability to represent positional relationships and spatial information. This limitation restricts their capacity to fully comprehend global contextual information and effectively manage long-range dependencies, which are critical for detecting 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 RCCAF track helps to extract local features from the input image, while the transformer track extracts the global features.
2. Cross ViT transformer used in the proposed model tackles the issue of diverse presentations of plant diseases. By leveraging the Cross ViT architecture, which utilizes multi-scale feature extraction and cross-attention mechanisms, the model enhances its ability to capture and interpret variations in color changes, spot sizes, and distributions, as well as differing levels of wilting.
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 RCSA and coordinate attention module. 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 combines Cross ViT with a Residual Cross Channel Attention Framework (RCCAF). Integrating these two distinct models helps in achieving effective classification of pear leaf diseases. The Custom CNN is enhanced with Residual Channel Shuffled Attention and a coordinate attention module.

B. CROSS VISION TRANSFORMER BLOCK

The proposed architecture includes a key component known as the Cross ViT. It is designed to identify 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 final classification head utilizes the processed representations for better classification. This proves the ability of the Cross ViT to tackle complex image processing challenges. Figure 2 represents the architecture of Cross ViT, highlighting its distinctive components within the framework.

C. RESIDUAL CROSS CHANNEL ATTENTION FRAMEWORK

The RCCAF is designed to enhance feature extraction and discrimination in pear leaf images. This architecture incorporates methods designed to capture and enhance both spatial and channel-wise interdependence. These are essential for accurate and dependable disease classification in agricultural environments. The

RCCAF incorporates RCSA and Coordinate Attention Module. 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. RCSA is better equipped 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 RCCAF 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 RC Attention Framework empowers the CNN to extract hierarchical features with 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).

$$y = RCSA(x) = f(W_1 \cdot \text{Shuffle}(W_2 \cdot x)) + x \quad (1)$$

Here, 'x' represents the input feature map or image, 'W₁' and 'W₂' 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).

FIGURE 3. ARCHITECTURE OF THE RCCAF

$$CA(x) = x \cdot \sigma(W_h \cdot \text{AvgPool}_h(x) + W_w \cdot \text{AvgPool}_w(x)) \quad (2)$$

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).

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

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).

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

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) 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 (RCCAF). 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 RC Attention

FIGURE 4. ARCHITECTURE OF THE CONDENSATION BLOCK

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).

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

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).

$$y = f_c(y) = W_2 * \text{ReLU}(W_1 * y + b_1) + b_2 \quad (6)$$

Here, 'W₁' and 'W₂' are weight matrices, 'b₁' and 'b₂' 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 RC Attention Framework, the condensed feature maps are processed through the classification pipeline. To manage the complexity and enhance model generalization, a strategy similar ¹ 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 ³ 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).

$$FL(p_i) = - \alpha_i (1-p_i)^{\gamma} \log(p_i) \quad (7)$$

Here, ' p_i ' is the model's estimated probability for the true class label, ' α_i ' 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 ⁴ighting easier samples and emphasizing harder examples, thereby en ⁴riching the model's ability to classify pear leaf diseases accurately. The learn ⁴ 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

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. Due to the prevalent class imbalance in the dataset, the adoption of data augmentation was considered to be important. This helped 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 ³ 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 ⁷ 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.

3 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 ⁵ accuracy of 85.31%. The training and validation loss, along with the training and validation accuracy, are presented in the Figure 5. The results demonstrate that the Cross ViT block enhances the ability of the model 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 RCCAF

To evaluate the role of the RCCAF in refining the features extracted from the images, experiment is conducted by 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 Figure 6. The results indicate that the RCCAF effectively enhances the performance of the model by emphasizing important features and suppressing irrelevant information.

(A)

(B)

FIGURE 6. ANALYSIS OF THE RCCAF (A) LOSS (B) ACCURACY

C. ANALYSIS OF RCCAF AND Cross ViT

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 Figure 7. These results demonstrate that combining the RCCAF with Cross ViT components substantially enhances the learning capacity of the model. 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

The overall performance of the proposed system, which integrates the Cross ViT, RCCAF, CB is analyzed through the experiment. The training was done over 25 epochs, resulting in a testing accuracy of 88.61%. The precision, recall, and F1 score for the testing set were 88.62%, 88.61%, and 88.61%, respectively. Figure 8 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 better accuracy.

(A)

(B)

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

E. GRAD CAM VIASULISATION

The Grad-CAM visualization highlights the important regions influencing the performance of the trained model. By applying Grad-CAM, we were able to identify the specific areas within the images from the DiaMOS plant dataset that significantly impact 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.

VI. DISCUSSION

This section provides an overview of the performance analysis, comparing state-of-the-art networks with existing studies. It also examines the limitations of the current works and highlights the 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. While these models are highly effective for general-purpose image classification tasks, they are often limited in capturing the specific features required to distinguish 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.

(A) (B)

(C) (D)

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

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

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.

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

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

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 RSCA and coordinate attention mechanisms to enhance feature extraction. The Cross-ViT path captures long-range dependencies and global contextual information, while the RCCAF 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 dual-track model, incorporating advanced feature extraction and attention mechanisms, constitutes a substantial contribution to the field of automated plant disease detection.

ORIGINALITY REPORT

13%

SIMILARITY INDEX

3%

INTERNET SOURCES

12%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

1

R. Karthik, R. Menaka, M. V. Siddharth, Sameeha Hussain, Bala Murugan, Daehan Won. "A deep feature fusion network using residual channel shuffled attention for cassava leaf disease detection", Neural Computing and Applications, 2023

Publication

3%

2

R Karthik, R Menaka, S Ompirakash, P Bala Murugan, M Meenakashi, Sindhia Lingaswamy, Daehan Won. "GrapeLeafNet: A Dual-Track Feature Fusion Network with Inception-ResNet and Shuffle-Transformer for Accurate Grape Leaf Disease Identification", IEEE Access, 2024

Publication

3%

3

R. Karthik, Sameeha Hussain, Timothy Thomas George, Rashmi Mishra. "A dual track deep fusion network for citrus disease classification using group shuffle depthwise feature pyramid and Swin transformer", Ecological Informatics, 2023

Publication

2%

4	Amir Shachar. "Introduction to Algogens", Open Science Framework, 2024 Publication	1 %
5	Arvind Dagur, Dhirendra Kumar Shukla, Nazarov Fayzullo Makhmadiyarovich, Akhatov Akmal Rustamovich, Jabborov Jamol Sindorovich. "Artificial Intelligence and Information Technologies", CRC Press, 2024 Publication	1 %
6	isteonline.in Internet Source	1 %
7	export.arxiv.org Internet Source	<1 %
8	getsmallbusinessideas.blogspot.com Internet Source	<1 %
9	fjs.fudutsinma.edu.ng Internet Source	<1 %
10	ris.utwente.nl Internet Source	<1 %
11	scipost.org Internet Source	<1 %

Exclude quotes On
Exclude bibliography On

Exclude matches < 10 words