Tanzania Journal of Engineering and Technology 2025, **43(3)**: 17-33

Websites: https://commons.udsm.ac.tz/tjet/; https://ajol.org/tjet

ISSN 1821-536X (**print**); ISSN 2619-8789 (**electronic**)
https://doi.org/10.52339/tjet.v44i3.1364

Regular Research Manuscript

CNN-Based Hybrid Model for Detecting Blight Diseases in Potato Crops with Advanced Image Processing Techniques

Farian S. Ishengoma†, Hussein R. Mkwazu, and Baraka W. Nyamtiga

Department of Informatics and Information Technology, College of Natural and Applied Sciences, Sokoine University of Agriculture, P.O. Box 3000, Morogoro, Tanzania.
†Corresponding author: farishen@sua.ac.tz; ; ORCID: 0000-0002-8093-2259

ABSTRACT

Potato production plays a vital role in global agriculture as a major food source for large populations. However, potato crops are highly susceptible to diseases, particularly Early Blight and Late Blight, which result in substantial yield losses. Timely detection and effective control of these diseases are essential for maintaining stable crop output. This study explores the integration of Convolutional Neural Networks (CNNs) and advanced image processing techniques to differentiate between diseased and healthy potato plants accurately. Two datasets comprising original and enhanced images were used to train four CNN models: InceptionV3, Xception, Densenet201, and Resnet152V2. The original images underwent background removal only, whereas the enhanced images were further processed using contrast enhancement and morphological transformation in addition to background removal to reduce noise, improve quality, and prepare the images for analysis. The CNN models were trained using these datasets, with their bottom layers fixed and the top layers fine-tuned to improve performance and reduce training time. Experimental results revealed that models trained on enhanced images achieved a 2.45% to 4.45% improvement in accuracy, precision, and sensitivity compared to those trained on original images. Moreover, a hybrid model that combined two high-performing CNNs achieved a 98.91% accuracy, marking up to 10.69% improvement over individual models. This approach offers significant potential for reducing crop yield losses while minimizing dependence on chemical treatments.

ARTICLE INFO
Submitted: Oct. 21, 2024
Revised: Aug. 18, 2025
Accepted: Aug. 30, 2025
Published: Oct. 2025

Keywords: *Potato, convolution neural network, morphological transformation, image processing.*

INTRODUCTION

Potato farming is globally vital as a primary food source, yet potato plants are highly susceptible to such diseases as Early Blight and Late Blight, which can severely reduce yields. Timely detection and management of these diseases are critical for agricultural yield. Failure to detect plant diseases and pests early can result in financial losses and decreased productivity despite the use of pesticides (Ishengoma & Lyimo, 2024; Sharma et al., 2019). Recent breakthroughs in artificial intelligence (AI), particularly deep learning (DL) and computer vision (CV) have enabled data-driven disease detection systems that analyze plant images with high accuracy, helping farmers minimize the spread and reduce pesticide reliance. Utilization of these technologies, therefore, boosts the adoption of sustainable and environmentally conscious farming methods (Shoaib et al., 2023; Tian et al., 2020).

Several CV studies have addressed potato leaf disease detection using either traditional machine learning (ML) (e.g., support vector machines (SVMs)) or DL techniques, specifically convolution neural networks (CNNs). SVM/ML-based approaches typically combine segmentation (e.g. K-means, graph-cut/Otsu thresholding) and texture features (e.g., Gray Level Co-occurrence Matrix (GLCM)) with classical classifiers. For example, (Islam et al., 2017) trained SVMs on 300 images from the publicly accessible Plant-Village dataset and reported a 95% accuracy; (Singh & Kaur, 2021) used an approach that integrates K-means, GCLM and multi-class SVM on the same dataset and reported a 95.99% accuracy; (Hou et al., 2021) extracted local binary patterns (LBP) features and, across five categories, by evaluating four classifiers (k-NN, SVM, ANN, and RF), trained on 2,840 images from the AI Challenger Global AI Contest, achieved 97.4% with SVM. Moreover, (Iqbal & Talukder, 2020) used Plant-Village database with over 450 images of healthy and diseased potato leaves to evaluate seven classification algorithms, with the Random Forest classifier and achieved a 97% accuracy. All these pipelines rely on hand-crafted features and often assume controlled imaging conditions.

On the other hand, the CNN-based approaches reduce manual feature design and generally improve robustness. (Khobragade et al., 2022) trained a CNN on a dataset of 5,162 original and 82,592 augmented images, achieving a classification accuracy of 98.07%. Nishad et al. (2022) evaluated such CNN models as VGG16, VGG19, and ResNet50 with K-means segmentation and augmentation for detecting and classifying potato leaf diseases. They utilized the Plant-Village and Mendeley datasets containing 2,580 images and were able to achieve an accuracy of 97% using VGG16 that outperformed the others (Nishad et al., 2022). (Sanjeev et al., 2021) reported a 96.5% accuracy using the Feed Forward Neural Network (FFNN) model on the Plant-Village dataset comprising a total of 2,152 images. Furthermore, Khalifa et al. (2021), proposed a CNN hierarchical architecture comprising 14 layers among which are two principal convolutional layers, tasked with extracting features, as well as two fully connected layers, for classification. The model underwent training using a dataset consisting of 1,722 images, and achieved a mean accuracy of 98.00% (Khalifa et al., 2021).

Collectively, prior studies depict high accuracy under controlled imaging, but most apply limited preprocessing techniques and evaluate only individual architectures separately, leaving open the questions whether selective image enhancement can improve CNN performance and how best to combine the various models. To address this gap, we integrate an image enhancement pipeline, including background removal, contrast adjustment, and morphological transformations (erosion and opening), with modern CNNs. We then train identical architectures on two matched datasets; a baseline set with background removal only and an enhanced set with additional contrast and morphology.

The contributions of our work are threefold, namely;

- A targeted image-enhancement pipeline, background removal, contrast adjustment, and morphological transformations (erosion/opening), tailored to visible symptoms of leaf diseases.
- A controlled comparison that trains identical CNN backbones on original versus enhanced images to quantify the effect of preprocessing.
- A lightweight parallel hybrid that fuses the two best-performing CNNs to improve accuracy and stability.

On a three-class potato leaf dataset (healthy, early blight, late blight), targeted enhancement consistently improved single-model accuracy, precision, and recall by 2.45% to 4.45%. A parallel hybrid that fuses the two top-performing CNNs achieved 98.91% accuracy, which is 10.69% higher than the lowest-performing individual model (88.22%). Taken together, these results indicate that principled preprocessing, by reducing image noise and improving pattern recognition on potato leaves, mitigates residual errors in CNN classifiers, and that a simple two-stream hybrid further improves reliability without heavy architectural changes.

The remaining sections of this work are organized as follows. The following Section describes the materials and methods, while the third section provides a description of the experimental setup. Section 4 contains the discussion and results of the experiment, and Section 5 concludes the study.

MATERIALS AND METHOD

Dataset

The study used a potato leaf disease dataset that is publicly available and can be accessed from Kaggle (Muhammad et al., 2020.; Rashid et al., 2021). The potato dataset contains 4072 images with 256x256 pixels that are divided into three categories: Early blight (Eab) with 1628 images, Healthy (He) with 1020 images, and Late blight (Lab) with 1412 images. Figure 1 shows the three image categories of potato leaves taken in a controlled environment.

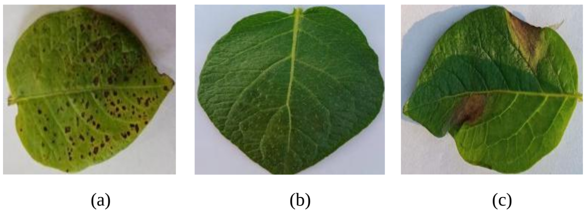


Figure 1: Potato leaf dataset (a) Early blight (b) Healthy (c) Late blight

Preprocessing method

We applied four preprocessing techniques to reduce image noise and highlight disease-related features: background removal, contrast enhancement, threshold segmentation, and morphological transformation (erosion and morphological opening), as illustrated in Figure 2. Background removal eliminates non-leaf areas, ensuring the model focuses on the leaf itself and avoids misleading textures. Contrast enhancement adjusts brightness and contrast, helping early CNN layers capture meaningful patterns more effectively. Thresholding generates a clean leaf mask, preventing background pixels from being included after segmentation. Erosion, the most effective method, removes small bright specks and trims noisy edges, sharpening lesion boundaries and enhancing feature clarity. Opening (erosion followed by dilation) breaks thin objects and smooths outlines but can also remove fine lesion details, which accounts for its slightly lower accuracy compared to erosion alone.

Backgrounds were removed using Python's "rembg" function, which applies a neural network model trained on a large dataset to remove backgrounds, thus reducing distracting noise. Next, we normalize exposure with contrast adjustment using equations (1) - (3). For each image, we compute its mean brightness B, set B_{ref} to the maximum per-image mean brightness on the training dataset, and rescale pixel intensities by the maximal intensity factor (IF), determined as $IF = \frac{B_{ref}}{B}$ with values clipped to [0, 255] (Demonstrated in Figure

3(a) and (b)). We then threshold the contrast-adjusted image to isolate the leaf foreground, operating per channel and combining masks as in equations (4) - (6); this step further suppresses background and makes foreground patterns more prominent.

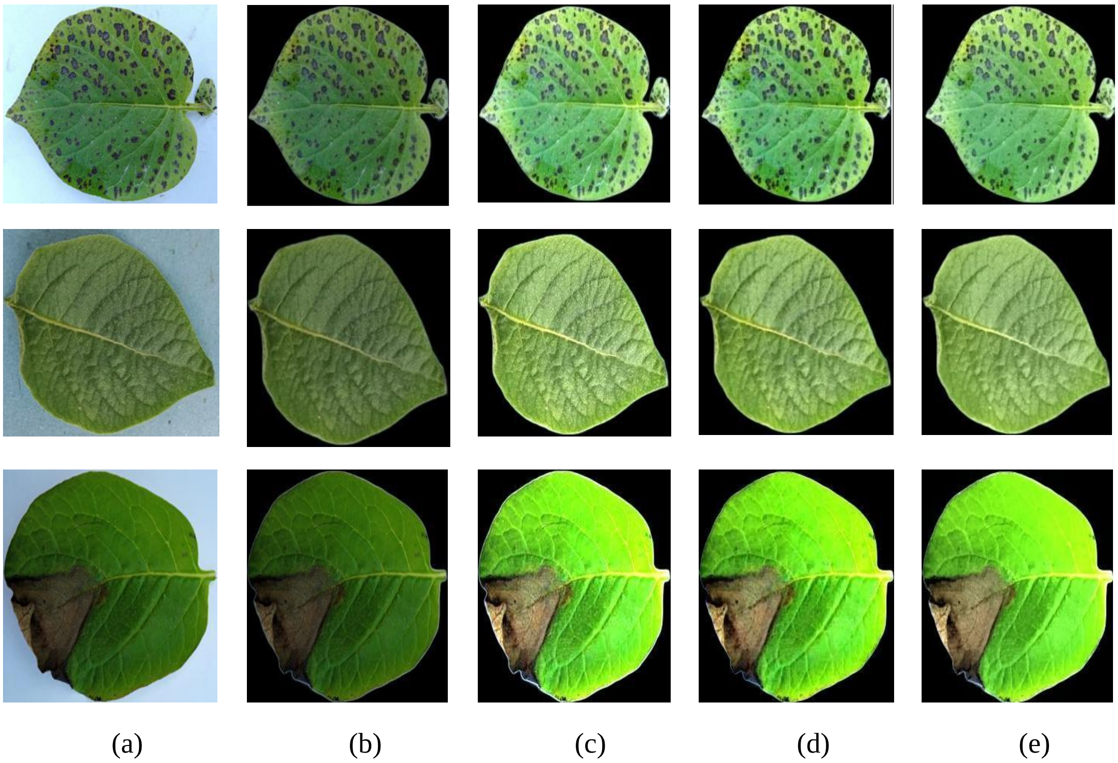


Figure 2: The pre-processing procedure used (a) Original image (b) Background removal (c) Brightness enhanced (d) Morphological transformation-erosion (e) Morphological transformation-open.

Finally, we apply morphological modification techniques to the thresholded image; the morphological-erosion and morphological-opening (erosion followed by dilation) to reduce small artifacts and smooth boundaries (as shown in Figure 3(c), (d), and (e)).

We evaluate two morphology variants; erosion-only and morphological opening, and the full preprocessing workflow is outlined in algorithm 1. It is worth noting that standalone dilation was not applied in our study. Although dilation can link bright regions or enhance features, in our setting it increased regional brightness and reduced efficiency, so we restrict its use to the dilation step within opening only.

$$B = \frac{1}{WH}\sum_{x=1}^{W}\sum_{y=1}^{H} I(x,y) \tag{1}$$

$$IF = \frac{B_{ref}}{B} \tag{2}$$

$$I'(x,y) = clip(IFI(x,y), 0, 255) \tag{3}$$

$$M_c(x,y) = \begin{cases} 1, & I'_c(x,y) > T_c \\ 0, & otherwise \end{cases} \quad c \in \{R, G, B\} \tag{4}$$

$$M(x,y) = M_R(x,y) \vee M_G(x,y) \vee M_B(x,y) \tag{5}$$

$$I_S(x,y) = I'(x,y) \cdot M(x,y) \tag{6}$$

Whereby; $I(x,y) \in [0{,}255]$ is the input intensity; $W \times H$ is image size in pixel and $clip(v, 0{,}255) = min(max(v, 0){,}255)$. The thresholds T_c can either be fixed or computed by Otsu. In our study, we adopt fixed global thresholds T_R , T_G, and T_B selected on the training set.

1. Equation (1) computes the average brightness of the image as the summation of all pixel intensities divided by the number of pixels.
2. Equation (2) defines an image-specific scaling factor. If the image is darker than the reference (i.e., $B < B_{ref}$), then $IF > 1$ and pixels will be brightened; if brighter, $IF < 1$ and pixels will be dimmed.
3. Equation (3) rescales each scale by IF and clips the result to the valid 8-bit range $[0{,}255]$, producing the contrast adjusted image I'.
4. Equation (4) builds a binary mask per channel (R, G, B), marking pixel as foreground (assigned value 1) where the contrast-adjusted value exceeds the channel threshold T_c ; otherwise background (value 0).
5. Equation (5) combines the three masks with a logical OR to get a single foreground mask M.
6. Equation (6) applies the mask to the contrast-adjusted image, keeping foreground pixels (where $M = 1$) and zero out background, yielding the segmented image I_S used for morphology.

***Algorithm* 1:** Image Preprocessing Pipeline (background removal, contrast adjustment, thresholding, morphology)

Step 1	Insert a 256x256 image.
Step 2	Using the rembg function, remove the background from the image
Step 3	Use Equation (1), to calculate the image's mean brightness B
Step 4	Use Equation (2) to calculate the IF as B_{ref}/B, whereby B_{ref} is the precomputed training-set reference brightness.
Step 5	Use Equation (3) to compute contrast adjustment (rescaling by IF and clip to $[0, 255]$) to obtain the enhanced image I'.
Step 6	Separate the enhanced image I' into red, green, and blue channels to create three grayscale images.
Step 7	Use Equations (4) to (6), to threshold each channel to build masks, combine the masks, and apply the combined mask to get the segmented image.
Step 8	Apply erosion on the image created in Step 7 with a 2x2 kernel.
Step 9	Combine the eroded channel images from Step 8 to create the eroded three-channel image.
Step 10	Apply dilation to the images from Step 8 with a 2x2 kernel (for the opening variant).
Step 11	Combine the images from Step 10 to get the dilated three-channel image.

Note: *Stop after Step 9 for the erosion-only variant; continue through Step 11 for the morphological opening variant.*

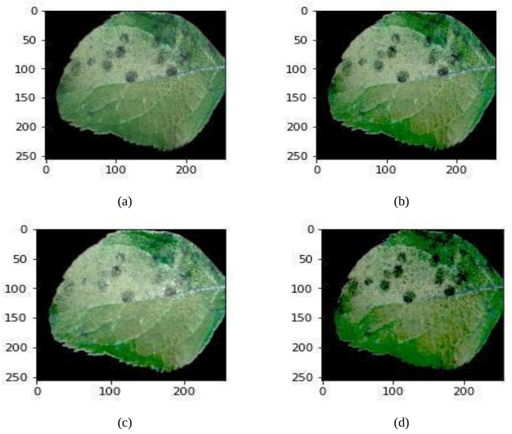


Figure 3: Comparison of images (a) Brightness enhanced (b) Threshold segmentation (c) Morphological-open (d) Morphological-erosion.

Convolution Neural Network

CNN plays a vital role in the field of DL and is widely utilized in the domain of computer vision. CNNs can detect visual patterns inside picture pixels with minimal preprocessing (Alzubaidi et al., 2021; Kattenborn et al., 2021; Tian et al., 2020). As a result, they have been widely employed in research endeavors pertaining to image-based detection.

In this study, we selected InceptionV3, Xception, DenseNet201, and ResNet152V2 to provide a diverse evaluation of CNN architectures. InceptionV3 was chosen for its ability to capture multi-scale features through parallel convolutional filters, making it suitable for detecting complex disease patterns (He et al., 2016). Xception employs depthwise separable convolutions, offering computational efficiency while maintaining strong feature extraction (Chollet, 2016). DenseNet201 leverages dense connectivity to promote feature reuse and mitigate vanishing gradients, which is advantageous when working with limited datasets (Huang et al., 2017). ResNet152V2, with its very deep residual learning framework, enables robust representation learning and effectively handles complex image variations (He et al., 2015). Together, these models represent complementary design strategies multi-scale feature learning, efficiency, dense connectivity, and deep residual learning providing a comprehensive basis for evaluating performance in crop disease classification.

The GoogleNet architecture presents a viable alternative to CNNs with the incorporation of an inception module, which introduces supplementary layers. The GoogleNet architecture employs a substitution of average pooling for fully connected layers located at the topmost portion of the convolutional network, leading to a notable decrease in the number of parameters (Szegedy et al., 2016). The architectural design of GoogleNet was built with the primary aim of reducing energy consumption and optimizing memory utilization. This study additionally examines InceptionV3, a model that is expected to supersede both InceptionV1 and InceptionV2 (He et al., 2016). The Xception on the other hand exhibited superior performance compared to InceptionV3, despite both models having an equal number of parameters. This notable difference in performance was observed specifically on the largest dataset. The main improvement in this model involves the substitution of conventional convolution layers with depth-wise separable convolutions. This modification entails decomposing the convolution operation into two distinct components: depth-wise convolutions and point-wise convolutions. This methodology effectively decreases the quantity of parameters and computations, while also upholding or enhancing the accuracy of the model (Chollet, 2016).

Huang et al., proposed the DenseNet-201 model, which is an expanded version of the DenseNet framework that exhibits dense connectivity patterns between layers. The architectural design, consisting of a total of 201 layers, effectively improves the utilization of features and the flow of gradients, all while ensuring computational efficiency (Huang et al., 2017). Consequently, this design exhibits exceptional performance in the domain of picture classification tasks.

ResNet model was proposed by Zhang, et al. to overcome the vanishing gradient problem in deep neural networks. The model incorporates residual blocks with skip connections, allowing for the training of exceptionally deep networks. This approach has demonstrated notable success in the realm of image recognition applications (He et al., 2015).

The CNN design is made up of three types of layers: convolution layers, pooling layers, and fully connected layers (FC). In a neural network, feature extraction is achieved using convolution and pooling layers, whereas the subsequent step of mapping these features into the final output, which is frequently represented by a softmax function, is accomplished using an FC layer. The CNN-based models presented in this study use input images with dimensions of 150x150 pixels, as shown in Figure 4. Convolutional techniques are applied to these images before they are connected to an FC layer of size 128. On the other hand, Figure 5 shows a hybrid CNN model with two models stacked in parallel and sharing the input data as well as output, which is connected with FC layers with 128 neurons and a softmax layer as an output.

EXPERIMENTAL SETUP

The framework of the proposed method is shown in Figure 6 with a 256 x 256-pixel input image. The input images are sent through the background removal block to keep the required section before being routed through the contrast enhancement block. This block computes the greatest intensity acquired from all images and uses it as the lowest intensity for all images. These RGB (red-green-blue) images are divided into three grayscale channels, and the threshold approach is used to separate the foreground and background. It is necessary to separate images into grayscale because morphological transformation procedures are applied to grayscale images. At this stage, morphological erosion is used first, followed by dilation as illustrated in Algorithm. The process of performing the erosion followed by dilation is known as Morphological-open. Later, the three grayscale image channels are combined to generate the RGB images that are sent into the CNN-based models. For quicker processing, the model's input block converts images to 150 x 150 pixels. Four sets of data are used to

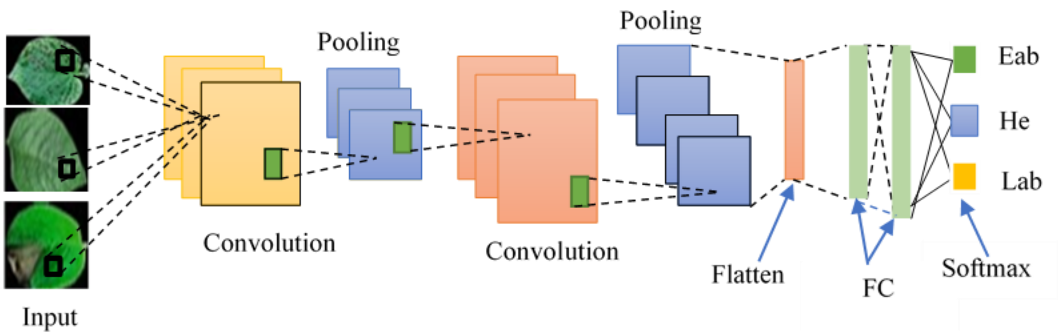


Figure 5:Convolution neural network architecture.

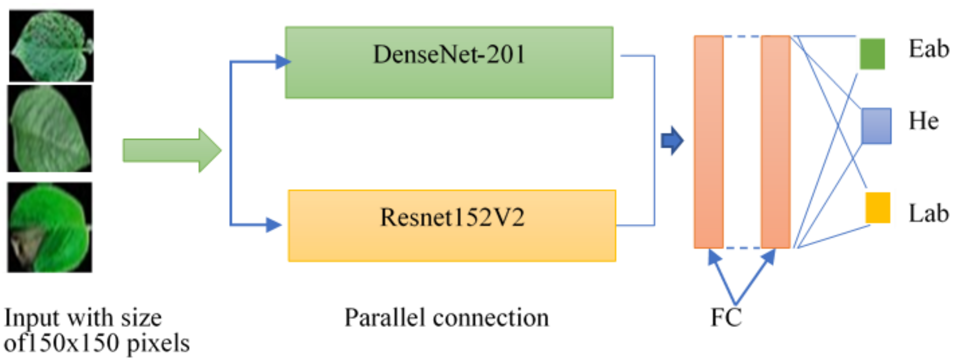


Figure 4:Proposed Hybrid convolution neural network architecture.

train models using the InceptionV3, Xception, DenseNet201, and ResNet152V2 for good comparison: original images, contrast enhanced images, morphological-erosion, and morphological-open. After testing the CNN-based models on four different sets of data, the images trained on morphological erosion outperformed the others by up to 4.45%. For further improvement and comparing with the state-of-the-art methods, the two models that performed better were merged in parallel using the same input and attained an overall accuracy of 98.91% and 0.98 of F1 score. The Adam optimizer was used for the training, with a 32-fixed batch size, a learning rate of 0.001, and a dropout rate of 0.15. All models were trained on a system powered by an Intel Core i5-7200u CPU (2.50 GHz, 2.71 GHz) and 16 GB of RAM. Furthermore, multiple augmentation techniques such as vertical flip, zoom range 0.5, and rotation for 90 degrees were applied during training for good generalization.

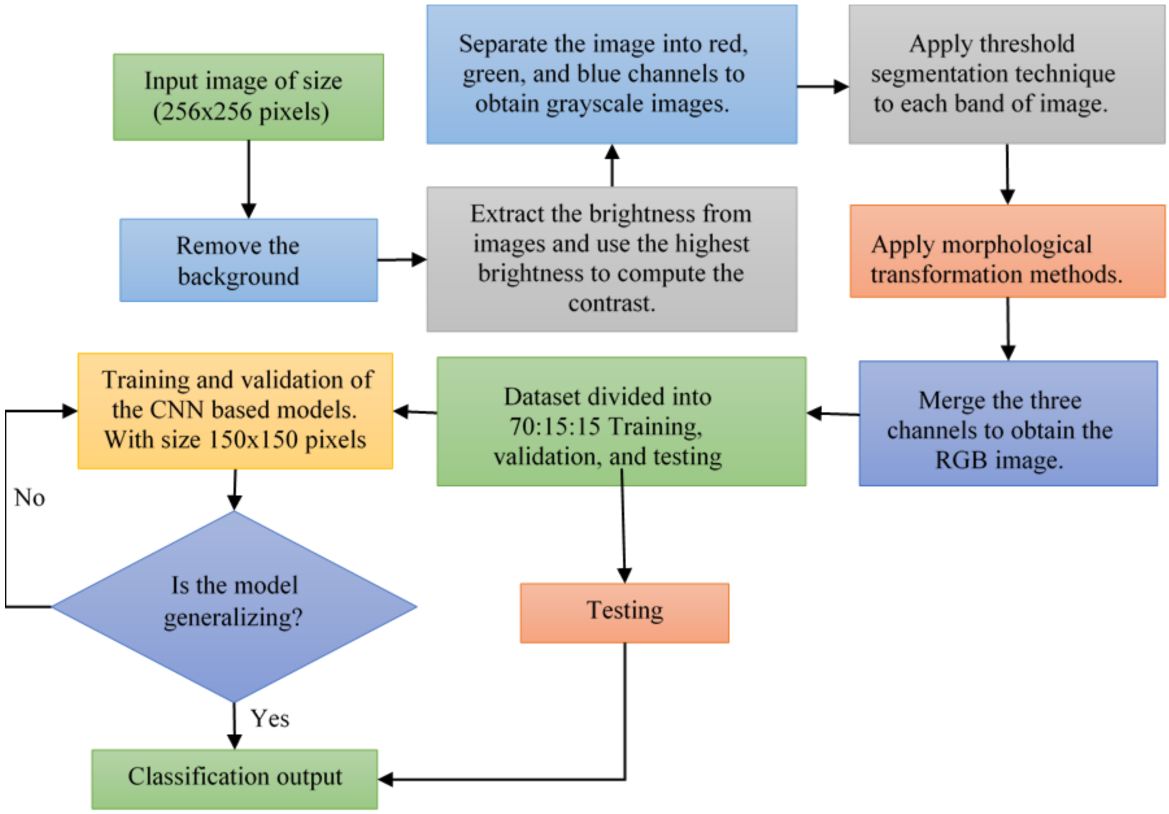


Figure 6: Framework of the proposed method.

Evaluation Parameters

To evaluate the classification models, we use four parameters listed in equations 7-10, namely accuracy (Acc), precision (Pre), recall (Rec), and F1-score (F_1). Accuracy measures the ratio of correctly classified images to the total number of images. Precision indicates the proportion of images predicted as a specific disease that are truly correct, meaning a high precision corresponds to fewer false positives. Recall represents the proportion of diseased images that are correctly identified, where a higher recall reduces false negatives. Finally, the F1-score combines precision and recall into a single measure, making it especially valuable when class distributions are imbalanced, as it reflects the model's ability to minimize both false positives and false negatives.

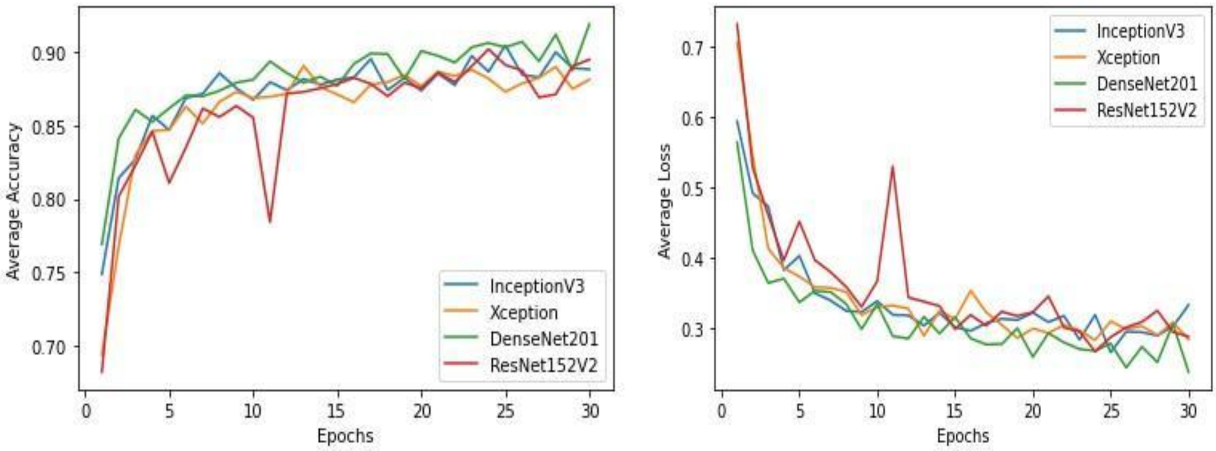
$$A_{cc} = \frac{T_{Ne} + T_{Po}}{T_{Ne} + T_{Po} + F_{Po} + F_{Ne}} \tag{7}$$

$$P_{re} = \frac{T_{Po}}{T_{Po} + F_{Po}} \quad (8)$$

$$R_{ec} = \frac{T_{Po}}{T_{Po} + F_{Ne}} \quad (9)$$

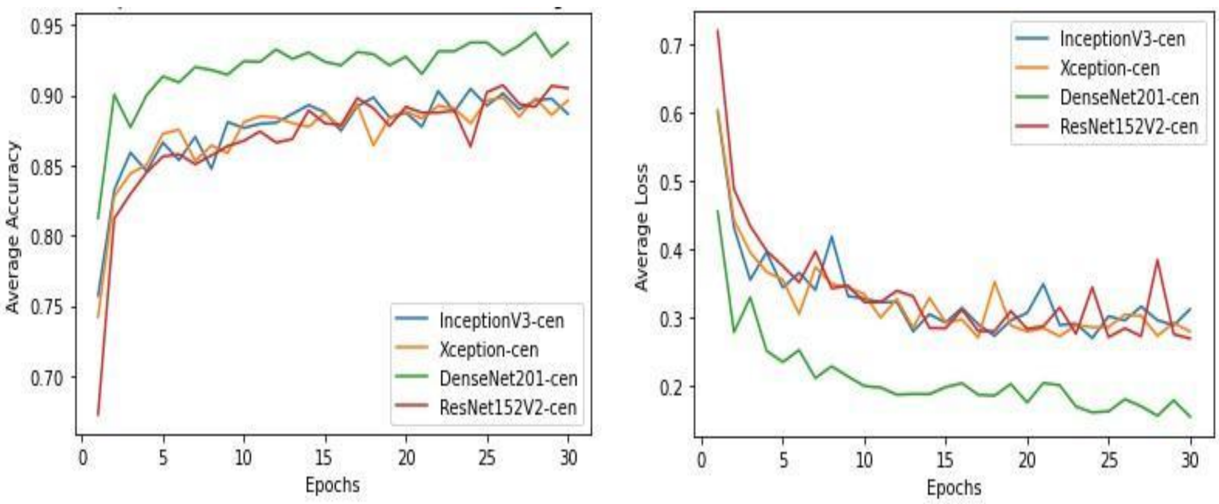
$$F_1 = 2((P_{re}.R_{ec})/(P_{re} + R_{ec})) \quad (10)$$

The term "True Negative (T_{Ne})" refers to the accurate prediction of the negative class by the model. A True Positive (T_{Po}) refers to a situation where the model accurately predicts the positive class. A false negative (F_{Ne}) refers to an incorrect prediction made by the model regarding the negative class. A false positive (F_{Po}) refers to a situation where a model incorrectly predicts a positive class.



(a) (b)

Figure 7:Models trained on original images which are InceptionV3, Xception, DenseNet201, and ResNet152V2. (a) Accuracy (b) Loss.



(a) (b)

Figure 8:Models trained on contrast-enhanced images which are InceptionV3-cen, Xception-cen, DenseNet201-cen, and ResNet152V2-cen. (a) Accuracy (b) Loss.

EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the experimental results of four sets of data, namely original images and three types of enhanced images: contrast enhanced, morphological erosion, and morphological-open. Additionally, this section discusses the comparison between the suggested method and current methods.

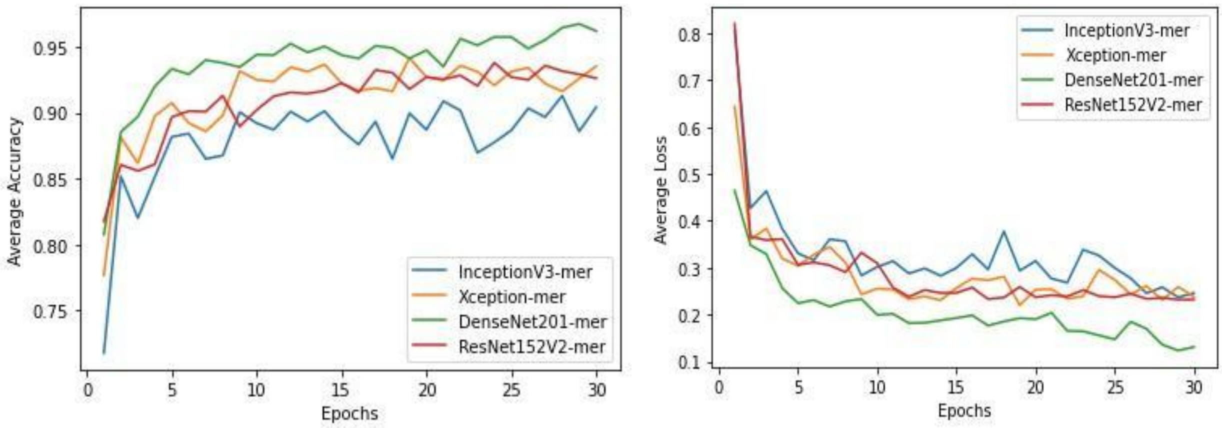
Experimental results

Figure 7 (a) and (b) show the accuracy and loss of models trained on the original images, which are InceptionV3, Xception, DenseNet201, and ResNet152V2. The average accuracy rates for the InceptionV3, Xception, DenseNet201, and ResNet152V2 models were 88.22%, 88.89%, 92.44%, and 90.22%, respectively, with mean losses of 0.35, 0.31, 0.23, and 0.28.

Figure 8 (a) and (b) show the accuracy and loss performance metrics of the models trained on contrast-enhanced images. These models are known as InceptionV3-cen, Xception-cen, DenseNet201-cen, and ResNet152V2-cen in this study. The average accuracy rates for the InceptionV3-cen, Xception-cen, DenseNet201-cen, and ResNet152V2-cen models, with mean losses of 0.33, 0.30, 0.18, and 0.28, were 88.67%, 89.78%, 93.56%, and 90.33%, respectively.

Figure 9 (a) and (b) show the accuracy and loss metrics of models trained on images enhanced with morphological erosion. These models are referred to as InceptionV3-mer, Xception-mer, DenseNet201-mer, and ResNet152V2-mer in this study. With mean losses of 0.24, 0.23, 0.12, and 0.25, the average accuracy rates for the InceptionV3-mer, Xception-mer, DenseNet201-mer, and ResNet152V2-mer models were 91.11%, 92.45%, 96.89%, and 92.67%, respectively.

Figure 10 (a) and (b) show the accuracy and loss metrics of models trained on images enhanced with morphological open. These models are known as ResNet152V2-mop, DenseNet201-mop, Xception-mop, and InceptionV3-mop in this study. The average accuracy rates for the InceptionV3-mop, Xception-mop, DenseNet201-mop, and ResNet152V2-mop models were 89.11%, 90.24%, 93.33%, and 88.89%, respectively, with mean losses of 0.36, 0.27, 0.21, and 0.34.



(a) (b)

Figure 9:Models trained on images enhanced with morphological erosion methods which are InceptionV3-mer, Xception-mer, DenseNet201-mer, and ResNet152V2-mer. (a) Accuracy (b) Loss.

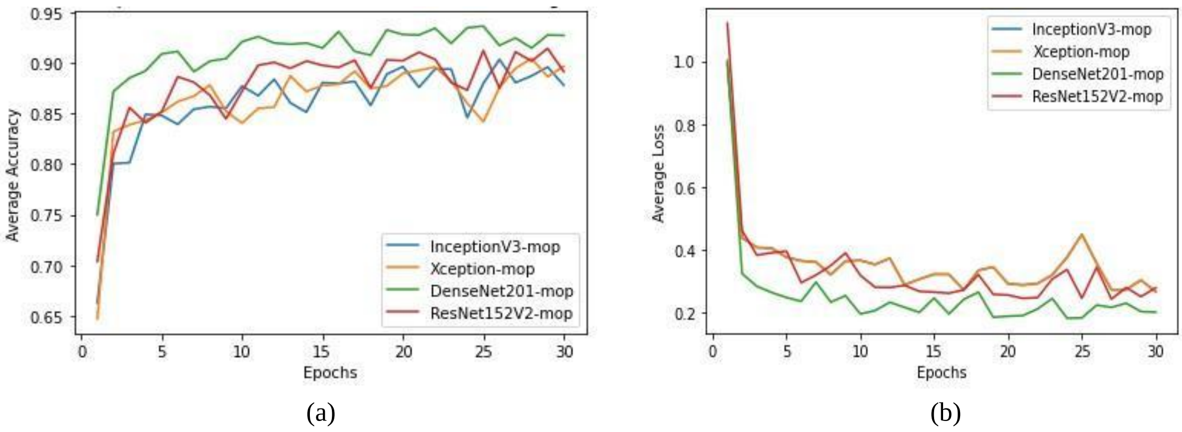


Figure 10:Models trained on images enhanced with morphological open method which are InceptionV3-mop, Xception-mop, DenseNet201-mop, and ResNet152V2-mop. (a) Accuracy (b) Loss

On the other hand, Table 1 through Table 5 shows the confusion matrices of the models trained on the four sets of data. Table 1 presents an overview of the InceptionV3, Xception, DenseNet201, and ResNet152V2 models. The InceptionV3 model exhibited poor performance, achieving an accuracy rate of 88.22%, when compared to the other three models. In contrast, the DenseNet201 model demonstrated the highest accuracy rate of 92.44%, surpassing both the Xception and ResNet152V2 models, which achieved accuracy rates of 88.89% and 90.22% respectively. Additionally, the Eab class in InceptionV3 has a poor recall of 0.82, while the He class has a low precision of 0.76 and an F1 score of 0.86. The precision, recall, and F1-score of the remaining classes vary between 0.83 and 0.99. In contrast, the precision, recall, and F1- score of the classes in the DenseNet201 model exhibit a range of values ranging from 0.9 to 0.95.

Table 1: Confusion matrix of all models trained on original images.

Class	Eab	He	Lab	Accuracy (%)	Precision	Recall	F1-score	Improvement (%)
InceptionV3								
Eab	123	24	3	88.22	0.99	0.82	0.90	-
He	0	149	1		0.76	0.99	0.86	
Lab	1	24	125		0.97	0.83	0.90	
Xception								
Eab	135	9	6	88.89	0.89	0.90	0.89	-
He	9	133	8		0.88	0.89	0.88	
Lab	8	10	132		0.90	0.88	0.89	
DenseeNet201								
Eab	139	5	6	92.44	0.95	0.93	0.94	-
He	7	138	5		0.90	0.92	0.91	
Lab	1	10	139		0.93	0.93	0.93	
ResNet152V2								
Eab	130	7	13	90.22	1.00	0.87	0.93	-
He	0	130	20		0.92	0.87	0.89	
Lab	0	4	146		0.82	0.97	0.89	

Table 2 shows the confusion matrix for the InceptionV3-cen, Xception-cen, DenseNet201-cen, and ResNet152V2-cen models. The InceptionV3-cen model showed poor performance, with an accuracy rate of 88.67%. In comparison, the DenseNet201-

cen model demonstrated the highest accuracy rate of 93.56% when compared to the Xception-cen and ResNet152V2-cen models, which achieved accuracy rates of 89.78% and 91.78%, respectively Additionally, it is observed that the He class of the InceptionV3-cen model has a precision value of 0.78. Similarly, the Lab class demonstrates a low F_1 score of 0.79 and a recall value of 0.87. The precision, recall, and F_1-score of the other classes vary between 0.89 and 0.97. In contrast, the precision, recall, and F_1-score of the classes of the DenseNet201-cen model exhibit a range of values spanning from 0.92 to 0.95. The empirical findings indicate that the utilization of contrast enhancement pre-processing techniques resulted in an enhancement of model performance ranging from 0.45% for InceptionV3-cen to 1.56% for ResNet152V2-cen.

Table 2: Confusion matrix of all models trained on contrast enhanced images.

Class	Eab	He	Lab	Accuracy (%)	Precision	Recall	F1-score	Improvement (%)
InceptionV3-cen								
Eab	134	15	1	88.67	0.96	0.89	0.92	0.45
He	1	146	3		0.78	0.97	0.87	
Lab	5	26	119		0.97	0.79	0.87	
Xception-cen								
Eab	137	8	5	89.78	0.95	0.91	0.93	0.89
He	4	136	10		0.85	0.91	0.88	
Lab	3	16	131		0.90	0.87	0.89	
DenseNet201-cen								
Eab	141	5	4	93.56	0.92	0.94	0.93	1.12
He	6	140	4		0.94	0.93	0.94	
Lab	6	4	140		0.95	0.93	0.94	
ResNet152V2-cen								
Eab	134	5	11	91.78	1.00	0.89	0.94	1.56
He	0	132	18		0.94	0.88	0.91	
Lab	0	3	147		0.84	0.98	0.90	

Table 3 shows the confusion matrix pertaining to the InceptionV3-mer, Xception-mer, DenseNet201-mer, and ResNet152V2-mer models. Among the three models evaluated, the Xception-mer model exhibited the lowest accuracy rate of 90.45%. Conversely, the DenseNet201-mer model demonstrated strong performance, with an accuracy rate of 94.89%. This surpassed the accuracy rates of the ResNet152V2-mer and InceptionV3-mer models, which achieved accuracy levels of 90.67% and 91.11% respectively. Furthermore, it is worth noting that the He class of InceptionV3-mer exhibits a precision value of 0.83. On the other hand, the Lab class showcases a relatively lower recall value of 0.81 and an F1-score of 0.88. The precision, recall, and F1-score of the remaining classes exhibit a range of values ranging from 0.90 to 0.97. On the other hand, the DenseNet201-mer model demonstrates class precision, recall, and F1-score values ranging from 0.95 to 1.00. The experimental results indicate that the utilization of erosion-based pre-processing techniques resulted in a notable improvement in model performance, ranging from 2.45% for ResNet152V2-mer to 4.45% for InceptionV3-mer.

Table 3: Confusion matrix of all models trained on Morphological erosion images

Class	Eab	He	Lab	Accuracy (%)	Precision	Recall	F1-score	Improvement (%)
InceptionV3-mer								
Eab	141	6	3	91.11	0.96	0.94	0.95	2.89
He	2	147	1		0.83	0.98	0.90	
Lab	4	24	122		0.97	0.81	0.88	
Xception-mer								
Eab	139	6	5	92.45	0.97	0.93	0.93	3.56
He	2	142	6		0.89	0.95	0.89	
Lab	3	12	135		0.92	0.90	0.89	
DenseeNet201-mer								
Eab	144	2	4	96.89	1.00	0.96	0.98	4.45
He	0	148	2		0.95	0.99	0.97	
Lab	0	6	144		0.96	0.96	0.96	
ResNet152V2-mer								
Eab	143	6	1	92.67	0.97	095	0.96	2.45
He	2	146	2		0.85	0.97	0.91	
Lab	3	19	128		0.98	0.85	0.91	

The confusion matrix for the models InceptionV3-mop, Xception-mop, DenseNet201- mop, and ResNet152V2-mop is shown in Table 4. The InceptionV3-mop model performed poorly, with an accuracy rate of 89.11%. When compared to the Xception-mop and ResNet152V2-mop models, which achieved accuracy rates of 90.24% and 88.89%, respectively, the DenseNet201-mop model achieved the highest accuracy rate of 93.33%. In addition, the He class of the InceptionV3-mop model has a precision value of 0.80, indicating a high number of false positives. Similarly, the Lab class has a poor recall score of 0.79, indicating that it has a high number of false negatives. Additionally, the Lab and He classes both achieved an F1-score of 0.88. The other classes' precision, recall, and F1-score range from 0.89 to 1.00. The precision, recall, and F1-score of the classes in the DenseNet201-mop model, on the other hand, range from 0.88 to 0.99.

The experimental results indicate that the performance of models trained with InceptionV3-mop, Xception-mop, and DenseNet201-mop improves by 0.89%, 1.35%, and 0.89%, respectively, when compared to models trained with original images. Surprisingly, the performance of ResNet152V2-mop dropped by 1.33% when compared to ResNet152V2. In addition, it was observed that the performance of these models was lower by a margin of 2% to 3.56% when compared to models trained using morphological erosion.

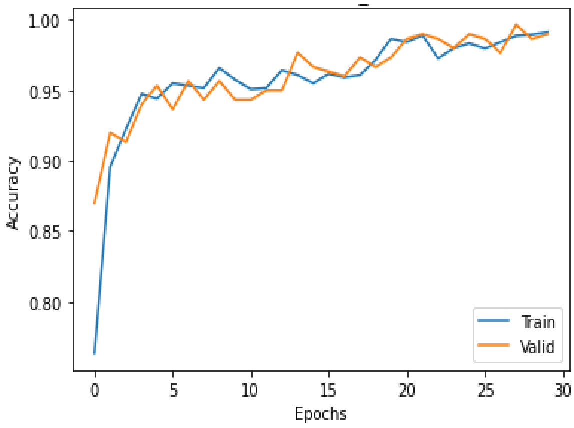
Morphological erosion method outperformed opening method because its sharpened lesion boundaries and removed small background noise without excessively altering the leaf structure. This allowed the model to focus on more discriminative disease patterns, leading to better feature learning. In contrast, opening (erosion followed by dilation) tended to smooth edges and eliminate fine lesion details, which reduced the availability of critical texture information for classification. As a result, erosion preserved key features while minimizing noise, yielding higher accuracy and lower loss.

Table 4: Confusion matrix of all models trained on Morphological-open images.

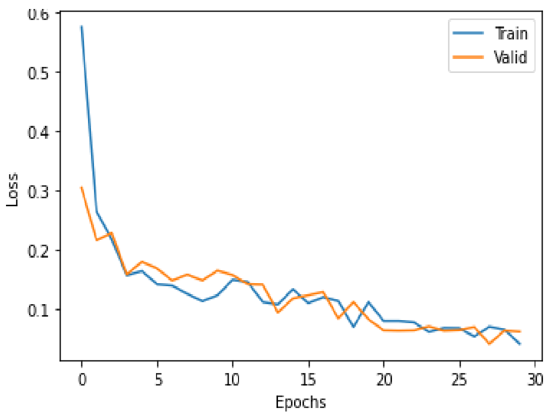
Class	Eab	He	Lab	Accuracy (%)	Precision	Recall	F1-score	Improvement (%)

InceptionV3-mop								
Eab	134	16	0	89.11	0.92	0.89	0.91	0.89
He	2	148	0		0.80	0.99	0.88	
Lab	10	21	119		1.00	0.79	0.88	
Xception-mop								
Eab	138	8	4	90.24	0.93	0.92	0.92	1.35
He	4	139	7		0.86	0.93	0.89	
Lab	7	14	129		0.92	0.86	0.89	
DenseeNet201-mop								
Eab	138	4	8	93.33	0.99	0.92	0.96	0.89
He	0	138	12		0.94	0.92	0.93	
Lab	1	5	144		0.88	0.96	0.92	
ResNet152V2-mop								
Eab	120	18	12	88.89	0.98	0.80	0.88	-1.33
He	2	139	9		0.84	0.93	0.88	
Lab	1	8	141		0.87	0.94	0.90	

To compare the proposed approach to current approaches, a hybrid CNN model was created by combining the two models that performed well which are DenseNet201-mer and ResNet152V2-mer. These models were joined in parallel and shared the same input and output. The hybrid CNN model was trained for 30 epochs and achieved an average accuracy of 98.91%, with a loss of 0.101 as shown in Figure 11 (a) and (b).



(a)



(b)

Figure 11: Accuracy and loss of the models trained on improved images (a) Hybrid CNN accuracy, (b) Hybrid CNN loss.

Table 5 shows the results of the proposed method; the model achieved an overall accuracy of 98.91%, representing a 10.67 improvement over InceptionV3. Furthermore, the class precision ranges from 0.98 to 1, the recall is 0.99, and the F1-score ranges from 0.98 to 0.99. It is worth noting that the Eab class outperformed the others, with precision score of 1 and recall and F1-scores of 0.99. Table 6 presents a comparative analysis of the performance of the proposed model and the existing state-of-the-art methodologies found in the literature for the purpose of classifying potato leaves. The table's data shows that the models provided in this study perform better than any previously recommended approaches. The proposed models achieved a precision level of 98.91%.

Table 5: Confusion matrix of the proposed hybrid CNN model

Class	Eab	He	Lab	Accuracy (%)	Precision	Recall	F1-score	Improvement (%)
Eab	148	1	1	98.91	1.00	0.99	0.99	10.69
He	0	149	1		0.98	0.99	0.98	
Lab	0	2	148		0.99	0.99	0.99	

Table 6: Comparative analysis of the performance of the proposed model and other models

Method	Applied Technique (s)	Plant disease (Classes)	Classification Accuracy
Proposed	**Hybrid-CNN**	Potato (3)	**98.91**
Khobragade et al., (2022)	CNN		98.07
Khalifa et al., (2021)	CNN		98.00
Iqbal & Talukder, (2020)	RF		97
Nishad et al., (2022)	CNN		97
Sanjeev et al, (2021)	FFNN		96.5
Singh A and Kaur H, (2021)	Multiclass SVM		95.99
Islam et al, (2017)	SVM		95.00
Hou et al., (2021)	SVM	Potato (5)	97.4

CONCLUSION

This study presented a method for detecting disease-induced patterns on potato leaves using efficient integration of CNN-based image processing and pre-processing techniques. By applying enhancement procedures such as background removal, contrast improvement, and morphological transformations (erosion and morphological open), the visibility of disease patterns on potato leaves was improved. The models trained on these enhanced images, particularly those processed with morphological transformation-erosion techniques, outperformed those trained on original images, depicting a 2.45% to 4.45% improvement in accuracy and F1-score. To further enhance the performance, the two best models, Densenet201-mer and Resnet152-mer, were combined in parallel to create a hybrid CNN model, which achieved an accuracy of 98.91% and an average F1-score of 0.98, representing a 10.69% improvement. This approach demonstrates the potential to transform disease detection in potato crops, reducing yield losses and minimizing the need for chemical interventions. The main limitation of this study lies in the sensitivity of the detection technique to background noise. When applied to images with complex or cluttered backgrounds, the method often misidentifies features, reducing its effectiveness under field conditions where such interference is common. For real-time image processing, background removal becomes necessary to improve detection, but this step introduces additional delays. Future work could explore the impact of these morphological transformations on a wider range of crop types to broaden the applicability of this approach.

Conflicts of interest: The authors claim to have no known conflicts of interest.

Data availability: The datasets analyzed in the paper are available in the Kaggle.com, https://www.kaggle.com/datasets/rizwan123456789/potato -disease-leaf -datasetpld/and https://drive.google.com/drive/folders/1FpcQA66pEg0XR8y5uEzWU REPpqSAPD

REFERENCES

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. In *Journal of Big Data* (Vol. 8, Issue 1). Springer International Publishing. doi:10.1186/s40537-021-00444-8

Chollet, F. (2016). Xception: Deep Learning with Depthwise Separable Convolutions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1800–1807.*, 1251–1258.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition*. http://arxiv.org/abs/1512.03385

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 770–778. doi:10.1109/CVPR.2016.90

Hou, C., Zhuang, J., Tang, Y., He, Y., Miao, A., Huang, H., & Luo, S. (2021). Recognition of early blight and late blight diseases on potato leaves based on graph cut segmentation. *Journal of Agriculture and Food Research*, *5*. doi:10.1016/j.jafr.2021.100154

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Iqbal, M. A., & Talukder, K. H. (2020). Detection of potato disease using image segmentation and machine learning. *2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, 43–47.

Ishengoma, F. S., & Lyimo, N. N. (2024). Ensemble model for grape leaf disease detection using CNN feature extractors and random forest classifier. *Heliyon*, ***10***(12). doi:10.1016/j.heliyon.2024.e33377

Islam, M., Dinh, A., Wahid, K., & Bhowmik, P. (2017). *Detection of Potato Diseases Using Image Segmentation and Multiclass Support Vector Machine*. 30–33. www.plantvillage.org

Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, ***173***(December 2020), 24–49. doi:10.1016/j.isprsjprs.2020.12.010

Khalifa, N. E. M., Taha, M. H. N., Abou El-Maged, L. M., & Hassanien, A. E. (2021). Artificial Intelligence in Potato Leaf Disease Classification: A Deep Learning Approach. In *Studies in Big Data* (Vol. 77, pp. 63–79). Springer Science and Business Media Deutschland GmbH. doi:10.1007/978-3-030-59338-4_4

Khobragade, P., Shriwas, A., Shinde, S., Mane, A., & Padole, A. (2022). Potato Leaf Disease Detection Using CNN. *2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, 1–5. doi:10.1109/SMARTGENCON56628.2022.10083986

Muhammad, R. S., & Javed, R. (n.d.). *Potato Disease Leaf Dataset(PLD)*. 2020. Retrieved October 11, 2023, from https://www.kaggle.com/datasets/rizwan123456789/potato-disease-leaf-datasetpld/

Nishad, M. A. R., Mitu, M. A., & Jahan, N. (2022). Predicting and Classifying Potato Leaf Disease using K-means Segmentation Techniques and Deep Learning Networks. *Procedia Computer Science*, ***212***(C), 220–229. doi:10.1016/j.procs.2022.11.006

Rashid, J., Khan, I., Ali, G., Almotiri, S. H., Alghamdi, M. A., & Masood, K. (2021). Multi-level deep learning model for potato leaf disease recognition. *Electronics (Switzerland)*, ***10***(17). doi:10.3390/electronics10172064

Sanjeev, K., Gupta, N. K., Jeberson, W. J., & Paswan, S. (2021). Early Prediction of Potato Leaf Diseases Using ANN Classifier. *Oriental Journal of Computer Science and Technology*, ***13***(0203), 129–134. doi:10.13005/ojcst13.0203.11

Sharma, A., Kumar, V., Shahzad, B., Tanveer, M., Sidhu, G. P. S., Handa, N., Kohli, S. K., Yadav, P., Bali, A. S., Parihar, R. D., Dar, O. I., Singh, K., Jasrotia, S., Bakshi, P., Ramakrishnan, M., Kumar, S., Bhardwaj, R., & Thukral, A. K. (2019). Worldwide pesticide usage and its impacts on ecosystem. *SN Applied Sciences*, ***1***(11), 1–16. doi:10.1007/s42452-019-1485-1

Shoaib, M., Shah, B., EI-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., Gechev, T., Hussain, T., & Ali, F. (2023). An advanced deep learning models-based plant disease detection: A review of recent research. In *Frontiers in Plant Science* (Vol. 14). Frontiers Media SA. doi:10.3389/fpls.2023.1158933

Singh, A., & Kaur, H. (2021). Potato plant leaves disease detection and classification using machine learning methodologies. *IOP Conference Series: Materials Science and Engineering*, ***1022***(1). doi:10.1088/1757-899X/1022/1/012121

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 2818–2826. doi:10.1109/CVPR.2016.308

Tian, H., Wang, T., Liu, Y., Qiao, X., & Li, Y. (2020a). Computer vision technology in agricultural automation —A review. *Information Processing in Agriculture*, ***7***(1), 1–19. doi:10.1016/j.inpa.2019.09.006