An Efficient Depthwise Multiscale Feature Learning Convolutional Network for Plant Leaf Disease Classification in Agriculture

Arun Mittal
ankitamttl064@gmail.com

Teerthanker Mahaveer University
Ashok Kumar
Teerthanker Mahaveer University
Shelendra Pal
Jaypee Institute of Information Technology

Research Article

Keywords: Transfer Learning Model, Computer vision, Diseases detection, Depth-wise convolution, Plant Disease

Posted Date: February 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8513197/v1

Additional Declarations: No competing interests reported.

An Efficient Depthwise Multiscale Feature Learning Convolutional Network for Plant Leaf Disease Classification in Agriculture

[1]Arun Mittal*, [2]Ashok Kumar and [3]Shelendra Pal
[1]ankitamttl064@gmail.com, [2]*ashchhonkar@gmail.com* and [3]*shelendra.pal@jiit.ac.in*
[1,2]*College of Computing Sciences & IT,*
Teerthanker Mahaveer University. Moradabad, India
[3]*Department of Computer Applications, Jaypee Institute of Information Technology*
Noida, India

Abstract

Plant disease detection and early disease treatment are essential for sustainable crop production. Computer vision for crop science is growing with the advancement in deep learning. The proposed work systematically addresses these issues through three datasets as Plant Village Maize Dataset (D1), Paddy Doctor Dataset (D2), and Sugarcane Leaf Image Dataset (D3) with different classes. The dataset contains 4188, 16225, and 6748 images from dataset sets D1, D2, and D3, respectively. This work has used a Generative Adversarial Network (GAN) to generate a synthetic dataset. Further use data preprocessing, and the data has been resized to $224 \times 224 \times 3$. The proposed model use Depth-wise Multiscale Feature Learning ConvoNet (DMFL-ConvoNet) model, which includes the Depth-Wise Convolutional Block (DCB) block of DMFL-ConvoNet with 3×3 and 5×5, facilitates the extraction of multiscale plant disease characteristics. Furthermore, it has added 2.5 million parameters. The proposed DMFL-ConvoNet model offers state-of-the-art performance and decreases computational complexity at 33 frames per second, making it ideal for real-time applications. The proposed DMFL-ConvoNet model has been compared with several transfer learning models, including ResNet50V2, InceptionResNetV2, NASNetMobile, EfficientNetV2L, and EfficientNetV2B0 models, and the proposed model has achieved 99.52% data accuracy in the multiple datasets.

Keywords

Transfer Learning Model, Computer vision, Diseases detection, Depth-wise convolution, Plant Disease

1. Introduction

A key component of the agricultural system that powers the world economy is crop production. It needs to be managed and planned properly. To increase crop productivity, it is crucial to develop new trends and scientific techniques. Techniques based on artificial intelligence are crucial to agricultural production [1]. Image processing is an underlying technology used by artificial intelligence approaches to improve yield output. Crop pictures are used in image processing to find and detect a variety of crop problems. Crop-related problems include managing the soil, weeds, and stress. Several types of leaf stress exist that are detrimental to crop productivity and should be identified early in the growth cycle [2]. The development of leaf stress in the crop will have an impact on its basic needs, which include water, soil, minerals, fresh air, sunlight, etc. A variety of methods have been developed for classifying and identifying plant stress. In computer vision, image processing, and precision farming, the difficult study topic of plant lesion recognition from images can lead to accurate, quick, and effective lesion detection. Using image processing techniques, the automatic identification of plant lesions provides quite acceptable and straightforward solutions [3,4]. These lesions are caused by various stresses in various plant parts. A plant can be affected by a wide range of stressors, which makes it challenging to identify it using visualization techniques. Different forms of stress, both biotic and abiotic, such as rain, pollution, viruses, bacteria, fungi, and insects, can influence a plant's leaves. According to the Food and Agriculture Organization (FAO) of the United Nations, the number of people suffering from hunger worldwide has been steadily rising since 2015 [1]. Current estimates indicate that nearly 680 million individuals, accounting for less than 9% of the global population, are affected by hunger. This growing crisis highlights the urgent need for efficient and sustainable agricultural practices. Moreover, plants play a vital role in maintaining ecological balance by producing oxygen through photosynthesis. However, when plant leaves are infected by disease, it can lead to the deterioration or death of the plant, ultimately hindering food production. Plant diseases have a profound impact on the growth of food crops and pose a serious threat to global agriculture. They significantly reduce both crop yield and quality, thereby endangering food security and the long-term sustainability of the agricultural sector [2]. Farmers traditionally rely on visual inspection and personal experience to identify diseases on plant leaves, a process that is time-consuming, labor-intensive, and demands specific expertise. However, with the advancement of information technology, computer vision-based methods for detecting plant diseases have attracted considerable interest. By analyzing images of leaves, these techniques can accurately and swiftly diagnose various diseases, greatly enhancing both the speed and accuracy of detection.

Although advancements in science and technology have promoted agricultural development, crop yields have not shown significant improvement due to several natural and human-induced factors. Among these challenges, crop diseases and insect pests account for the largest share of yield loss. Currently, research on crop diseases can be broadly classified into two main approaches [3]. The first is the traditional physical method, which primarily relies on spectral detection to differentiate between various diseases. Since pests and pathogens affect leaves in different ways, the resulting damage alters their spectral absorption and response, allowing a distinction between healthy and infected crops. The second approach employs computer vision techniques for image-based identification. Images captured in real-world scenarios contain a lot of noise and varying levels of illumination. Therefore, if the classifier provides the actual image directly, disease classification suffers. To solve these issues, several preprocessing procedures are used to improve the samples' visual representation. One such technique is image segmentation. The segmentation process, which aimed to eliminate the diseased section of the plant leaf, allowed for the accurate identification of the Region of Interest (ROI). It then uses it as input for recognition. In computer vision, segmentation is the process of splitting an existing image into segments or sections to extract information and identify objects or diseases. By simplifying the input picture data and the space, the segmentation technique also facilitates the recall of aberrant regions. This process includes obtaining the required ROI and the image's bounds. Data can be segmented using a variety of methods, including Otsu's threshold algorithm, K-means clustering, and border detection algorithms.

In this method, disease-related features are extracted from plant images using computational techniques, and classification is achieved by analyzing the differences between healthy and diseased plants. DL has come up as a transformative approach for crop disorder identification with the help image-based categorization systems. Padhi J. et al. [4] put forth a hybrid 'Convolutional Neural Network (CNN)' model that combines simulated thermal imaging to increase the diagnostic accuracy of paddy leaf disorder under changing environmental conditions. This approach raised firmness and decreased false categorizations in 'real-field imagery'. Subbarayudu C. et al. [5] put forth an 'Automated Hybrid DL' architecture that composites multiple CNN architectures, giving higher categorization accuracy and efficiency in identifying complex paddy leaf disorders. In the context of maize leaf disorders, Paul H. et al. [6] displayed a mobile use case using a pre-trained 'VGG16 CNN' architecture that balances categorization accuracy with deployment efficiency, especially in low-resource environments. Meng Y. et al. [7] introduced a rapid maize disorder identification technique related to 'YOLO-MSM', with the help of multi-scale modules and attention layers to confirm faster convergence and raised object localization. Shandilya G. et al. [8] explained a combined 'CNN-ViT (Vision Transformer)' model that uses 'Transformer Encoders' for pulling off contextual relations in maize leaf disorder patterns, displaying increased performance over classical CNNs. Sun L. et al. [9] uncovered the effectiveness of 'CASF-Mnet', a cross-attention and multi-scale fusion network, which toughened spatial representation and pulled off higher performance in early-stage disorder detection. Additional innovations have raised disorder localization and ‘Fine-grained’ categorization in maize. Rai C. K. et al. [10] rated deep CNN models to detect and separate Northern maize leaf blight, pulling off precise region-level disorder segmentation for good agronomic insights. Akkamahadevi C. et al. [11] rated a DL model focused at raising paddy production by using advanced pattern extraction and ensemble decision-making, resulting in sizable gains in precision and recall for disorder categorization. Bachhal P. et al. [12] put forth a hybrid categorization idea combining ‘Probabilistic Random Forest (PRF)’ and ‘Support Vector Machine (SVM)’, leading to tough generalization across different maize disorder data records. Khan I. et al. [13] explained the use of ‘Deep Transfer Learning’ for ‘Fine-grained’ maize leaf disorder categorization, which notably increased the identification of visually similar disorder types while decreasing training time.

Sugarcane disorder detection has also profited from these developments. Srinivasan S. et al. [14] put forth a ‘Deep Neural Network (DNN)-based’ categorization structure for sugarcane leaf disorders that pulled off high accuracy with less preprocessing. Rajput K. et al. [15] rated a hybrid ‘CNN-SVM’ system composited for sugarcane pathology, enabling more interpretable and exact categorization. Qaadan S. et al. [16] displayed a stacked ensemble learning idea that avoids class imbalance and raises F1 scores, making it suitable for real-world smart farming use cases. Jardawala H. A. et al. [17] put forth a CNN model based on the ‘Xception’ architecture, outlining efficient depth wise separable convolutions for quicker and more accurate disorder identification in precision agriculture. Finally, Ruprah T. S. et al. [18] explained an all-inclusive pipeline that combines DL methods to detect and categorize sugarcane leaf disorders, showing potential for large scale field deployment.

The major contribution of our work is as follows:

- The disease image is classified using sixty different categories from three different datasets: the Sugarcane Leaf Image Dataset (D3), the Paddy Doctor Dataset (D2), and the Plant Village Maize Leaf Dataset (D1). A 99.52% data accuracy was obtained when the suggested approach was applied to the dataset.
- It is recommended to use multiscale feature learning as DMFL-ConvoNet. The ConvoNet (DMFL-ConvoNet) model includes the DCB block with 3 × 3 and 5 × 5, which facilitates the extraction of multiscale plant disease characteristics.
- With just 2.5 million parameters, the proposed DMFL-ConvoNet model offers state-of-the-art performance and decreases computational complexity at 33 frames per second, making it ideal for real-time applications. It has been integrated with several transfer learning models, including ResNet50V2, InceptionResNetV2, NASNetMobile, EfficientNetV2L, and EfficientNetV2B0.

The rest of this paper is organized as follows. Section 2 discusses about the Plant Disease Dataset Collection. Section 3 explain about the methodology, whereas Section 4 explains the data result using different models in terms of accuracy, precision, recall, and F1-score. Finally, Section 5 concludes the research results and proposes some future work.

2. Plant Disease Dataset Collection

The current work created a real-time plant disease dataset for the three major food grains—rice, wheat, and maize—based on leaf photographs. Diseases with comparable symptoms, a high yield loss, a large number of accessible samples, and other attributes were chosen to generate the dataset. For the diseases considered under Plant Village Maize Dataset (D_1), Paddy Doctor Dataset (D_2), and Sugarcane Leaf Image Dataset (D_3). Furthermore, Figure 1 shows the complete workflow of this work, which is followed in this work. The descriptions of the dataset are given in Section 2.1.

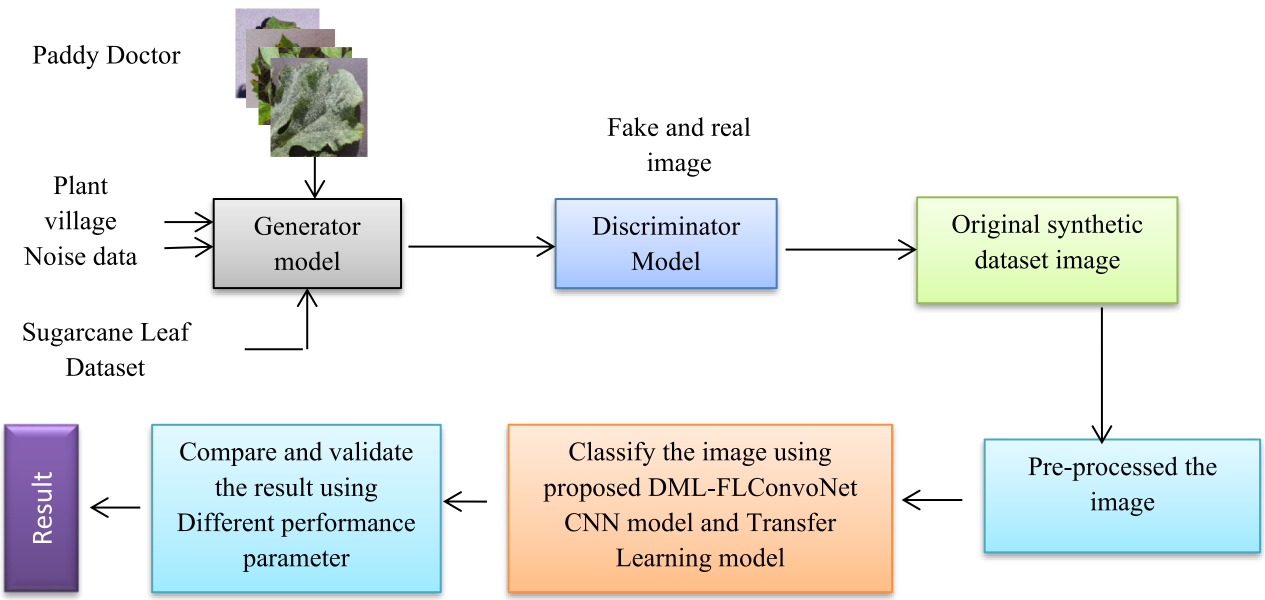


Figure 1. Complete dataflow of this work

2.1 Description of Dataset

This dataset, which focuses on the detection and classification of illnesses in chili plants, is an extensive resource for scholars and experts in the domains of computer vision, machine learning, and agriculture. It is ideal for deep learning applications since it offers high-resolution pictures of both healthy and sick leaves.

A. Dataset D_1: Plant Village Maize Dataset

The maize leaf dataset comprises a total of 4,188 labeled images categorized into four classes, representing both diseased and healthy conditions of maize plants. It includes 1,306 images of Common Rust, characterized by reddish-brown pustules on the leaf surface, 574 images of Gray Leaf Spot, showing elongated rectangular gray to tan lesions, 1,146 images of Blight, marked by irregular necrotic patches causing tissue damage, and 1,162 images of Healthy leaves, free from any disease symptoms [19]. This dataset provides a diverse representation of maize leaf health conditions and can be effectively utilized for developing and evaluating deep learning models aimed at automatic disease detection and classification in maize crops.

B. Dataset D_2: Paddy Doctor Dataset

16,225 captioned paddy leaf photos from 13 different classes—including 12 distinct paddy diseases and healthy leaves—are included in the Paddy Doctor collection. It is the greatest collection of expertly annotated photos for computer vision algorithm testing and evaluation. A smartphone camera with a high quality (1,080 x 1,440 pixels) was used to capture the photographs of rice leaves from actual paddy fields. An agronomist assisted with the meticulous cleaning and labeling of the gathered photos. Additional information is available on the Paddy Doctor Project's website [20], which may be accessed at https://paddydoc.github.io. A selection of the photos in the dataset is displayed in Figure 2.

Dataset D1: Plant village

Dataset D_2: Paddy Doctor **Dataset**

Dataset D_3: Sugarcane Leaf Dataset

Figure 2. The sample image is a leaf disease dataset

C. Dataset D_3: Sugarcane Leaf Dataset

6,748 high-resolution photos of sugarcane leaves make up the varied Sugarcane Leaf Dataset. The pictures have a resolution of 768 × 1024 pixels and are saved in the JPEG format [21]. Nine illness categories, a healthy leaf category, and a dry leaf category are among the dataset's eleven classifications. The disease categories encompass common sugarcane leaf diseases such as banded chlorosis, brown spot, brown rust, pokkah boeng, mosale, grassy shoot, smut, yellow leaf disease, and sett rot (9 class 6,748 high-resolution photos of sugarcane leaves make up the varied Sugarcane Leaf Dataset. The pictures have a resolution of 768 × 1024 pixels and are saved in the JPEG format [21]. Nine illness categories, a healthy leaf category, and a dry leaf category are among the dataset's eleven classifications. With a 72-dpi resolution, the images in the dataset are of exceptional quality, ensuring that the sugarcane leaf samples are displayed in a clear and detailed manner. Figure 2 shows a sample of the photos, and Table 1 has the explanation.

Table 1. Dataset description

S.N.	Data Set Sources	Plant Category	Leaf Category (Healthy / Diseased)		Total No. of Plant Images
			Healthy	Diseased	
1	Paddy Doctor Dataset	Paddy	2405	13820	16,225
2	Sugarcane Leaf Dataset	Sugarcane	597	6151	6748
3	Plant Village Maize Dataset	Maize	1162	3026	4188

The table outlines three distinct plant leaf datasets—Paddy, Sugarcane, and Maize—showing the number of healthy and diseased samples in each. The Paddy Doctor Dataset includes a total of 16,225 images, with 2,405 labeled as healthy and 13,820 as diseased, offering a comprehensive source for rice disease analysis. The Sugarcane Leaf Dataset provides 6,748 images, comprising 597 healthy and 6,151 diseased leaves, with a strong bias toward diseased cases. The Plant Village Maize Dataset contains 4,188 images, of which 1,162 are healthy and 3,026 are diseased, covering major maize leaf diseases such as rust, blight, and gray leaf spot. Collectively, these datasets are dominated by diseased samples, making them highly useful for training and testing deep learning models aimed at crop disease identification and classification.

3. Methodology

This work has used three different datasets, and the images are in different sizes. So, images need to be preprocessed before executing the transformer learning model.

3.1 Leaf Image Generator

In order to expand the dataset and lessen over-fitting, the "Generative Adversarial Network (GAN)" may be employed as a data augmentation technique in this study. In GAN, noise is converted into a matrix using conventional convolutional layers. The discriminator model and the generator model are the two separate models that make up a GAN [22]. While the discriminator is responsible for identifying fake images, the generator is in charge of creating them. The generator and discriminator engage in competition and practice at the same time. The discriminator's job is to make sure that the images produced by the generator are as accurate as possible. The input, embedding, dense, input, reshaping, and concatenate layers comprise the GAN discriminator model. After that, four convolutional layers are used. A Leaky ReLU layer takes the place of the ReLU layer after each convolutional layer. A thick layer, a dropout layer, and a flat layer come after the final Leaky ReLU layer. There are 771,454 trainable parameters in the model overall [23]. Table 2 displays the discriminator model parameters and network layers.

Table 2: GAN Discriminator model description

Layer Name	Discriminator model No. of Parameters	**Generator** model No. of Parameters
Input	[224x224x3]	[4,4,3]
Dense	sigmoid activation	sigmoid activation
Leaky ReLU	max(0,x)	max(0,x)
Conv2D	[32x32], [16x16], [8x8], [4x4], stride=[2x2], padding same	[32x32], [16x16], [8x8], [4x4], stride=[2x2], padding same
Flatten	[0,2048]	

Dropout	[0,2048]

An embedding layer, a Leaky ReLU layer, a dense layer, a reshape layer, a concatenation layer, and input layers make up the GAN generator model. A last convolutional layer comes after the last Leaky ReLU layer [24]. A total of 1,735,904 parameters is available for training the model. The GAN generator model (G_e), uses random noise and latent points as inputs to produce fake pictures. Distinguishing between the authentic and fraudulent pictures produced by the (G_e) model is the aim of the discriminator model (E_i). Assume that the noise distribution in the (G_e) model is N_d and that the data and supplementary class label. The generator model in Equation 1 generates both synthetic and real pictures, and it maximizes the probability that both types of images will be properly identified by using the E_i.

$$\mathrm{m}_{in}\ \_m_{ax} P(E_i, G_e) = A_{data}[\log E_i(I|L)] + [\log(1 - E_i G_e(Z|L)] \quad \text{(eq. 1)}$$

The suggested method generates synthetic pictures of various leaf diseases using a GAN. Before utilizing the GAN model to produce synthetic pictures, use the photographs to train it. The discriminator model in a GAN is trained using a real picture of a plant leaf and a label, whereas the generator model is trained with noise and a label. A fictitious picture of a leaf disease is now produced by the generator model and sent into the discriminator model.

3.2 Data Preprocessing

Pre-processing is necessary to enhance the quality of the input photographs before they are used by deep learning or machine learning algorithms. The significant percentage of damaged and pictures in the Plant Village maize dataset, Paddy Doctor Dataset, and Sugarcane Leaf Dataset collection necessitates cautious processing. Every leaf picture is scaled and normalized to a common dimension to ensure consistency and compatibility with deep learning models. To standardize input sizes and lessen the computing strain, image scaling, often referred to as resizing, is necessary [25]. Then, normalization and augmentation procedures are used to further improve the photographs' visual quality. These pre-processed photos are then run through a density predictor and impulse noise detector to detect and fix noisy data, ensuring that the input is cleaner for a precise illness diagnosis. Collecting the crop leaf images from the Plant Village collection is the first step. In order to preserve the information, scaling, normalization, and augmentation techniques are required because the Plant Village collection includes some damaged and indistinct images [26]. All leaf images were scaled to 224 × 224 × 3 pixels to guarantee uniformity and model conformance. Image scaling is required to standardize the input dimensions and lower the processing cost of the framework. The pre-processed pictures are then transmitted to the noise detector and impulse density predictor to preserve image quality. Further, in section 3.3 has discussed the proposed model is discussed.

3.3 Proposed Model: Depth-wise Multiscale Feature Learning ConvoNet (DMFL-ConvoNet)

The goal of the suggested approach is to create a low-weight network for plant disease detection. The main goal is to develop a reliable CNN architecture that can be used for real-time applications and run on low-resource hardware.

An effective DMFL-ConvoNet architecture has the following characteristics:

(a) fewer parameters;

(b) lower memory costs (model size)

(c) decreased computational complexity (total multiplication and addition operations)

(d) enhanced performance.

3.3.1 DMFL-ConvoNet in deep learning

A key component of deep learning models, especially for image and video processing applications, is regular convolution. However, its computing requirements may limit its real-time application. Equation (1), seen in Figure 3, provides the parameters and formulae for determining the computing cost of simple convolution. In order to overcome this limitation, depth-wise convolution was developed, which reduces the number of parameters and computing load by processing input channels individually using different kernels. DMFL-ConvoNet is a well-liked technique for more effectively creating deeper neural networks by combining depth-wise and binary-wise (1x1) convolutions [27]. Figure 3 displays the variables and computing cost for DMFL as determined by Equation (2).

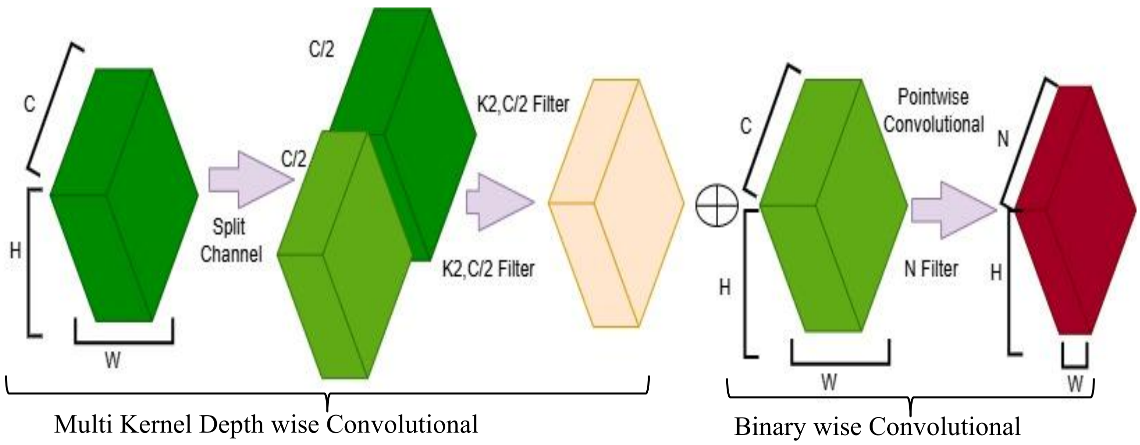


Figure 3. Depth-wise Convolutional Layer

This block reduces computation by splitting channels and applying group convolutions before using a 1×1 convolution for channel mixing—common in mobile-efficient CNNs like Shuffle Net. Although it works well, the depth-wise separable convolution has limitations when it comes to extracting multi-scale features. To get around this restriction, we suggested the DMFL-ConvoNet [28]. It retrieves multistate characteristics of plant diseases by using several kernels in a depth-wise convolution process. Thus, learning spatial correction at various scales is aided by multi-scale depth-wise convolution. Equation (eq 2) may be used to determine the input channels' group size.

$$H_t = \frac{D_m}{L_j} \qquad \text{(eq 2)}$$

where H_t = Group size, D_m Is the number of input channels, and L_j Is the number of distinct kernels. DMFL-ConvoNet is comparable to conventional depth-wise convolution when $H_t = 1$. For every channel group, a variety of kernels employing depth-wise convolution methods are employed. The input channels are separated into two groups, with a group size of $H_t = 2$ as the proposed MDsConv contains both 3×3 and 5×5 kernels. The number of channels in the input feature map is half that of the 3x3 and 5x5 kernels. Since $[\,C_1\,, C_2\,]$ The input feature map is split into two groups. The C_1 is a feature map, which represents a 3 x 3 DMFL-ConvoNet, and the second C_2 feature map, which represents 5×5 DMFL-ConvoNet. Figure 4 displays every element of the DMFL-ConvoNet. Algorithm 1 presents the whole pseudo-code for DMFL-ConvoNet [29]. The following formula yields the whole computation and parameters required for DMFL-ConvoNet:

$$\text{DMFL} - \text{ConvoNet} = \left\{\left(3 \times 3 \times \frac{C}{2} \times H \times W\right) + \left(5 \times 5 \times \frac{C}{2} \times H \times W\right)\right\} + (C \times N \times H \times W) \quad \text{(eq. 3)}$$

Table 3 shows the abbreviations which has been used in Algorithm 1 as DMFL-ConvoNet.

Table 3. Abbreviations which has been used in Algorithm 1 as DMFL-ConvoNet

Abbreviation	Description
$A(P_d, S_l, P_l)$	Input image
C, W, H	Channel, Width, and Height
G_{conv}	Convolution Operations of Convnext Backbone
H_t	Group size
D_m	Number of input channels
L_j	Number of unique kernels
P_d	Paddy Doctor Dataset
S_l	Sugarcane Leaf Dataset
P_l	Plant Village Maize Dataset
T_e, L_1,L_2	Tensor, Kernal Size1,and Kernal Size2
Sp	Split

Algorithm 1. Depth-wise convolutional block of DMFL-ConvoNet (DCB)

Input: $A(P_d, S_l, P_l)$, T_e, L_1,L_2,

Output: B_{out}

1. Take input as $A(P_d, S_l, P_l)$ in Tensor(T_e) and L_1,L_2
2. F= $A(P_d, S_l, P_l)$ with L_1,L_2,
3. Split the $A(P_d, S_l, P_l)$ =Sp(A, F, A_x= -1)
4. Set P, array as out[ ], P=out[]
   for A, $K_i in$ zip(S_p, L_1) do
5. $MSD_{convo} = D_e(L, L_i)A$
6. P. concanate(out),
7. P=out.BN2D(out)
8. P=out.BN2D(out)
9. P=out.Convo((out. , L_1,L_2)=1)
10. $B_{out} = BatchNorm\ 2D(out)$

Take input as $A(P_d, S_l, P_l)$ in Tensor(T_e) and L_1,L_2 and take F= $A(P_d, S_l, P_l)$ with L_1,L_2, and split the $A(P_d, S_l, P_l)$ =Sp(A, F, A_x= -1). Furthrmore, set P, array as out[ ], P=out[] for A, $K_i in$ zip(S_p, L_1) do MSD_{convo}. Finally, concentrate (out), out.BN2D(out), P=out.BN2D(out) P=out.Convo((out. , L_1,L_2)=1) and return $BatchNorm\ 2D(out)$. The final includes DMFL-ConvoNet blocks, are shown in Figure 4.

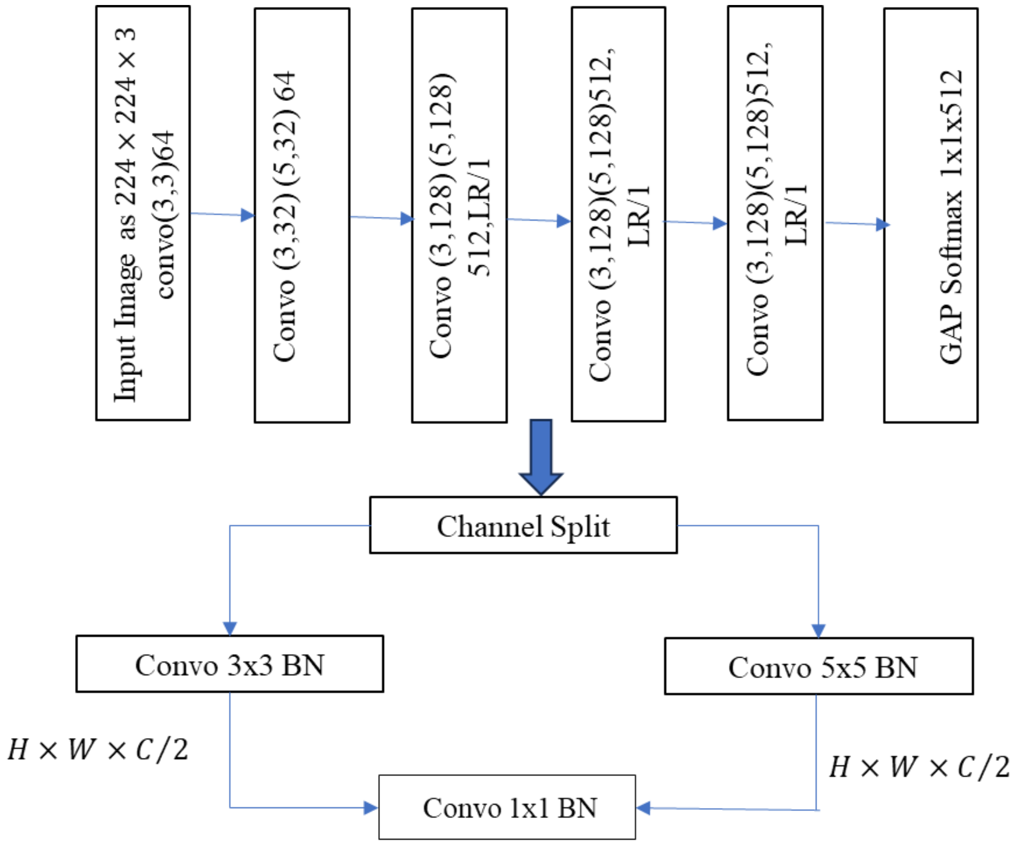


Figure 4. The proposed DMFL-ConvoNet blocks architecture includes DCB blocks.

The input dataset is enlarged to 224 × 224 × 3 before being sent for feature extraction. Except for the final layer, every batch normalization layer in the proposed network uses the LeakyReLU activation function. Using a stride of two rather than maximum pooling reduces the input feature's spatial dimensionality. Additionally, 1 × 1 pixel point-wise convolution employing 512 filters is used in place of the fully linked layers in order to minimize the number of parameters without compromising accuracy [30]. After the global average pooling layers, the softmax layer generates the class labels.

4. Result and Discussion

Each model's learning rate was set at 0.001, and 32 was selected as the batch size. To improve the models' generalization and avoid over-fitting, a dropout operation was applied to each of the pre-trained models on the DMFL-ConvoNet. The DMFL-ConvoNet model is enhanced by the DCB blocks when batch sizes are 32 and the learning rate is 0.001.

4.1 Classification Report

The results of all the approaches used in this study were combined into an exceedingly matrix notation with several measurements, including "accuracy", "precision", "recall", and "F1-score". Ascertain the neural community's accuracy through the utilization of affirmative test detection. Equation 4 defines precision as the extent to which a neural network can correctly identify a target that possesses an accumulation of positive values [13].

$$Pre = \frac{Tr_{Po}}{Tr_{Po}+Fa_{Po}} \quad \text{Eq. 4}$$

Diseased particles are those that an algorithm fails to detect. False positives Fa_{Ne} It occurs when pixels with the disease are incorrectly identified as healthy or non-diseased pixels. True negatives Tr_{Po} Predict the location of leaf diseases with precision [14]. The recall value [15] is computed as the sum of the cumulative values in equation 5 and the accuracy with which the neural network detects a goal.

$$Rec = \frac{Tr_{Po}}{Tr_{Po}+Fa_{Ne}} \quad \text{Eq. 5}$$

Equation 6 illustrates that F1 is the weighted average of recall and accuracy; it signifies the proportion of accurate positive predictions. Each of these measures evaluates how well the segmentation and classification phases performed [16].

$$F1 - Score = \frac{2 \times Pre \times Rec}{Pre+Rec} \quad \text{Eq. 6}$$

Accuracy of a proposed system is defined as the ratio of accurately predicted samples to the total number of samples. The prediction's accuracy is calculated by following equation 7:

$$Accuracy = \frac{\sum Tr_{Po}+\sum Fa_{Ne}}{Tr_{Po}+Fa_{Ne}+Tr_{Po}+Fa_{Po}} \quad \text{(Eq. 7)}$$

For segmentation evaluation, these metrics are computed at the pixel level; for classification evaluation, they are computed at the image level. To perform simple segmentation, the average intersection is employed rather than the mean intersection over union (mIoU) [17]. For evaluating segmentation, the mioU index is utilized. The proximity between a Ground-Truth named mask (GT) and its corresponding segmented mask (SM) or forecast mask (FM) is quantified by the mIoU index [18]. With 99.55% data accuracy, this effort has produced improved results. The suggested model weighs little, has a 9.6 M parameter, and takes 6 hours to run. The computationally effective optimization approach employed in the proposed model employs a tailored optimization and adjusts to the shifting loss landscape.

4.2 Comparison with the standard CNN model

We examine DMFL-ConvoNet's parameter requirements and computing cost for a single convolution operation. The addition and multiplication operations are added together to determine the cost. We will examine a 224 × 224 × 3 input feature map with input channels (C) = 16, padding = 1, stride = 1, and number of filters (N) = 64. Conventional convolution with kernel size K = 3 × 3 yields the following results, whereas multi-kernel depth-wise convolution kernel size is L1 = 3 × 3. Two kernels, L2 = 5 × 5, have the same parameters and parameter, 1575, but the depth-wise convolution kernel size has a parameter of 9280 and costs 921.6 M. Compared to the standard convolution details in figure 5, DMFL-ConvoNet has 55 times reduced processing cost and seven times fewer parameters.

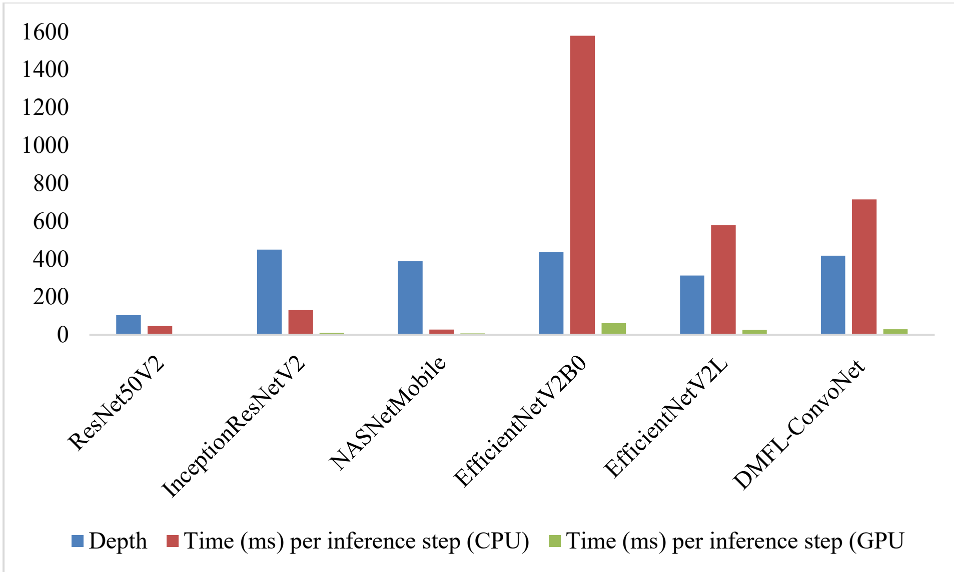


Figure 5: DMFL-ConvoNet model weighted description

The horizontal bar chart compares six CNN models—ResNet50V2, InceptionResNetV2, NASNetMobile, EfficientNetV2B0, EfficientNetV2L, and DMFL-ConvoNet—based on model depth, CPU inference time, and GPU inference time. Depth (blue bars) varies from low in NASNetMobile to high in EfficientNetV2L, while GPU inference times (green bars) are consistently low across all models. CPU inference times (red bars) are significantly higher, with EfficientNetV2B0 showing the longest time (~1600 ms) despite moderate depth, whereas NASNetMobile has the lowest CPU inference time. Overall, DMFL-ConvoNet offers a balanced profile with relatively high depth and moderate CPU time, while GPU times remain efficient for all models. Less time is spent on inference since faster answers are obtained with fewer calculations and parameters. The suggested architecture maintains low memory and computational complexity while achieving state-of-the-art performance in plant disease classification.

4.3 Comparison of the Model on Public Datasets

Three publicly accessible datasets—the Plant Village Maize Leaf Dataset, the Paddy Doctor Dataset, and the Sugarcane Leaves Dataset—are used to assess the suggested approach. As indicated in Table 10, the comparison is undertaken in terms of accuracy and total parameters. In every dataset, the proposed architecture performs better than the existing work. Since all of the photos in the plant village dataset were taken inside with a constant background and without any of the multiscale variation examples seen in Figure 6, all of the techniques perform well, reaching over 99%.

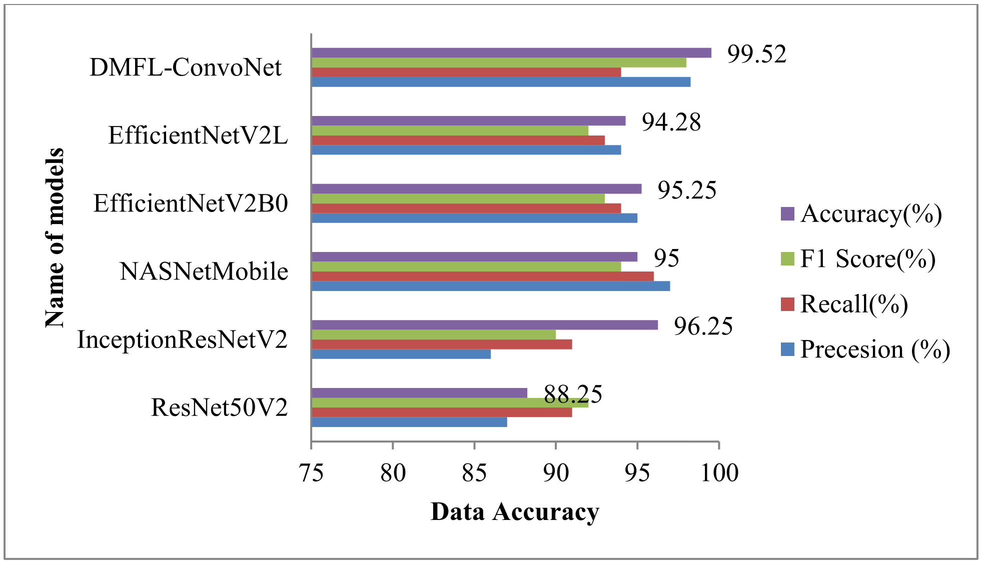
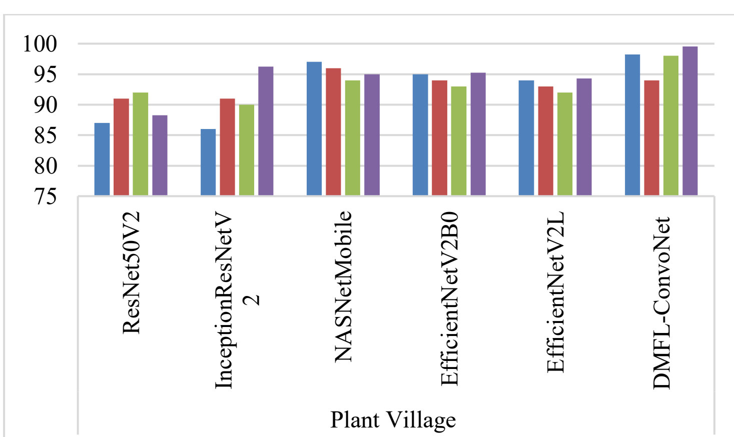
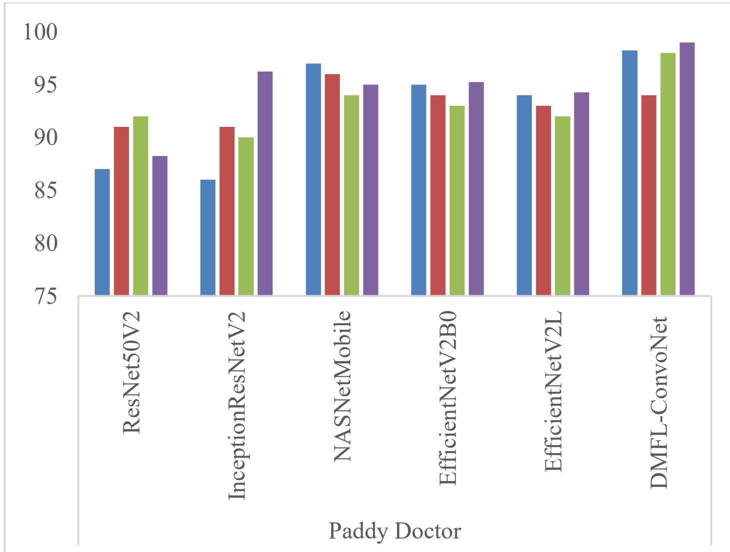


Figure 6. Comparison of the proposed model with different transfer learning models

The bar chart compares six CNN models—DMFL-ConvoNet, EfficientNetV2L, EfficientNetV2B0, NASNetMobile, InceptionResNetV2, and ResNet50V2 are using four performance metrics: Accuracy, F1 Score, Recall, and Precision. DMFL-ConvoNet outperforms all others, achieving the highest accuracy of 99.52%, followed closely by EfficientNetV2L and EfficientNetV2B0, which also show consistently high results. NASNetMobile maintains good performance but is slightly behind in recall, while InceptionResNetV2 and ResNet50V2 have comparatively lower recall and F1 scores despite decent precision, indicating a trade-off between sensitivity and overall accuracy.

The graph compares the performance of six deep learning models: DMFL-ConvoNet, EfficientNetV2L, EfficientNetV2B0, NASNetMobile, InceptionResNetV2, and ResNet50V2 across three datasets: Sugarcane, Paddy Doctor, and Plant Village Maize, using Accuracy, F1 Score, Recall, and Precision as evaluation metrics. Across all datasets, DMFL-ConvoNet consistently achieves the highest performance, with values close to or above 95% in all metrics, demonstrating strong robustness and generalization. EfficientNetV2L and ResNet50V2 also show competitive results, particularly in accuracy and precision, while NASNetMobile and InceptionResNetV2 perform slightly lower, especially in recall. Overall, all models maintain high performance above 85% in most cases, with accuracy and precision generally being the strongest metrics across the board. Figure 7 shows the Comparison of the transfer learning model concerning the D1, D2, and D3 datasets.





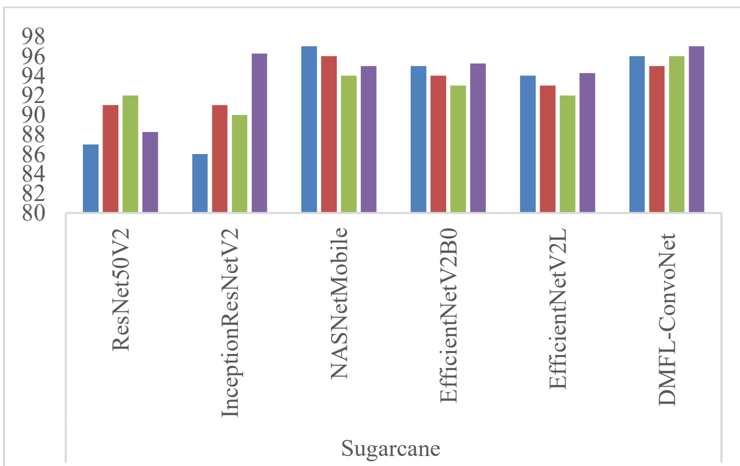


Figure 7. Comparison of the transfer learning model concerning D1, D2, and D3 datasets.

The given figure 7 shows comparison of the performance of six deep learning models: DMFL-ConvoNet, EfficientNetV2L, EfficientNetV2B0, NASNetMobile, InceptionResNetV2, and ResNet50V2 across three different datasets (Sugarcane, Paddy Doctor, and Plant Village Maize). Accuracy (99.52%), F1 Score (98.25%) (green), Recall (97.99 %) (red), Precision (98.21%) (blue). The picture shows that the suggested DMFL-ConvoNet design may focus on the affected regions as compared to the conventional state-of-the-art techniques. InceptionResNetV2, ResNet50V2, EfficientNetV2L, EfficientNetV2B0, and NASNetMobile all probably focus too much on diseased areas while neglecting other beneficial elements of the condition as shown in figure 8. Given its ability to collect data at many scales, the suggested multi-kernel strategy concentrates on discriminating regions.

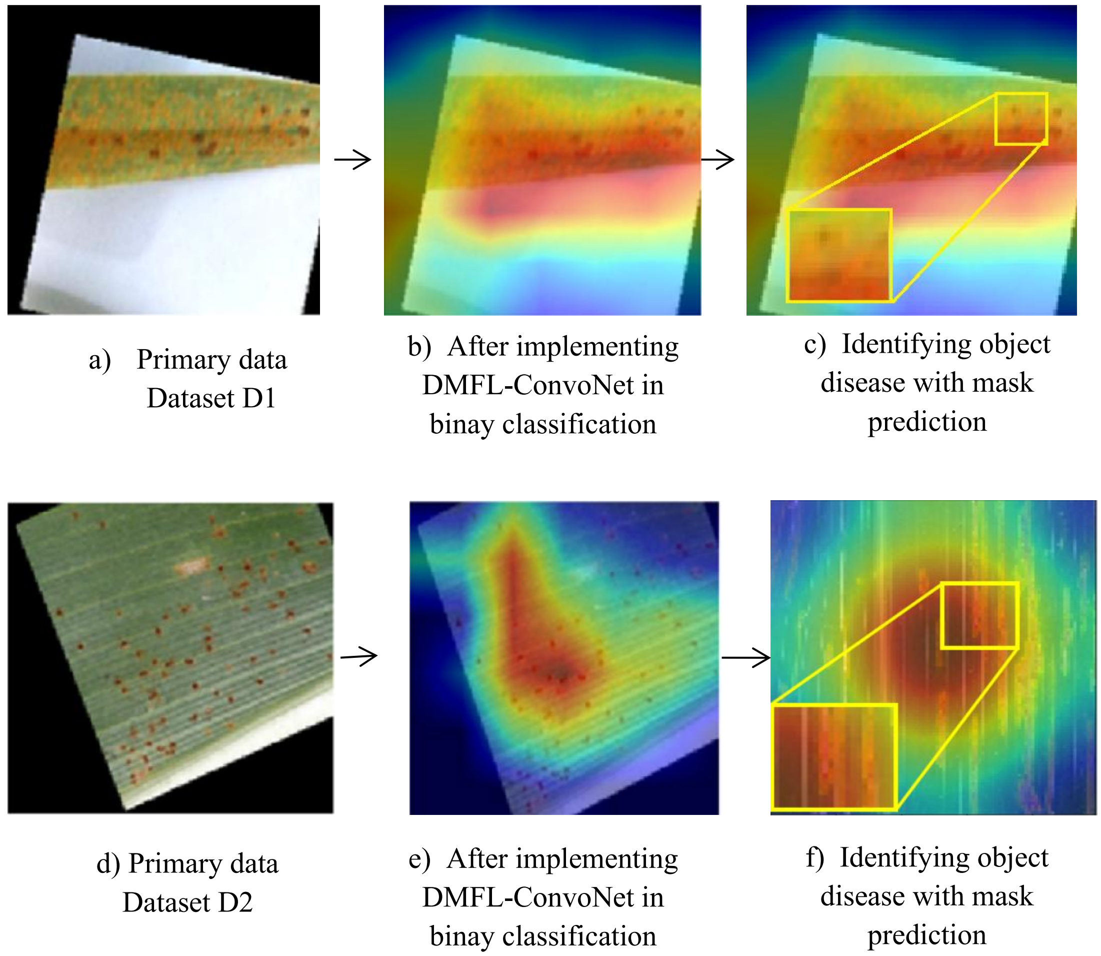


Figure 8. The affected parts of the disease in a leaf are identified using activation maps using the proposed DMFL-ConvoNet model. a) Primary data Dataset D1, b) After implementing DMFL-ConvoNet in binay classification c) Identifying object disease with mask prediction d) Primary data Dataset D2, e) After implementing DMFL-ConvoNet in binary classification f) Identifying object disease with mask prediction

5. Conclusion

The utility of the collection might be increased by using image segmentation algorithms to separate leaves from the photos. We believe that the development of scalable, computer vision-based plant disease identification requires this dataset as a prerequisite. 2.6 million parameters are needed for this plant disease detection algorithm. The suggested architecture meets memory requirements and real-time computations with remarkable accuracy and efficiency. The multi-kernel depthwise convolution improves the multiscale feature extraction for plant disease diagnostics by employing 3 × 3 and 5 × 5 kernels. The computational complexity and properties of the network are lessened by the depthwise convolution. The proposed method uses sixty-three times fewer parameters than the proposed DMFL-ConvoNet architecture, ResNet50V2, InceptionResNetV2, NASNetMobile, EfficientNetV2B0, and EfficientNetV2L. Furthermore, the suggested DMFL-ConvoNet architecture achieves 99.52% accuracy and a processing speed of 35 frames per second (FPS). The proposed model has outperformed ResNet50V2, InceptionResNetV2, NASNetMobile, EfficientNetV2B0, and EfficientNetV2L in terms of accuracy and speed. The D1, D2, and D3 datasets—which have respective accuracy rates of 99.52%, 98.89%, and 97%—as well as the new D1 dataset that we developed, are used to test the proposed method. In order to determine how the proposed method impacts real-time plant disease identification, it may be tested in the future utilizing devices with limited resources, including cellphones and agricultural robots.

Data Availability Statement: The datasets used during the current study are available from the corresponding author on reasonable request.

Acknowledgment: I pay thanks to my Supervisor Dr. **Ashok Kumar and** Co-Supervisor **Dr. Shelendra Pal** who have given guidance and light to me during this Paper. Their versatile knowledge have helped me in the critical times during the span of this Research paper.

Authors Contributions: Author 1(Arun Mittal* worked on Section 3 and Section 4 Methodology and results), Author 2 (Ashok Kumar contributed in Introduction and Literature Review) and Author 3(Shelendra Pal contributed in Literature review and in conclusion.

Funding: No Funding

Competing Interests: The authors declare no competing interests.

References

1. Mishra S, Sachan R, Rajpal D. Deep convolutional neural network based detection system for real-time corn plant disease recognition. Procedia Computer Science. 2020 Jan 1;167:2003-10.

2. Shoaib M, Hussain T, Shah B, Ullah I, Shah SM, Ali F, Park SH. Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease. Frontiers in plant science. 2022 Oct 7;13:1031748.

3. Das S, Biswas A, Sinha P. Deep learning analysis of rice blast disease using remote sensing images. IEEE Geoscience and Remote Sensing Letters. 2023 Feb 13;20:1-5.

4. Padhi J, Mishra K, Ratha AK, Behera SK, Sethy PK, Nanthaamornphong A. Enhancing paddy leaf disease diagnosis-a hybrid CNN model using simulated thermal imaging. Smart Agricultural Technology. 2025 Mar 1;10:100814.

5. Subbarayudu C, Kubendiran M. An automated hybrid deep learning framework for paddy leaf disease identification and classification. Scientific Reports. 2025 Jul 24;15(1):26873.

6. Paul H, Udayangani H, Umesha K, Lankasena N, Liyanage C, Thambugala K. Maize leaf disease detection using convolutional neural network: a mobile application based on pre-trained VGG16 architecture. New Zealand Journal of Crop and Horticultural Science. 2025 Apr 3;53(2):367-83.

7. Meng Y, Zhan J, Li K, Yan F, Zhang L. A rapid and precise algorithm for maize leaf disease detection based on YOLO MSM. Scientific Reports. 2025 Feb 19;15(1):6016.

8. Shandilya G, Gupta S, Mohamed HG, Bharany S, Rehman AU, Hussen S. Enhanced Maize Leaf Disease Detection and Classification Using an Integrated CNN-ViT Model. Food Science & Nutrition. 2025 Jul;13(7):e70513.

9. Sun L, He J, Zhang L. CASF-MNet: multi-scale network with cross attention mechanism and spatial dimension feature fusion for maize leaf disease detection. Crop Protection. 2024 Jun 1;180:106667.

10. Rai CK, Pahuja R. Northern maize leaf blight disease detection and segmentation using deep convolution neural networks. Multimedia Tools and Applications. 2024 Feb;83(7):19415-32.

11. Akkamahadevi C, Adaickalam V. An effective deep learning framework for diseases prediction to enrich paddy production. International Journal of System Assurance Engineering and Management. 2025 Jul 24:1-0.

12. Bachhal P, Kukreja V, Ahuja S, Lilhore UK, Simaiya S, Bijalwan A, Alroobaea R, Algarni S. Maize leaf disease recognition using PRF-SVM integration: a breakthrough technique. Scientific Reports. 2024 May 3;14(1):10219.

13. Khan I, Sohail SS, Madsen DØ, Khare BK. Deep transfer learning for fine-grained maize leaf disease classification. Journal of Agriculture and Food Research. 2024 Jun 1;16:101148.

14. Srinivasan S, Prabin SM, Mathivanan SK, Rajadurai H, Kulandaivelu S, Shah MA. Sugarcane leaf disease classification using deep neural network approach. BMC Plant Biology. 2025 Mar 4;25(1):282.

15. Rajput K, Manwal M, Chauhan RK, Kukreja V, Mehta S. Transforming sugarcane leaf diseases pathology with convolutional neural networks and SVM. In2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) 2024 Mar 14 (pp. 1-6). IEEE.

16. Qaadan S, Alshare A, Ahmed A, Altartouri H. Stacked ensembles powering smart farming for imbalanced sugarcane disease detection. Applied Sciences. 2025 Mar 5;15(5):2788.

17. Jardawala HA. Deep learning-based sugarcane leaf disease detection using Xception: a precision agriculture approach. InIET Conference Proceedings CP920 2025 May 2 (Vol. 2025, No. 7, pp. 1654-1660). Stevenage, UK: The Institution of Engineering and Technology.

18. Ruprah TS, Gupta A, Jadwal P, Mokashi V. Sugarcane Disease Detection and Classification Using Deep Learning Techniques. In2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI) 2025 Apr 4 (Vol. 3, pp. 246-250). IEEE.

19. https://www.kaggle.com/datasets/smaranjitghose/corn-or-maize-leaf-disease-dataset

20.https://ieee-dataport.org/documents/paddy-doctor-visual-image-dataset-automated-paddy-disease-classification-and-benchmarking

21. https://www.kaggle.com/datasets/nipunrajput/sugarcane-dataset

22. Ai Y, Sun C, Tie J, Cai X. Research on recognition model of crop diseases and insect pests based on deep learning in harsh environments. IEEE Access. 2020 Sep 21;8:171686-93.

23. Zeng Q, Ma X, Cheng B, Zhou E, Pang W. Gans-based data augmentation for citrus disease severity detection using deep learning. IEEE Access. 2020 Sep 18;8:172882-91.

24. Thomas S, Kuska MT, Bohnenkamp D, Brugger A, Alisaac E, Wahabzada M, Behmann J, Mahlein AK. Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective. Journal of Plant Diseases and Protection. 2018 Feb;125(1):5-20.

25. Jiang P, Chen Y, Liu B, He D, Liang C. Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. Ieee Access. 2019 May 6;7:59069-80.

26. Chen WL, Lin YB, Ng FL, Liu CY, Lin YW. RiceTalk: Rice blast detection using Internet of Things and artificial intelligence technologies. IEEE Internet of Things Journal. 2019 Oct 16;7(2):1001-10.

27. Haque MA, Marwaha S, Deb CK, Nigam S, Arora A, Hooda KS, Soujanya PL, Aggarwal SK, Lall B, Kumar M, Islam S. Deep learning-based approach for identification of diseases of maize crop. Scientific reports. 2022 Apr 15;12(1):6334.

28. Zhang X, Qiao Y, Meng F, Fan C, Zhang M. Identification of maize leaf diseases using improved deep convolutional neural networks. Ieee Access. 2018 Jun 6;6:30370-7.

29. Sun J, Yang Y, He X, Wu X. Northern maize leaf blight detection under complex field environment based on deep learning. IEEE access. 2020 Feb 13;8:33679-88.

30. Rauf HT, Saleem BA, Lali MI, Khan MA, Sharif M, Bukhari SA. A citrus fruits and leaves dataset for detection and classification of citrus diseases through machine learning. Data in brief. 2019 Oct 1;26:104340.