IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 23 December 2023, accepted 29 January 2024, date of publication 1 February 2024, date of current version 8 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3361044

RESEARCH ARTICLE

GrapeLeafNet: A Dual-Track Feature Fusion Network With Inception-ResNet and Shuffle-Transformer for Accurate Grape Leaf Disease Identification

R. KARTHIK[1], R. MENAKA[1], S. OMPIRAKASH[2], P. BALA MURUGAN[2], M. MEENAKASHI[2], SINDHIA LINGASWAMY[3], AND DAEHAN WON[4]

[1]Centre for Cyber Physical Systems (CCPS), Vellore Institute of Technology, Chennai 600127, India
[2]School of Electronics Engineering, Vellore Institute of Technology, Chennai 600127, India
[3]Department of Computer Applications, National Institute of Technology Tiruchirappalli, Tiruchirappalli 620015, India
[4]System Sciences and Industrial Engineering, Binghamton University, Binghamton, NY 13902, USA

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

ABSTRACT Grapes are a widely cultivated crop in the horticultural industry, renowned for their unique flavor and nutritional benefits. However, this crop is highly susceptible to various diseases that can cause significant reductions in yield and quality, resulting in considerable financial losses. Therefore, it is imperative to identify these diseases to effectively manage their spread. Traditionally, the identification of grape leaf diseases has relied on scientific expertise and observational skills. However, with the advent of deep learning methods, it is now feasible to recognize disease patterns from images of infected leaves. In this research, we propose a novel dual-track feature fusion network titled 'GrapeLeafNet' for detecting grape leaf disease. It employs a dual-track feature fusion approach, combining Inception-ResNet blocks with CBAM for local feature extraction and Shuffle-Transformer for global feature extraction. The first track uses Inception-ResNet blocks to represent features at multiple scales and map significant features, and CBAM captures significant spatial and channel dependencies. The second track employs Shuffle-Transformer to extract long-term dependencies and complex global features in images. The extracted features are then fused using Coordinate attention, enabling the network to capture both local and global contextual information. Experimental results on the Grape leaf disease dataset from Plant Village demonstrate the effectiveness of the proposed network, achieving an accuracy of 99.56%.

INDEX TERMS Grape leaf disease, deep learning, transformer, CNN, attention.

I. INTRODUCTION

Grapes constitute a significant crop to produce wine, with an estimated global production of approximately 77.4 million metric tonnes [1]. The leading grape-producing nations comprise China, Italy, the United States of America, France, and Spain [2]. Grapes are the third most valuable horticultural crop grown widely in the Mediterranean region [3]. Despite their global popularity, grapes are prone to diseases caused by fungi and bacteria, which can result in up to the maximum economic loss of 80% [4]. Out of the different diseases, Grape Black Rot, Grape Black Measles, and Isariopsis Leaf Spot are the most common diseases that significantly affect grape leaves currently [5], [6].

Grape Black Measles is caused by fungi namely Phaeomoniella spp. The typical symptoms of this disease are rounded, irregular spots that spread across the veins of the leaves. A few leaves affected by this disease can result in a severe fungal infection throughout the entire plant, emphasizing its severity [7]. Isariopsis leaf spot, or Grape Leaf

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao.

Blight, is caused by Pseudocerospora Vitis and develops on leaves. This disease exhibits light brown symptoms initially and dark brown symptoms later, depending on the season [8]. Grape Black Rot is caused by Guignardia Bidwellii and grows rapidly during humid and early spring conditions. The disease initially appears as black spots throughout the veins of the leaves and alters the chemical composition by increasing the sugar concentration [9], [10].

Conventional disease detection methods rely on manual inspection and subjective symptom assessment, which can lead to errors and delays in diagnosis [11]. DNA-based serological methods offer improved pathogen identification but are time-consuming and require specialized expertise [11]. As the morphological characteristics of diseased spots can vary considerably, there is a need for more efficient and accurate disease detection methods [12]. Computer vision techniques, coupled with high-resolution sensors, smartphone integration, and deep learning algorithms, have emerged as a promising approach for disease diagnosis [13], [14]. The ability of Deep learning to extract and analyze complex patterns from images makes it well-suited for automated disease detection. This research proposes a deep learning-based method for accurately diagnosing and classifying grape leaf diseases.

II. RELATED WORKS

In recent years, various methods have been proposed for detecting diseases in grape leaves. These methods can be classified into two categories: Traditional Machine Learning methods and Deep Learning.

Nitesh et al. presented an approach for grape leaf disease detection using Gray-Level Occurrence Matrix-based features and Support Vector Machine (SVM). This method transformed the input images into HSV color space for precise feature extraction [15]. Nasiri and Nejand proposed another method for grape leaf disease detection using Gray-Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), and Local Binary Pattern (LBP) based features [16]. A few methods were proposed based on segmenting the diseased portions of the leaves and then processing them further for classification. Jaisakthi et al. presented a Grabcut segmentation approach for grape leaf disease detection [17]. GLCM-based features were extracted from the segmented region for further classification. Javidan et al. employed k-means clustering to segment the region of interest for grape leaf disease detection [18]. Then, feature extraction is performed in different color spaces namely RGB, HSV, and LAB. SVM was used for classifying the type of disease. Adeel et al. applied a local contrast haze reduction technique to enhance the input grape leaf images for effective disease detection [19]. Color, geometric, and Local Binary Pattern based features were considered for classification using SVM. Another method was proposed by Shantkumari and Uma for grape leaf disease detection using histogram gradient features and Improvised K-Nearest Neighbor (KNN) [20]. Zamani et al. used Principal Component Analysis (PCA) to extract significant features for plant leaf disease detection and these extracted features are classified using Random Forest [44].

Researchers have increasingly adopted deep learning (DL) methods over machine learning (ML) methods based on manual feature extraction for grape leaf disease classification, as the latter is limited by small dataset sizes and requires more data for accurate classification. Despite the limited dataset size challenge in DL, data augmentation techniques are available to address this issue. Unlike ML, DL automatically extracts features, eliminating the need for domain expertise in feature selection.

Most of the recent works for grape leaf disease detection focus on implementing deep learning techniques, including pre-trained architectures and transfer learning. Some works use deep learning networks for feature extraction and traditional machine learning for classification. Thet et al. proposed the use of the VGG16 architecture as the base model for feature extraction in their grape leaf disease classification model [21]. The VGG16 model has a deep network structure, consisting of 16 convolutional layers and 3 fully connected layers, which can capture and extract complex features from the input images. Subsequently, SVM was applied to classify the extracted features into specific disease classes. Similarly, Jain and Periyasamy also adopted the VGG16 architecture to extract features from grape leaf images [22]. The extracted features were then trained with a Random Forest classifier that can handle multi-class classification tasks with high accuracy.

Huang et al. presented an ensemble model for grape leaf disease using VGG16, MobileNet, and AlexNet [23]. Peng et al. proposed another hybrid method for grape leaf disease detection with ResNet50 and ResNet101 as feature extractors and SVM for classification [24]. Many other studies have also utilized pre-trained deep learning-based architectures to classify grape leaf diseases. Ji et al. proposed a united model using GoogLeNet and ResNet architectures for grape leaf disease detection [25]. The United Model was employed to classify grape leaves as healthy or diseased, specifically identifying common diseases such as black rot, esca, and isariopsis leaf spot. Ashokkumar et al. proposed an approach for grape leaf disease detection using Faster Region-based Convolutional Neural Network [26]. This approach combined attention-based multilayer convolutional feature creation, object identification, & categorization. Fraiwan et al. presented a comprehensive study for grape leaf disease detection with 11 different networks comprising Googlenet, Inception, MobileNet, ResNet architectures, etc [27]. The study reported DarkNet architecture yielding the maximum accuracy when compared to other networks. Uttam presented a transfer learning-based approach for grape leaf disease detection based on EfficientNet architecture [29]. Rayhan and Setyohadi presented an approach for grape leaf disease detection using the VGG16 network [30].

Generative Adversarial Networks (GAN) were applied to generate synthetic images to combat the data imbalance issues. Liu et al. presented a GAN-based approach to generate images of four different grape leaf diseases for training identification models [28]. Tang et al. presented a lightweight approach for grape leaf disease detection using a channel-wise attention mechanism [31]. The network was proposed to diagnose grape diseases, including black rot, black measles, and leaf blight. Lauguico et al. presented a transfer learning approach with Alexnet based on Regions with Convolutional Neural Networks [32]. Kavala and Pothuraju analyzed the impact of the transfer learning approach with VGG16 and VGG19 networks for grape leaf disease detection [33]. Another work was proposed based for grape leaf disease detection based on the VGG network [34]. Xie et al. presented a deep learning-based approach for grape leaf disease detection using the Inception-v1 module, Inception-ResNet-v2 module, and SE-blocks [35]. Guo et.al implemented data augmentation using both rigid transformation and deep convolutional generative adversarial networks for feature extraction. Modified ResNet and GoogleNet are employed for classification [45]. In the recent years, Customized CNN are employed for plant's environmental concerns by incorporating attention networks and cross-layer extraction structures [40], [41], [42], [43], [46].

Although deep learning models have shown promising results in detecting grape leaf diseases, this research has highlighted some limitations that need to be addressed in the context of grape leaf disease detection which are tabulated in Table 1. These limitations include class imbalance, inadequate class-wise adaptability, and insufficient attention to critical leaf features. The proposed work will address these research gaps, and the following subsection will discuss the steps taken to address these limitations.

A. RESEARCH GAPS AND MOTIVATION

The following are the research gaps that were observed in the existing works related to grape leaf disease detection.

Previous studies have mainly focused on either local or global feature extraction, which can lead to suboptimal performance in detecting grape leaf diseases. To overcome this issue, it is recommended to use feature fusion techniques.

The model architectures proposed in existing research are often large and complex, making it difficult to deploy them on low-resource devices. The linear arrangement of convolutional layers in these models can result in intricate computations and a higher number of trainable parameters, limiting their practical usefulness in real-world applications.

CNNs are commonly utilized in previous studies due to their high accuracy in detecting and classifying various plant diseases. However, CNNs may not be suitable for capturing global contextual information and handling long-range dependencies. This limitation can be particularly problematic in cases where the disease symptoms are subtle and require a more comprehensive understanding of the context.

B. RESEARCH CONTRIBUTIONS

The following are the research contributions towards the proposed work.

The proposed network aims to improve the classification of grape 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 Shuffle Transformer to extract global features.

The proposed network incorporates Inception-ResNet for better feature extraction, resulting in reduced computation. This is achieved by performing feature fusion of different feature scale convolutions in parallel with input features.

The proposed network applies both CBAM and Coordinate Attention to capture both spatial and positional information in the image. Specifically, the CBAM module can capture interdependencies between different feature channels, while the Coordinate Attention can capture positional relationships between different features.

The Shuffle Transformer track used in the proposed network is effective in capturing long-range dependencies in images. By incorporating channel shuffling techniques and utilizing both global and local information, it can capture complex features that may be important for identifying subtle symptoms of grape leaf diseases.

III. PROPOSED SYSTEM

This section presents an overview of the different modules employed in the proposed network.

A. ARCHITECTURE OVERVIEW

This section presents an overview of the different modules employed in the proposed network.

The proposed network 'GrapeLeafNet' has two distinct feature extraction tracks for precise classification of grape leaf disease detection. The first track consists of a linear CNN integrated with modules like Inception-ResNet block and Convolutional Block Attention Module (CBAM). This track uses Inception-ResNet blocks to achieve an optimal balance between performance and computational efficiency by employing filters of various sizes, such as 1×1, and 3×3, in parallel within the same layers. CBAM integrates the Channel Attention and Spatial Attention mechanisms, enabling the network to selectively focus on informative channels and spatial regions of the input feature maps, which in turn improves the network's feature representation capability. The second track involves the use of a Shuffle transformer to effectively extract the long-range dependencies. The features from the two different tracks were concatenated and subjected to the Coordinate Attention module to further enhance the feature representation for precise classification. The architecture of the proposed network is presented in Figure 1.

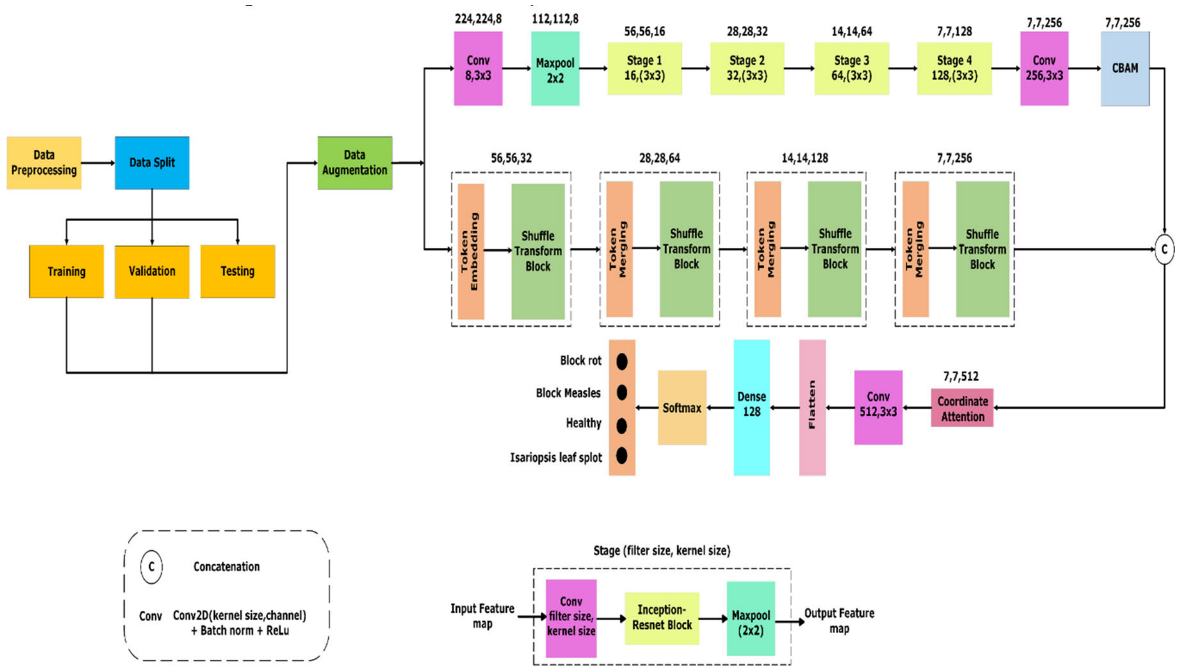


FIGURE 1. Architecture of the proposed network.

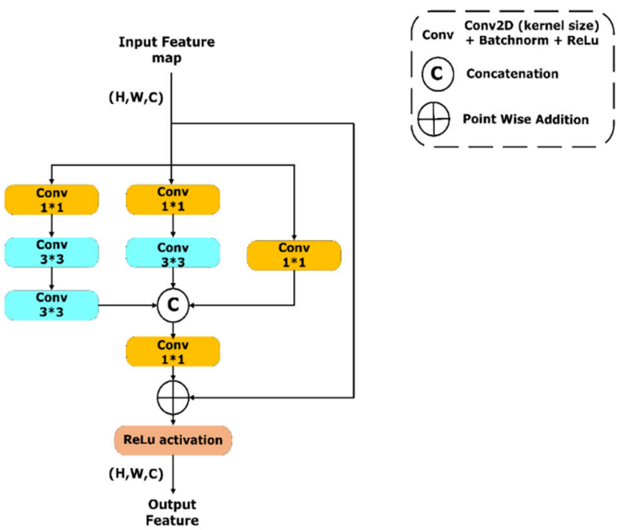


FIGURE 2. Schematic overview of inception-Resnet.

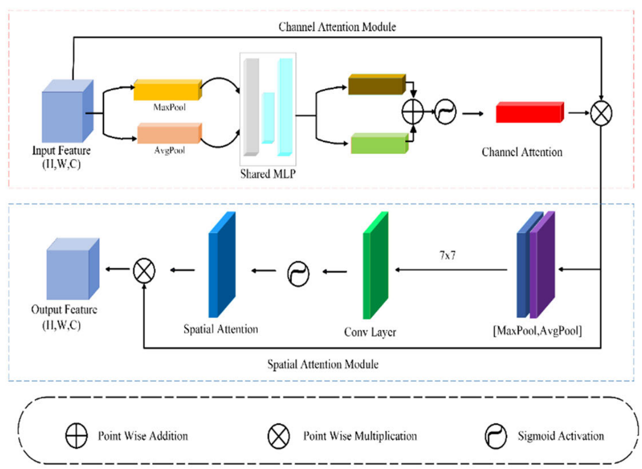


FIGURE 3. Architecture of the CBAM block.

B. INCEPTION-RESNET BLOCK

Convolutional Neural Networks (CNNs) generally require larger convolutional layers to improve feature extraction and classification performance. However, increasing the number of layers in the network can lead to longer computation times. Conversely, a model designed for the low computational cost may not achieve satisfactory classification accuracy. To address these challenges, the Inception-ResNet block has been incorporated into the baseline CNN module [36]. The Inception-ResNet block employs a series of parallel convolutional layers with different kernel sizes to extract features at multiple scales from the input tensor. The outputs of these parallel layers are concatenated along the channel dimension, generating a high-dimensional feature map. Next, a bottleneck layer consisting of 1×1 convolutions reduces the feature map's dimensionality. Figure 2 illustrates the architecture of the Inception-ResNet block.

C. CONVOLUTIONAL BLOCK ATTENTION MODULE (CBAM)

The CBAM module is a neural network component that enhances the performance of convolutional neural networks (CNNs) by focusing on relevant image features [37]. It comprises two modules: the channel attention module and the spatial attention module. The channel attention module uses

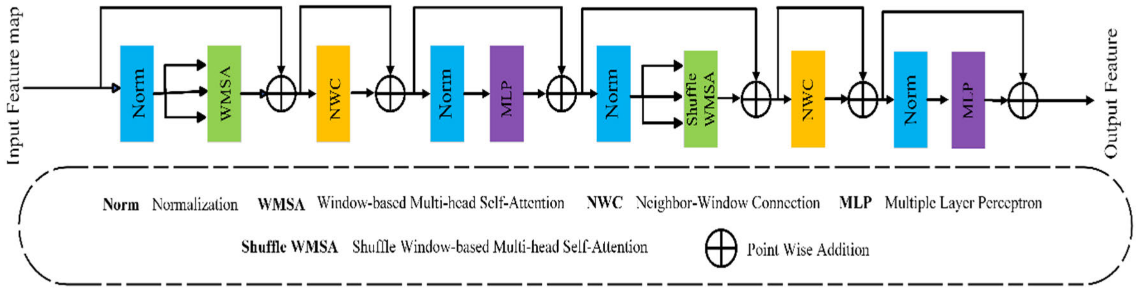


FIGURE 4. Architecture of the shuffle-transformer block.

average pooling and max pooling to obtain the global information of each channel. This information is utilized to calculate channel-wise attention weights by concatenating the average and maximum pooling values and passing them through several fully connected layers. These weights are then employed to highlight the essential channels and suppress the less significant ones by rescaling the channel features.

The spatial attention module of the CBAM module follows a similar approach to the channel attention module, but it rescales the spatial features rather than the channel features. The spatial attention module applies average pooling and max pooling to the feature maps of each channel to determine the spatial attention weights. These weights are then used to rescale the spatial features by emphasizing the significant spatial locations and reducing the unimportant ones. By integrating the channel and spatial attention modules, the CBAM module can selectively focus on important features, resulting in improved performance in tasks such as image classification. Figure 3 illustrates the architecture of the CBAM.

D. SHUFFLE TRANSFORMER

The Shuffle Transformer is a convolutional neural network (CNN) that employs three modules to learn global and local relationships between feature maps in an image [38]. The first module is the window-based multi-head self-attention module (WMSA), which converts the feature maps into a 2D grid and uses self-attention over the rows and columns to capture long-range dependencies between different parts of the image. The second module is the neighbor window connection module (NWC), which connects the feature maps of neighboring pixels within a small window to capture local relationships and improve performance while reducing computational complexity. Finally, the shuffle window-based multi-head self-attention module (shuffle-WMSA) is used to shuffle the channels after applying self-attention to capture diverse information from the feature maps and improve the network's ability to detect complex features.

The combination of these three modules allows the Shuffle Transformer to achieve state-of-the-art performance in image classification. By using a combination of global and local information and leveraging channel shuffling techniques, the Shuffle Transformer is capable of capturing complex features in images efficiently and accurately. Figure 4 illustrates the architecture of the Shuffle-Transformer.

E. COORDINATE ATTENTION MECHANISM BLOCK

The Coordinate Attention Block is a convolutional neural network module that attends to features selectively, based on their spatial locations in an image [39]. This is achieved by separately applying global pooling along the x and y axes to capture the mean and variance of the feature maps along each axis. Specifically, the module computes the mean of the feature maps at each position along the x-axis and y-axis through global average pooling. These two mean values are then concatenated and fed into a feedforward network to calculate attention weights for each spatial location. The attention weights are then utilized to rescale the original feature maps, highlighting the significant locations and de-emphasizing the less important ones. The Coordinate Attention Block selectively focuses on critical spatial locations in an image, leading to improved performance of CNNs in various computer vision tasks, such as object detection, image classification, and semantic segmentation. Additionally, the module requires fewer parameters than other attention mechanisms, resulting in computational efficiency and ease of implementation in CNN architectures. The following architecture is represented in Figure 5.

IV. RESULTS AND DISCUSSION

This section presents an overview of the dataset description, data augmentation techniques, experimental setup, model training, and validation. The final subsection evaluates the performance of the model using several performance metrics for each ablation experiment performed.

A. DATASET DESCRIPTION

This study utilized the grape leaf dataset, which includes 4062 labeled images with dimensions of 256 by 256 pixels, sourced from the plant village dataset. The grape leaf dataset is separated into two subsets, healthy and diseased, and further categorized into three distinct diseases: Black Rot, Black

TABLE 1. Sample images for each class from the plant village dataset.

Input class	Image samples		
Healthy			
Black Measles			
Black Rot			
Isariopsis leaf spot			

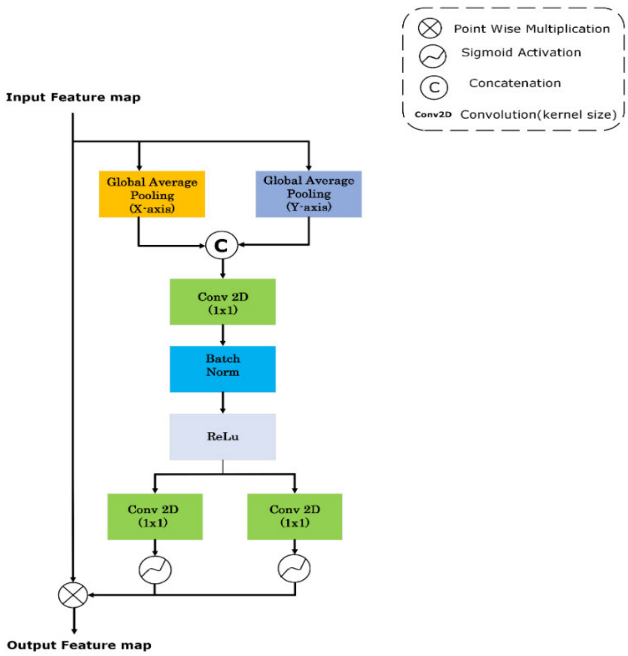


FIGURE 5. Architecture of the coordinate attention block.

Measles, and Isariopsis leaf spot. Table 1 presents the sample images for each class.

B. DATA AUGMENTATION

Data augmentation is a technique used to increase dataset diversity for leaf disease detection. It creates new data points by applying transformations such as rotations, flips, scaling, and changes in brightness and contrast. This technique can help to balance an imbalanced dataset and improve classifier accuracy. It can also improve the generalization of the classifier by exposing it to a wider range of variations in the appearance of diseased leaves. Overall, data augmentation is useful for improving the accuracy and robustness of the classifier in limited or imbalanced datasets. It can help distinguish between healthy and diseased leaves in real-world scenarios. Data augmentation is done in this work using various techniques to create new data points from existing ones. These techniques include: (1) rotating the image randomly between 0 to 40 degrees, (2) shearing the image randomly between 0 to 0.2, (3) shifting the width of the image randomly between 0 to 0.2, (4) flipping the image randomly horizontally and vertically, (5) adjusting the brightness randomly between 0.5 to 1.5, and (6) applying a random gaussian blur with sigma values between 0 to 0.8.

The dataset was divided into three distinct sets, with the training set containing 60% of the data, the validation set containing 20%, and the remaining 20% comprising the test set. After the split, augmentation was carried out on both the training and validation sets.

C. ENVIRONMENTAL SETUP

The proposed network was developed utilizing PyTorch, which is an open-source deep-learning framework. The model employed the Adam optimization technique, with a learning rate of 0.0001 and weight decay of 0.0001, to minimize the loss. The training process was carried out on an Azure virtual machine that was equipped with an NVIDIA Tesla P40 GPU, specifically designed for deep learning tasks.

D. EVALUATION METRICS

The following metrics are evaluated in this context: Accuracy, Precision, Recall, F1 Score, Specificity, and Area Under Curve (AUC). These parameters are used to determine the quality of a model. Accuracy is a statistical metric that measures how accurately a predicted value matches its actual value. Precision is the ratio of true positives to the total number of positive predictions made by the model. Specificity measures the model's ability to correctly identify true negative values, and it is calculated as the ratio of true negatives to the total number of negatives. Recall is the ratio of true positives to the total number of positive samples in the dataset. The F1 score is a combination of precision and recall and measures the model's ability to identify positive samples. The evaluation metrics are determined using the True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values, which are expressed in (1)-(5) below. Overall, these metrics help in evaluating the performance of a model and identifying areas for improvement. The area under Receiver Operating Characteristic Curve (ROC) is used to describe the probabilistic variation between recall and False Positive.

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{1}$$

$$\text{Precision} = \frac{T_p}{T_p + F_p} \tag{2}$$

$$\text{Recall} = \frac{T_p}{T_p + F_n} \tag{3}$$

$$\text{F1 Score} = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

$$\text{Specificity} = \frac{T_n}{T_n + F_p} \tag{5}$$

E. ABLATION STUDIES

Ablation studies are used in CNN to analyze the importance of different components by removing or disabling them. This helps to understand their individual contributions to the model's overall performance and identify areas for improvement. The proposed model comprises of a 7-layer CNN as the base model, integrated with several other blocks including Inception-ResNet, Convolutional Block Attention Module (CBAM), Shuffle Transformer block, and coordinate attention block. Sequential analysis was performed to understand the impact of integrating these blocks into the baseline CNN. Ablation studies were conducted on each model to assess the effect of each block on the network's performance. An experiment was then conducted to measure the accuracy, precision, recall, F1 score, and area under the ROC curve for each model sequentially, and the results were presented in Table 2.

1) ANALYSIS OF THE 7-LAYER LINEAR CNN

The performance of the baseline 7-layer CNN network was evaluated in this experiment. The linear layered CNN structure, without additional modules, was trained for 120 epochs. The proposed network architecture consists of convolutional blocks followed by batch normalization and Rectified Linear Unit (ReLU) activation functions. Three out of the seven convolutional blocks are accompanied by max pooling and dropout layers to enhance the model's generalization ability. Fig. 6 illustrates the performance and ROC curve of the baseline 7-layer CNN is presented. The evaluation of the trained baseline model on a separate dataset yielded an accuracy of 92.05%.

2) EFFECTIVENESS OF THE INCEPTION-ResNet

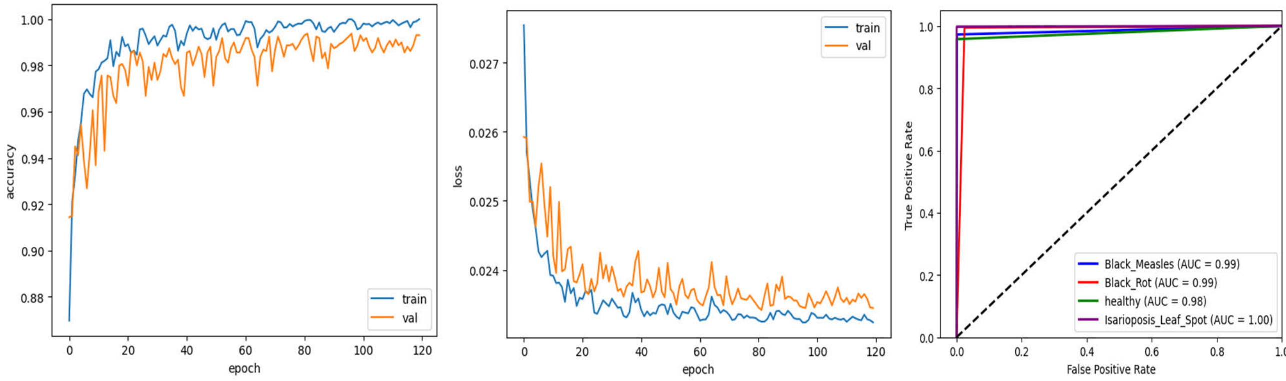
This study investigates the impact of incorporating Inception-ResNet blocks into the linear CNN model used in the previous experiment. By utilizing parallel convolution with different kernel sizes, Inception-ResNet blocks lower the network's computational expense and enhance feature extraction by combining the input image with multiple parallel convoluted images. The trained model achieves a test set accuracy of 98.50%, with additional metric results provided in Table 2. Figure 7 displays the model's performance and ROC curve.

3) EFFECTIVENESS OF THE CBAM BLOCK

In this study, the impact of incorporating a CBAM block into the network used in the previous experiment is analyzed. CBAM enhances spatial and channel features, leading to improved accuracy and generalization for significant feature extraction. After training the model with CBAM, it achieved an enhanced accuracy of 98.75% on a new testing dataset, which is 0.25% higher than the previous ablation study. Figure 8 presents the accuracy, loss, and ROC curve of this proposed model, while the other parameters are presented in Table 2.

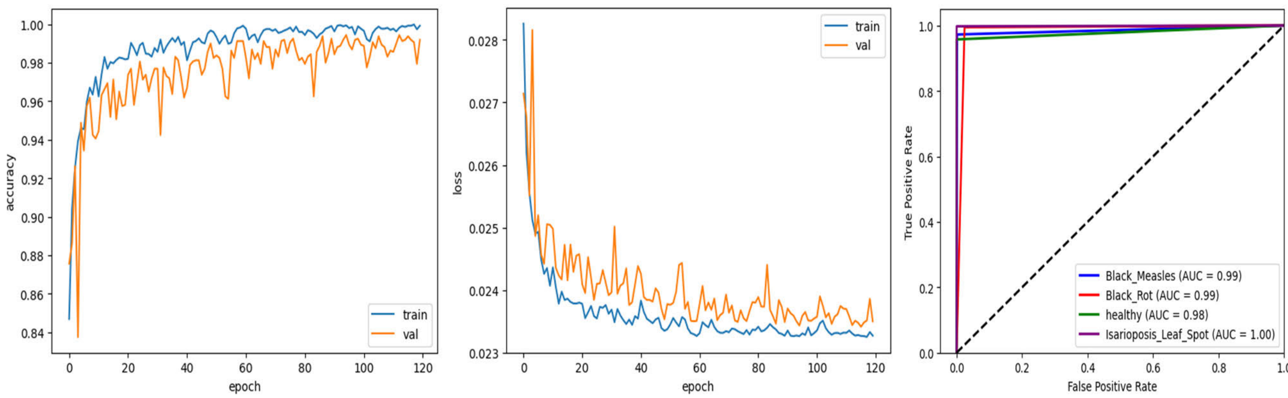
4) ANALYSIS OF THE SHUFFLE TRANSFORMER BLOCK

This study examines how incorporating the shuffle transformer track affects the final classification.



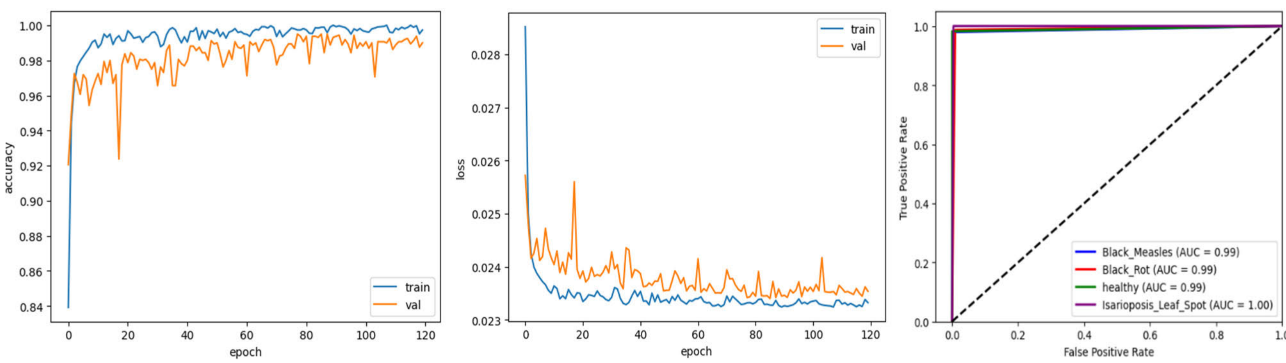
(a) (b) (c)

FIGURE 6. Analysis of the baseline CNN network (a) Accuracy (b) Loss (c) ROC curve.



(a) (b) (c)

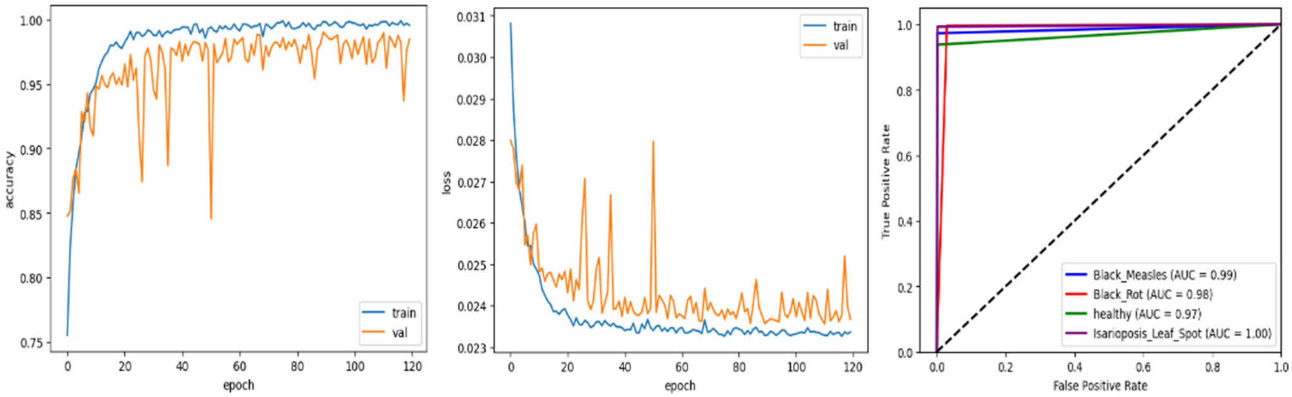
FIGURE 7. Analysis of the baseline CNN network + Inception-Resnet Block (a) Accuracy (b) Loss (c) ROC curve.



(a) (b) (c)

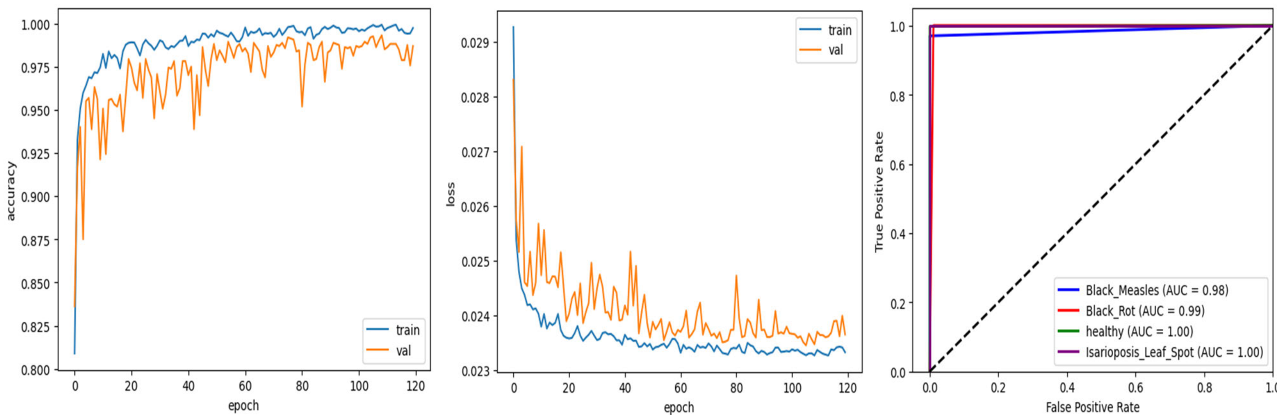
FIGURE 8. Analysis of the baseline CNN + Inception-Resnet + CBAM module Block (a) Accuracy (b) Loss (c) ROC curve.

By utilizing global and local information and incorporating channel shuffling techniques, the Shuffle Transformer can effectively and accurately capture complex features in images. The network was trained for 120 epochs and achieved a test set accuracy of 97.44%. Figure 9 displays the accuracy, loss, and ROC curve of this proposed model, while the other parameters are listed in Table 2.



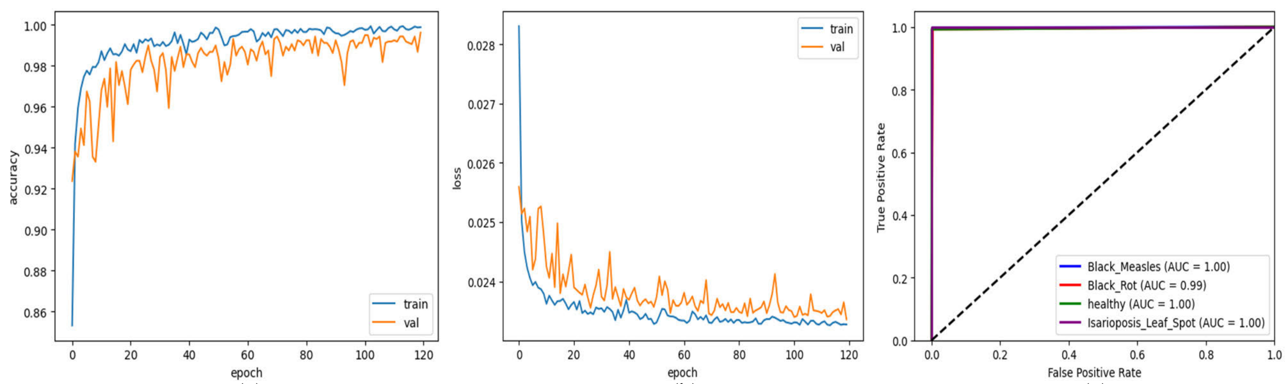
(a) (b) (c)

FIGURE 9. Analysis of the shuffle transformer track (a) Accuracy (b) Loss (c) ROC curve.



(a) (b) (c)

FIGURE 10. Analysis of the proposed network without CA Block (a) Accuracy (b) Loss (c) ROC curve.



(a) (b) (c)

FIGURE 11. Analysis of the proposed model (a) Accuracy (b) Loss (c) ROC curve.

5) ANALYSIS OF THE PROPOSED NETWORK WITHOUT CA BLOCK

This study investigates the impact of concatenating the features extracted from two distinct branches, namely the CNN track (Section IV-E3) and the shuffle transformer track (Section IV-E4). The features were combined and trained without utilizing Coordinate attention. The resulting accuracy for the test set was 99.19%, which represents an improvement of almost 0.44% over all previous ablation studies. Figure 10 illustrates the accuracy, loss, and ROC

TABLE 2. Sample images for each class from the plant village dataset.

Model	Accuracy	Precision	Recall	F1-Score	Specificity	AUC
Baseline CNN	0.9206	0.9217	0.9206	0.9208	0.9225	0.9201
Baseline CNN + Inception-Resnet	0.9850	0.9856	0.9850	0.9851	0.9750	0.9904
Baseline CNN + Inception-Resnet + CBAM	0.9875	0.9877	0.9875	0.9875	0.9725	0.9925
Shuffle Transformer	0.9744	0.9759	0.9744	0.9745	0.9725	0.9850
Baseline CNN + Inception-Resnet + CBAM + Shuffle Transformer	0.9919	0.9921	0.9919	0.9919	0.9750	0.9925
Proposed Network	**0.9956**	**0.9956**	**0.9956**	**0.9956**	**0.9975**	**0.9975**

curve of this proposed model, and Table 2 lists the other parameters.

6) ANALYSIS OF THE PROPOSED NETWORK

The network used in the previous experiment was modified by adding coordinate attention and then trained for 120 epochs. This architecture comprises a linear CNN as the base model, Inception-ResNet to reduce the computational cost, CBAM to enhance generalization by improving spatial and channel features, Shuffle Transformer to extract long-range dependency features, and coordinate attention to focus on the x-axis and y-axis using global average pooling (GAP) to provide precise feature weighting. The proposed model achieved an improved accuracy of 99.59% on the test set, which is a 0.40% increase compared to the previous model without coordinate attention. Figure 11 shows the accuracy, loss, and ROC for the proposed model, and additional metrics are presented in Table 2.

F. PERFORMANCE ANALYSIS

Table 3 presents the performance evaluation between the existing works and the proposed model. Existing works predominantly concentrated on pre-trained architectures, which performed modestly on a limited amount of data. The proposed model integrates various components, such as Inception-ResNet, Convolutional Block Attention Module (CBAM), Shuffle Transformer, and Coordinate Attention mechanism. Transfer learning-based approaches were quite successful in disease detection and these methods have an accuracy in the range of 95 to 98%. Although transfer learning can be an effective method to achieve high accuracy rates, it is crucial to acknowledge that pre-trained architectures may not always be suitable for every task. In contrast to prior studies that did not incorporate attention or context mechanisms, the proposed customized CNN model was more successful in identifying grape leaf diseases. The proposed system leverages the connections among features and contextual information, allowing it to acquire features with greater accuracy and attain higher levels of precision. Additionally, Table 4 illustrates a comparison between the number of trainable parameters in the proposed network and those in state-of-the-art architectures. The analysis indicates that the

TABLE 3. Comparative analysis of the performance of the proposed work with that of other existing works.

Sl.no	Source	Methodology	Accuracy (in %)
1	Yohan Rayhan et al. [30]	VGG16	87.5
2	Sajjad Nasari et al. [16]	GLCM with SVM	89.93
3	Alishba Adeel et al. [19]	Low contrast haze reduction-neighbourhood component analysis with SVM	91.34
4	Arie Hasan et al. [34]	VGG16	95
5	Adi et al. [33]	VGG6 and VGG19	98
6	Khaing et al. [21]	VGG16	98.4
7	Miaomiao et al. [25]	GoogleNet & ResNet50	98.57
8	Zhe et al. [31]	AlexNet	99.01
9	Xiaoyue et al. [35]	Inception-Squeeze and Excitation-Resnet Model	99.47
10	**Proposed network**	**Dual track deep feature network**	**99.59**

TABLE 4. Parameter analysis of the proposed model.

Model	No. of Parameters	Accuracy	Precision	F1-Score	Recall
VGG16	134,276,932	0.9444	0.9444	0.9472	0.9439
AlexNet	56,337,156	0.9500	0.9500	0.9528	0.9504
Inception-ResNetV3	54,342,884	0.9844	0.9844	0.9849	0.9844
ResNet50	23,595,908	0.9550	0.9550	0.9567	0.9548
InceptionV3	21,810,980	0.9675	0.9675	0.9696	0.9671
Xception	20,869,676	0.9781	0.9781	0.9794	0.9781
Proposed network	**8,836,697**	**0.9956**	**0.9956**	**0.9956**	**0.9956**

proposed network outperforms its counterparts while using fewer trainable parameters.

V. CONCLUSION

Grapes offer numerous nutritional benefits, including high levels of antioxidants, vitamins, and minerals. However, diverse climatic conditions during grape cultivation make them susceptible to different diseases. Current disease detection methods rely on manual visual recognition of symptoms, leading to delayed treatment and potential crop loss. Advanced computer vision techniques can address this challenge by providing automatic and accurate disease classification. This research proposes a novel network for grape leaf disease detection with two feature extraction tracks for precise classification. The first track utilizes Inception-ResNet blocks and CBAM to achieve optimal balance and enhance feature representation. CBAM integrates Channel and Spatial Attention mechanisms to focus on informative channels and spatial regions. The second track employs a Shuffle transformer to effectively extract long-range dependencies. The features from both tracks are combined and subjected to the Coordinate attention module to further refine feature representation for precise classification. The proposed model, trained on the Grape leaf disease dataset from Plant Village, achieves an accuracy of 99.56%. The proposed network is limited to detect disease patterns acquired from constrained leaf portions under experimental setup. Though it performs well for this constrained setup, as an extension, further research is required to detect diseases from large scale landscape images acquired through drones for real-time deployment. Other potential future research directions include refining segmentation algorithms to improve the accuracy of outlining disease markers in grape leaf images and extending the proposed framework to other plant types, enabling computer-aided disease classification across various crops.

REFERENCES

[1] C. Venkitasamy, L. Zhao, R. Zhang, and Z. Pan, "'Grapes," in *Integrated Processing Technologies for Food and Agricultural By-Products*. Amsterdam, The Netherlands: Elsevier, 2019, pp. 133–163, doi: 10.1016/b978-0-12-814138-0.00006-x.

[2] A. Seccia, F. G. Santeramo, and G. Nardone, "Trade competitiveness in table grapes," *Outlook Agricult.*, vol. 44, no. 2, pp. 127–134, Jun. 2015, doi: 10.5367/oa.2015.0205.

[3] J. M. Alston and O. Sambucci, *Grapes in the World Economy* (Compendium of Plant Genomes). Cham, Switzerland: Springer, 2019, pp. 1–24, doi: 10.1007/978-3-030-18601-2_1.

[4] S. G. Crandall, J. Spychalla, U. T. Crouch, F. E. Acevedo, R. P. Naegele, and T. D. Miles, "Rotting grapes don't improve with age: Cluster rot disease complexes, management, and future prospects," *Plant Disease*, vol. 106, no. 8, pp. 2013–2025, Aug. 2022, doi: 10.1094/pdis-04-21-0695-fe.

[5] K. Sanghavi, M. Sanghavi, and A. M. Rajurkar, "Early stage detection of downey and powdery mildew grape disease using atmospheric parameters through sensor nodes," *Artif. Intell. Agricult.*, vol. 5, pp. 223–232, Apr. 2021, doi: 10.1016/j.aiia.2021.10.001.

[6] J. Zhu, A. Wu, X. Wang, and H. Zhang, "Identification of grape diseases using image analysis and BP neural networks," *Multimedia Tools Appl.*, vol. 79, nos. 21–22, pp. 14539–14551, Jan. 2019, doi: 10.1007/s11042-018-7092-0.

[7] M. Ji and Z. Wu, "Automatic detection and severity analysis of grape black measles disease based on deep learning and fuzzy logic," *Comput. Electron. Agricult.*, vol. 193, Feb. 2022, Art. no. 106718, doi: 10.1016/j.compag.2022.106718.

[8] D. Molitor and B. Berkelmann-Loehnertz, "Simulating the susceptibility of clusters to grape black rot infections depending on their phenological development," *Crop Protection*, vol. 30, no. 12, pp. 1649–1654, Dec. 2011, doi: 10.1016/j.cropro.2011.07.020.

[9] N. Kellner, E. Antal, A. Szabó, and R. Matolcsi, "The effect of black rot on grape berry composition," *Acta Alimentaria*, vol. 51, no. 1, pp. 126–133, Feb. 2022, doi: 10.1556/066.2021.00195.

[10] L. W. Kuswidiyanto, H.-H. Noh, and X. Han, "Plant disease diagnosis using deep learning based on aerial hyperspectral images: A review," *Remote Sens.*, vol. 14, no. 23, p. 6031, Nov. 2022, doi: 10.3390/rs14236031.

[11] A. Sinha and R. S. Shekhawat, "A novel image classification technique for spot and blight diseases in plant leaves," *Imag. Sci. J.*, vol. 68, no. 4, pp. 225–239, May 2020, doi: 10.1080/13682199.2020.1865652.

[12] A. Ali, S. Ali, M. Husnain, M. M. S. Missen, A. Samad, and M. Khan, "Detection of deficiency of nutrients in grape leaves using deep network," *Math. Problems Eng.*, vol. 2022, pp. 1–12, Jan. 2022, doi: 10.1155/2022/3114525.

[13] A.-K. Mahlein, "Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping," *Plant Disease*, vol. 100, no. 2, pp. 241–251, Feb. 2016, doi: 10.1094/pdis-03-15-0340-fe.

[14] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers Plant Sci.*, vol. 7, Sep. 2016, doi: 10.3389/fpls.2016.01419.

[15] N. Agrawal, J. Singhai, and D. K. Agarwal, "Grape leaf disease detection and classification using multi-class support vector machine," in *Proc. Int. Conf. Recent Innov. Signal Process. Embedded Syst. (RISE)*, Oct. 2017, pp. 238–244, doi: 10.1109/RISE.2017.8378160.

[16] S. Nasiri and M. Z. Nejand, "A method based on image processing for the automatic diagnosis of grapevine leaf disease," *Biosystem Eng. Iran*, vol. 53, no. 1, Apr. 2022, doi: 10.22059/ijbse.2022.327192.665432.

[17] S. M. Jaisakthi, P. Mirunalini, D. Thenmozhi, and Vatsala, "Grape leaf disease identification using machine learning techniques," in *Proc. Int. Conf. Comput. Intell. Data Sci. (ICCIDS)*, Feb. 2019, pp. 1–6, doi: 10.1109/ICCIDS.2019.8862084.

[18] S. M. Javidan, A. Banakar, K. A. Vakilian, and Y. Ampatzidis, "Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning," *Smart Agricult. Technol.*, vol. 3, Feb. 2023, Art. no. 100081, doi: 10.1016/j.atech.2022.100081.

[19] A. Adeel, M. A. Khan, M. Sharif, F. Azam, J. H. Shah, T. Umer, and S. Wan, "Diagnosis and recognition of grape leaf diseases: An automated system based on a novel saliency approach and canonical correlation analysis based multiple features fusion," *Sustain. Comput., Informat. Syst.*, vol. 24, Dec. 2019, Art. no. 100349, doi: 10.1016/j.suscom.2019.08.002.

[20] M. Shantkumari and S. V. Uma, "Grape leaf image classification based on machine learning technique for accurate leaf disease detection," *Multimedia Tools Appl.*, vol. 82, no. 1, pp. 1477–1487, Apr. 2022, doi: 10.1007/s11042-022-12976-z.

[21] K. Z. Thet, K. K. Htwe, and M. M. Thein, "Grape leaf diseases classification using convolutional neural network," in *Proc. Int. Conf. Adv. Inf. Technol. (ICAIT)*, Nov. 2020, pp. 147–152, doi: 10.1109/ICAIT51105.2020.9261801.

[22] B. Jain and S. Periyasamy, "Grapes disease detection using transfer learning," 2022, *arXiv:2208.07647*.

[23] Z. Huang, A. Qin, J. Lu, A. Menon, and J. Gao, "Grape leaf disease detection and classification using machine learning," in *Proc. Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData) IEEE Congr. Cybermatics (Cybermatics)*, Nov. 2020, pp. 870–877, doi: 10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics50389.2020.00150.

[24] Y. Peng, S. Zhao, and J. Liu, "Fused-deep-features based grape leaf disease diagnosis," *Agronomy*, vol. 11, no. 11, p. 2234, Nov. 2021, doi: 10.3390/agronomy11112234.

[25] M. Ji, L. Zhang, and Q. Wu, "Automatic grape leaf diseases identification via UnitedModel based on multiple convolutional neural networks," *Inf. Process. Agricult.*, vol. 7, no. 3, pp. 418–426, Sep. 2020, doi: 10.1016/j.inpa.2019.10.003.

[26] K. Ashokkumar, S. Parthasarathy, S. Nandhini, and K. Ananthajothi, "Prediction of grape leaf through digital image using FRCNN," *Meas., Sensors*, vol. 24, Dec. 2022, Art. no. 100447, doi: 10.1016/j.measen.2022.100447.

[27] M. Fraiwan, E. Faouri, and N. Khasawneh, "Multiclass classification of grape diseases using deep artificial intelligence," *Agriculture*, vol. 12, no. 10, p. 1542, Sep. 2022, doi: 10.3390/agriculture12101542.

[28] B. Liu, C. Tan, S. Li, J. He, and H. Wang, "A data augmentation method based on generative adversarial networks for grape leaf disease identification," *IEEE Access*, vol. 8, pp. 102188–102198, 2020, doi: 10.1109/access.2020.2998839.

[29] A. K. Uttam, "Grape leaf disease prediction using deep learning," in *Proc. Int. Conf. Appl. Artif. Intell. Comput. (ICAAIC)*, May 2022, pp. 369–373, doi: 10.1109/ICAAIC53929.2022.9792739.

[30] Y. Rayhan and D. B. Setyohadi, "Classification of grape leaf disease using convolutional neural network (CNN) with pre-trained model VGG16," in *Proc. Int. Conf. Smart Gener. Comput., Commun. Netw. (SMART GENCON)*, Oct. 2021, pp. 1–5, doi: 10.1109/SMARTGENCON51891.2021.9645862.

[31] Z. Tang, J. Yang, Z. Li, and F. Qi, "Grape disease image classification based on lightweight convolution neural networks and channelwise attention," *Comput. Electron. Agricult.*, vol. 178, Nov. 2020, Art. no. 105735, doi: 10.1016/j.compag.2020.105735.

[32] S. Lauguico, R. Concepcion, R. R. Tobias, A. Bandala, R. R. Vicerra, and E. Dadios, "Grape leaf multi-disease detection with confidence value using transfer learning integrated to regions with convolutional neural networks," in *Proc. IEEE REGION 10 Conf. (TENCON)*, Nov. 2020, pp. 767–772, doi: 10.1109/TENCON50793.2020.9293866.

[33] S. A. P. N. Kavala and R. Pothuraju, "Detection of grape leaf disease using transfer learning methods: VGG16 & VGG19," in *Proc. 6th Int. Conf. Comput. Methodologies Commun. (ICCMC)*, Mar. 2022, doi: 10.1109/iccmc53470.2022.9753773.

[34] M. A. Hasan, Y. Riyanto, and D. Riana, "Grape leaf image disease classification using CNN-VGG16 model," *Jurnal Teknologi dan Sistem Komputer*, vol. 9, no. 4, pp. 218–223, Oct. 2021, doi: 10.14710/jtsiskom.2021.14013.

[35] X. Xie, Y. Ma, B. Liu, J. He, S. Li, and H. Wang, "A deep-learning-based real-time detector for grape leaf diseases using improved convolutional neural networks," *Frontiers Plant Sci.*, vol. 11, Jun. 2020, doi: 10.3389/fpls.2020.00751.

[36] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," 2016, *arXiv:1602.07261*.

[37] M. Canayaz, "C+EffxNet: A novel hybrid approach for COVID-19 diagnosis on CT images based on CBAM and EfficientNet," *Chaos, Solitons Fractals*, vol. 151, Oct. 2021, Art. no. 111310, doi: 10.1016/j.chaos.2021.111310.

[38] Z. Huang, Y. Ben, G. Luo, P. Cheng, G. Yu, and B. Fu, "Shuffle transformer: Rethinking spatial shuffle for vision transformer," 2021, *arXiv:2106.03650*.

[39] S. Cheng, L. Wang, and A. Du, "Asymmetric coordinate attention spectral–spatial feature fusion network for hyperspectral image classification," *Sci. Rep.*, vol. 11, no. 1, Aug. 2021, doi: 10.1038/s41598-021-97029-5.

[40] Z. Tang, X. He, G. Zhou, A. Chen, Y. Wang, L. Li, and Y. Hu, "A precise image-based tomato leaf disease detection approach using PLPNet," *Plant Phenomics*, vol. 5, Jan. 2023, doi: 10.34133/plantphenomics.0042.

[41] J. Li, G. Zhou, A. Chen, C. Lu, and L. Li, "BCMNet: Cross-layer extraction structure and multiscale downsampling network with bidirectional transpose FPN for fast detection of wildfire smoke," *IEEE Syst. J.*, vol. 17, no. 1, pp. 1235–1246, Mar. 2023, doi: 10.1109/jsyst.2022.3193951.

[42] J. Zhan, Y. Hu, G. Zhou, Y. Wang, W. Cai, and L. Li, "A high-precision forest fire smoke detection approach based on ARGNet," *Comput. Electron. Agricult.*, vol. 196, May 2022, Art. no. 106874, doi: 10.1016/j.compag.2022.106874.

[43] L. Zhang, C. Lu, H. Xu, A. Chen, L. Li, and G. Zhou, "MMFNet: Forest fire smoke detection using multiscale convergence coordinated pyramid network with mixed attention and fast-robust NMS," *IEEE Internet Things J.*, vol. 10, no. 20, pp. 18168–18180, Oct. 2023, doi: 10.1109/jiot.2023.3277511.

[44] A. S. Zamani, L. Anand, K. P. Rane, P. Prabhu, A. M. Buttar, H. Pallathadka, A. Raghuvanshi, and B. N. Dugbakie, "Performance of machine learning and image processing in plant leaf disease detection," *J. Food Quality*, vol. 2022, pp. 1–7, Apr. 2022, doi: 10.1155/2022/1598796.

[45] Z. Guo, H. Zheng, X. Xu, J. Ju, Z. Zheng, C. You, and Y. Gu, "Quality grading of jujubes using composite convolutional neural networks in combination with RGB color space segmentation and deep convolutional generative adversarial networks," *J. Food Process Eng.*, vol. 44, no. 2, Dec. 2020, doi: 10.1111/jfpe.13620.

[46] V. Meshram, K. Patil, V. Meshram, D. Hanchate, and S. D. Ramkteke, "Machine learning in agriculture domain: A state-of-art survey," *Artif. Intell. Life Sci.*, vol. 1, Dec. 2021, Art. no. 100010, doi: 10.1016/j.ailsci.2021.100010.

R. KARTHIK received the master's degree from Anna University, India, and the Ph.D. degree from the Vellore Institute of Technology, India. He is currently an Associate Professor with the Research Centre for Cyber Physical Systems, Vellore Institute of Technology, Chennai. 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 for journals published by Elsevier, IEEE Springer, and Nature.

R. MENAKA received the master's degree in applied electronics and the Ph.D. degree from Anna University, Chennai, India. She is currently a Professor with the Center for Cyber Physical Systems, Vellore Institute of Technology, Chennai. She has published around 80 papers in peer-reviewed journals and conferences. Her research interests include biomedical image and signal processing, neural networks, and fuzzy logic.

S. OMPIRAKASH received the bachelor's degree in electronics and communication engineering from the Vellore Institute of Technology, Chennai. His research interests include deep learning, machine learning in the agricultural field, and chatbot for customer service. Recently, he has published research work on chatbots.

P. BALA MURUGAN received the Bachelor of Technology degree in electronics and communication engineering from VIT, Chennai. He is currently pursuing the master's degree in data science and artificial intelligence with the University of Liverpool. His research interests include deep learning, image recognition, data science, and machine learning.

M. MEENAKASHI is currently an Electronics and Communication Engineer, passionate about bringing innovative solutions for the real-world problem in a more efficient way. Her current research interest includes bringing an optimizing solution using deep learning technologies for agriculture-based problems.

SINDHIA LINGASWAMY is currently an Assistant professor with the Department of Computer Applications, National Institute of Technology Tiruchirappalli, Tiruchirappalli. She has published around ten papers in peer-reviewed journals and conferences. Her research interests include computer vision, image processing, artificial intelligence, and machine learning.

DAEHAN WON is currently an Associate Professor with the Department of Systems Science and Industrial Engineering, Binghamton University, Binghamton, NY, USA. His research interests include large scale mathematical programming, data analytics/mining in healthcare, and designing smart manufacturing systems to advance industry 4.0.