Received: August 24, 2023. Revised: November 17, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Parallel Custom Deep Learning Model for Classification of Plant Leaf Disease Using Fusion of Features

Sunil A. Patel[1]* **Pratik A. Barot[2]**

[1]*Computer Engineering Department, L. D. College of Engineering, Ahmedabad, India*
[2]*Computer Engineering Department, Government Engineering College, Gandhinagar, India*
* Corresponding author's Email: sunilpatel.ce@ldce.ac.in

Abstract: Machine learning models have been used to protect the plant from leaf disease by early detection and classification in the field of agriculture. Farmers' principal source of income is agriculture, and plant leaf disease causes them to lose a lot of crop each year. It is essential to provide food for the whole world by detecting disease at an early stage in the farm for future decision-making. However, there is currently no such kind of system in place to forecast a disease at an early stage. The primary goal of this research is to inform farmers about a novel strategy for preventing leaf disease in plants. Through this research and earlier diagnosis, the losses incurred by farmers might be decreased. In this study, we have suggested a unique custom parallel deep convolutional neural network-based model for the classification of leaf diseases on the farm. This proposed CNN model is trained using the plant village dataset with 10 different categories of leaf disease classes. The patterns of leaf images at particular times are utilized in conjunction with computer vision techniques to detect plant leaf diseases. The tomato plant is taken into consideration for current scientific efforts in disease identification, categorization, and diagnosis. Image resizing, cropping, segmentation using k-means clustering and thresholding, and normalization are five different types of data preprocessing techniques used. The suggested approach may accurately diagnose diseases by automatically extracting features using implicit CNN. During comparisons with the testing of the given data, the output of the proposed model is 99.65%, which is better than transfer learning models available in the literature. This suggested model may be checked for consistency and reliability, and farmers can use it as a practical tool to protect their leaf plants from any kind of disease.

Keywords: CNN, Deep learning, Disease detection, Classification, Plant village.

1. Introduction

The country's agricultural landmass has a significant impact on both its economic and social development by producing and providing enough food to the entire population of the world. However, it is not 100% possible to maintain crop, vegetable, and fruit safety and security. According to the United Nations Food and Agriculture Organization, pesticides and diseases pose a threat to global food security and are responsible for a loss of 20–40% of the world's food production [1]. According to the survey [2-3], the entire world has a $220 billion loss due to pest disease and $70 billion loss due to invasive insects annually. Smallholder farmers that rely only on agricultural productivity for their livelihood face difficulties from factors including climate change, the loss of pollinators, plant diseases, and others.

The utilization of analysis and detection techniques for fast and accurate plant disease identification employing current technology aids farmers in solving such issues, which is fundamentally important for agriculture's growth [4]. In our research, we have employed tomato plants to precisely and quickly identify early disease, which enables early prevention. Therefore, early disease identification and detection are crucial for selecting the best course of action and preventing the spread of the disease. These methods' primary goal is to identify diseases in their earliest stages so that

Received: August 24, 2023. Revised: November 17, 2023.

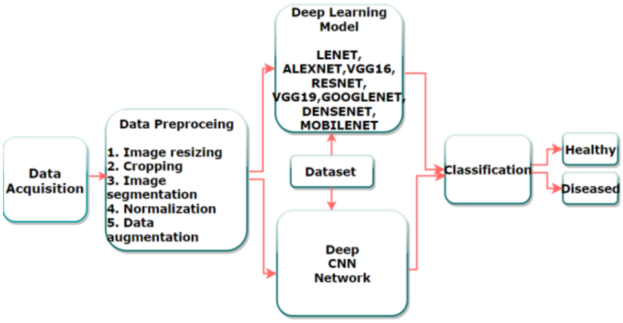


Figure. 1 Workflow of the deep CNN model

appropriate treatment can be taken when it is needed. To combat these types of agricultural diseases, technologies like computer vision and machine learning are used [5]. In this research, we have used a deep learning mechanism to provide an effective treatment for plant diseases. While examining plant leaf diseases research articles, the major problems and challenges are described below.

- The quality of a leaf image must be excellent for proper identification of disease.
- A large volume of images with a publicly available dataset is necessary for improving accuracy.
- Sometimes noise is present in the data, impacting the leaf disease classification.
- Data augmentation is required to increase the number of samples.
- Proper segmentation is required for the identification of disease, and it does not overfit or underfit during training and testing.
- A common model for the identification of all diseases is more challenging because different plants can have a variety of diseases.
- The reduced accuracy found in some scenarios calls for further optimization of the current models.
- More diseases need to be explored due to the limited number of diseases available in the dataset.

Deep learning methods in computer vision have attracted a lot of attention recently, and they have been used in various fields like detection, classification, and machine translation [6]. All the deep learning methods and models mostly fall within the category of supervised learning. This study proposes a deep learning system to automatically identify leaf diseases using a set of labelled training examples while keeping all of the aforementioned factors in mind. CNN's based deep learning algorithm automatically detects diseases from the crop images; therefore, the end user does not need to be an expert in disease recognition. The method described in this study is a novel approach for using a customized parallel deep CNN to detect plant diseases. Fig. 1 depicts the suggested process model for classification and diagnosis.

The remaining sections are arranged as follows: A comprehensive description of current leaf disease approaches is provided in section 2. The proposed technique is presented in section 3. In section 4, the outcomes of the suggested model on samples of tomato leaves are investigated. Section 5 concludes the leaf disease detection model.

2. Related work

We have seen in the past that lots of work has been conducted by various researchers for plant leaf disease recognition using image processing. This research's goal is to demonstrate how several popular machine learning techniques may effectively address these various but closely connected agricultural diseases on leaves. In this section, we have covered various papers on bean, cucumber, wheat, tea, maize, and tomato leaf plants for disease recognition in the field of agriculture. Table 1 Shows summary of literature review on all the disease with deep learning approaches.

The author has proposed a novel technique for two kinds of plant diseases that includes linear contrast stretching using min-max to enhance the quality of the image, segmentation using K-Means to extract regions of interest, GLCM for feature extraction, and a SVM for bean disease classification [7]. In paper [8], multiple descriptors are used for the identification of bean leaf disease using CNN. The author has suggested a semantic segmentation model based on convolutional neural networks for pixel-level segmentation of the powdery mildew on cucumber leaf pictures [9]. In this [10] paper, Only eleven features are analyzed during the classification stage using support vector machines for tea leaf disease identification and reaches the recognition accuracy by more than 90%. In [11], Various statistical operations like mean, median, standard deviation, and mode have been applied to find some meaningful information in the image. Finally, K-means clustering and SVM were utilized to classify the wheat disease. In [12] suggested a research approach on the pipeline of CNN for yield loss in maize leaf plants caused by northern leaf blight (NLB) and achieved 96.7% accuracy.

We have referred many papers on tomato leaf plant disease recognition using conventional and

Received: August 24, 2023. Revised: November 17, 2023.

Table 1. Summary of literature review

Paper ID	Application	Technique	Plant Type	Class	Sampl e Size	Acc.	Drawback / Future work
[5]	Detection, Classification	DWT+PCA+ GLCM+CNN	Tomato	10	600	99.09	Apply the fusion technique for the extraction of features
[7]	Detection, Classification	K-Means+ GLSM+SVM	Bean	2	40	100	Increase the sample size and number of classes in the dataset
[8]	Classification	Multiple descriptors with CNN	Bean	3	1295	99.09	Through a mobile application, send the latest information to the farmer
[9]	Segmentation, Classification	U-Net architecture	Cucumb er	2	50	96.08	Prepare a dataset in an uncontrolled environment and increase the sample size
[10]	Detection, Classification	11 manual Feature + SVM	Tea	3	200	93	Improved segmentation technique and different classifiers
[11]	Detection	K-Means + SVM	Wheat	4	90	-	Apply this method to other leaf plants for the detection of disease
[12]	Classification	Three Stage CNN	Maize	2	1796	96.7	Reduce the number of parameters, Apply this method to other plants
[13]	Classification	VGG16	Tomato	10	18835	95.5	Design custom CNN model for performance improvement
[14]	Classification	Hybrid CNN	Tomato	10	18160	99.17	Design a categorization model for further improvement
[15]	Disease Prediction	Extreme Learning Machine	Tomato	-	-	89.19	Increase sample size by GAN and balance dataset
[16]	Detection	OR-AC-GAN Network	Tomato	2	60	96.25	Integrate other information like leaf temperature and chlorophyll content
[17]	Classification	CNN	Tomato	10	18160	95.65	Extraction of infected parts from a complex background
[18]	Classification, Visualization	CNN	Tomato	9	14828	99.18	Reduce the learning parameter and design a small deep CNN model
[19]	Recognition	CNN with Fast Enhance Learning	Tomato	10	10000	99.40	To add more diseases and train the model

deep learning methods, but few of them are described here. The features are extracted using contour tracing, DWT, PCA, and GLCM in the tomato leaf sample and classified using CNN, SVM, and the K-Nearest Neighbour algorithm. This algorithm achieved 99% in CNN, 97% in KNN, and 88% in SVM accuracy for tomato disorder samples [5]. In [13], a pretrained VGG16 model was used for unhealthy tomato leaf disease identification with a transfer learning approach, and the model reached 99% accuracy for early detection of a disease. This proposed approach uses robust hybrid CNN and INCEPTION modules for tomato leaf disease prediction with 99.17% accuracy [14]. In [15], the author has prepared a real-time powdery mildew disease dataset for tomato plant disease recognition using an extreme machine learning algorithm. This model obtains 89.19% recognition accuracy and 88.57% AUC. The research suggests a novel hyperspectral analysis-proximal sensing approach based on generative adversarial networks. It achieved 96.25% accuracy for proper disease classification on images [16]. In [17], deep learning architecture is used to detect diseases that occur on tomato plants in 10 different classes in the greenhouse. In [18], author proposed a CNN-based algorithm for disease identification using a large volume of tomato leaf datasets. The results were positive, with an accuracy rate of 99.18%. In [19], CNN is used for tomato disorder recognition and achieves 99.40% accuracy. In our study, we trained and tested a custom parallel deep CNN model for identifying diseases using a set of images of plant leaves. By referring to all the papers with their

Received: August 24, 2023. Revised: November 17, 2023.

limitations and future work, we have identified that the feature fusion technique with proper segmentation will boost the performance at some level. We have overcome this limitation by applying k-means clustering and thresholding to color and grayscale images with parallel custom CNN feature fusion. We will also discuss how a proper segmentation approach increases the accuracy of disease classification. The next part provides an overview of the models that have been used as well as training and testing on the datasets.

3. Proposed deep learning method

We go into further depth about the entire designing, training, and validation processes for the parallel deep CNN-based architecture for identifying leaf diseases in tomato plants. In the next subsections, the complete research approach is divided into a number of stages, beginning with dataset details, preprocessing steps, the proposed parallel deep CNN model, classification, and training.

3.1 Dataset details

In our proposed research work, we have used the plant village dataset with healthy and diseased leaf images [20]. It has a total of 54305 images in 38 different categories on 13 different plant leaves. But We have used tomato leaf plants for our research work, which includes 10 various categories of classes with a total of 18835 samples without augmentation and 81315 samples with augmentation. For data augmentation, we have applied rotation, scaling, flipping, translation, and brightness adjustment to increase the images. Table 2 includes all the classes along with their original and augmented images. Fig. 2 shows sample diseased and healthy images of each class of tomato leaf plants.

3.2 Pre-processing step

The performance of a model depends not only on the architecture but also on the quality of the data given for processing. Thus, the image preprocessing step is required for data preparation and leads to better accuracy. Preprocessing enables us to get rid of unwanted distortions and enhance specific features that are crucial for the application. The consistency of the model is also increased by proper feature extraction.

We have used four preprocessing steps before sending the image to the CNN model for training. Firstly, CNN requires the same size of data for model input and subsequent layers; therefore, it resized all the dataset images. Second, each image is divided into a 123 x 123-pixel square that is equal in size and is referred to as the top left, top right, center, bottom left, and bottom right. Fig. 3 shows a sample cropping operation on an image.

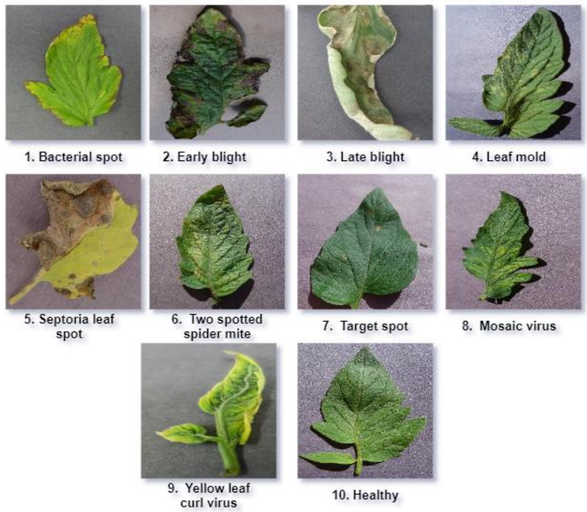


Figure. 2 Sample images of all the diseases

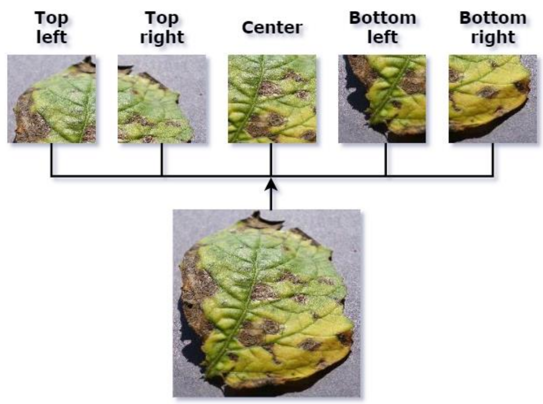


Figure. 3 Sample leaf cropping image

By cropping, we can easily find those regions that are less visible and infected at the corner or center. Third, the proposed model extracts feature parallelly from color and grayscale images. Therefore, we have used k-means and thresholding operations for the segmentation of an image. K-means helps to find objects of interest present in the leaf image to find out the affected part by disease [21]. Fig. 4 displays samples of leaves after applying the segmentation algorithm. Image thresholding separates the foreground and background of an image, making it easier to spot elements that aren't easily apparent in the picture.

Fourth, we have applied range scaling as a part of data normalization [22] for faster convergence and stability of the model. The normalization transforms all the image values to a similar scale in

 DOI: 10.22266/ijies2024.0430.05

Received: August 24, 2023. Revised: November 17, 2023.

Table 2. Summary of the dataset

Sr. No	Class Name	Total Sample	Training Sample	After Data Augmentation	Validating Sample	Testing Sample
1	Bacterial spot	2127	1501	10635	1060	414
2	Early blight	1000	700	3500	500	200
3	Late blight	1909	1339	6695	950	380
4	Leaf mold	1000	700	3500	500	200
5	Septoria leaf spot	1771	1240	6200	885	354
6	Two spotted spider mite	1676	1175	5875	835	334
7	Target spot	1404	984	4920	700	280
8	Mosaic virus	1000	700	3500	500	200
9	Yellow leaf curl virus	5357	3752	18760	1675	1070
10	Healthy	1591	1114	5570	795	318
Total		18835	13205	69155	8400	3760

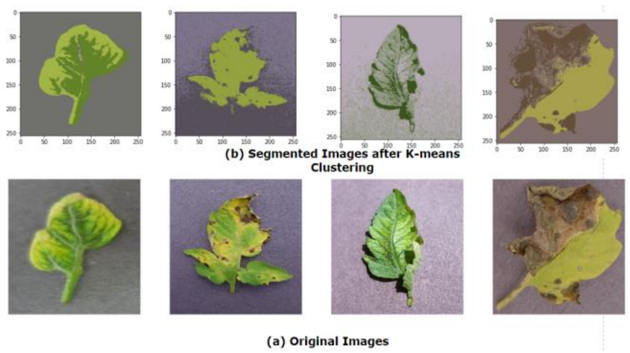


Figure. 4 Sample leaf of the original and k-means clustered image

the range of 0 to 1. At last, the final target dataset is prepared after applying all the aforementioned actions to each image while maintaining proper image quality and without losing the content.

3.3 Parallel custom deep CNN architecture

All the deep CNN models have been used in various agricultural and other image classification applications, as per the literature survey. In our work, we have implemented a custom parallel deep CNN model with the fewest possible layers and minimum learning parameters without compromising the performance of the model. Fig. 5 represents the proposed custom parallel deep CNN architecture for the leaf disease image recognition task. The proposed feed-forward network consists of a number of convolutions, pooling, and a fully connected layer for image classification. The primary benefit of utilizing CNN is the convolution layer's ability to extract features directly from images without the need for manual extraction. During the learning phase, different numbers of kernels are used for the adjustment of the weight at the time of training. We have used the activation function to add the nonlinearity in the image to read the complex information present in the leaf. The network becomes large when the number of feature maps increases, creating a huge volume of parameters and raising computation costs. Therefore, we have reduced the size of an image by max-pooling, which also decreases computation costs and network learning parameters. Dropout and regularization, which remove some nodes from the network automatically to minimize overfitting, are also crucial. We have used a 50% dropout rate during training. Regularization is used for faster convergence of the network. The last, fully connected (FC) layer is used to flatten all the values received from the preceding layer. It is a single vector for classifying the image into 10 different classes depending on the prediction made by a model after training. The neural network is built using all of these features, and the output layer's SoftMax activation function is applied to classify images based on the calculated probability value.

The deep CNN model passes through a number of epochs and internally trains numerous models to find the best one. During training, we have utilized gradient descent optimization in batch for the adjustment of the best weight, which improves the efficiency of the model and, conversely, reduces the cost. This proposed model uses different kinds of grayscale and color images for feature extraction through individual channels.

For merging the appropriate features, we have applied max and average fusion techniques. This merging of features strongly represents the proper information present in the image. In our work, we have used a 123 x 123 x 3-pixel leaf image for input to the network. The 32, 64, 128, and 256 convolution filters with sizes of 3 X 3 are used for learning the information present in the image. For reducing the dimension of the image and learnable parameters, we have used a pooling layer with a size

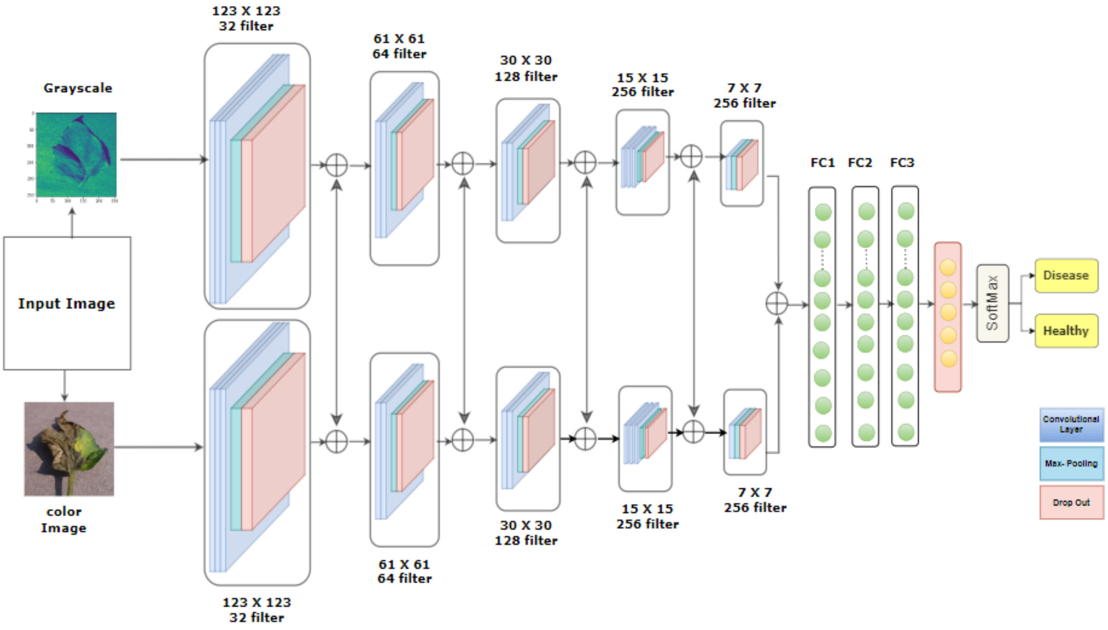


Figure. 5 Proposed parallel custom deep CNN model for leaf disease recognition

of 2 X 2. It reduces the image size from 123 X 123 to 61 X 61. To increase the model's capacity for learning and accuracy, we have designed our own model by hyperparameter adjusting and reconfiguring the layers through various experiments.

3.4 Classification and training

The CNN model is composed of a series of convolution, pooling, fully connected, and flattening layers for processing the set of images. In CNN, For the multiclass classification problem, the SoftMax function receives the value obtained by the flatten layer and classifies it. The main objective is to predict the output sample's diseased or healthy class by using the generated probability value of this function. In our experimentation, we have a total of 10 different classes given by K = 1, 2,... 10. The formula for the SoftMax function is described below.

$$\text{SoftMax}(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (1)$$

In the above Eq. (1), K represents the set of classes, and z represents the set of values generated by the flatten layer and fed to the SoftMax for classification. The final sum of the generated probability values is equal to 1. We have used the argmax (SoftMax) function to obtain the highest probability score that belongs to the predicted class for a given input sample x. We have used batch gradient descent with min-batch to calculate the loss function for updating the weight of the model. We have also applied regularization techniques such as weight decay, drop-out ratio, call-back, learning rate, and early stopping criteria for increasing accuracy.

Table 3. Hardware and software specifications

Hardware/Software	Size/Number
Processor	Intel Xenon E5-2680 v4 dual socket with 14 cores
RAM	64 GB ECC (8 GB X 8) DDR4 – 2400 MHz RAM
HDD	8 TB SATA/NL-SAS disks
OS	Ubuntu
Graphics Card	NVIDIA Quadro 5000 (16 GB)
Language	Python, OpenCV, NumPy, Keras, TensorFlow etc..

3.5 Experimental setup

All the deep learning models have more training parameters; therefore, they require a high-end GPU-based machine to do the task efficiently. We have used the Param Savak DL GPU system (Supercomputer) to carry out this task. The software and hardware specifications are described in Table 3.

3.6 Evaluation parameter

The evaluation parameter plays a major role in describing the quality of the model. We have calculated precision, recall, F1-measure, and Accuracy using the below formula for this suggested model.

$$\text{Accuracy } (\%) = \frac{\text{TP+TN}}{\text{TP + FP+TN+FN}} \quad (2)$$

Received: August 24, 2023. Revised: November 17, 2023.

$$\text{Precision (\%)} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall (\%)} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{F1-Measure (\%)} = \frac{2*\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

4. Results and discussion

For result analysis, we have used the plant village leaf dataset for testing this suggested parallel custom deep CNN model. We have divided the entire dataset into three different categories: training at 70%, validation at 10%, and testing at 20%. To evaluate the efficiency of the classification network, accuracy, precision, recall, and f1-score metrics are calculated. Moreover, we have also compared the result with the existing DL algorithm using the transfer learning approach. At last, we have shown the result in the form of a confusion matrix, performance with respect to category of class, compared to different categories of dataset, and comparisons with existing methods.

4.1 Performance comparison on deep learning models

The most popular approach for reusing a trained model on a new problem is called transfer learning. It is mostly used in machine learning tasks where a huge amount of computational complexity is required to find a proper solution. By comparing the results, we can see from Fig. 6 that our proposed model achieved 99.65% accuracy for the leaf disease identification problem, which is superior to existing deep learning models. Finally, it is observed that our suggested parallel custom deep CNN model outperforms when comparing the results with existing deep learning models.

4.2 Class wise performance on various metrics

We have calculated accuracy, recall, precision, and f1-score for all classes as performance indicators to check the efficiency of the model. The performance metrics are depicted category-wise in Fig. 7 for illustration purposes. As a result, we have achieved more than 98% accuracy for every class, making the model more robust. This model shows 100% accuracy for mosaic virus and spider mites. It is seen that most of the class describes higher precision, which indicates the proportion of positive identification in each sample. It is obvious that this proposed model is highly suitable for real-time application for leaf disease recognition.

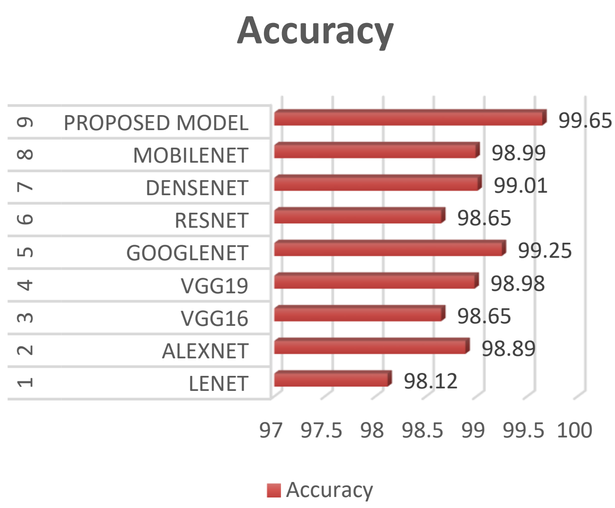


Figure. 6 Performance comparison with existing deep learning models

4.3 Confusion matrix for proposed model

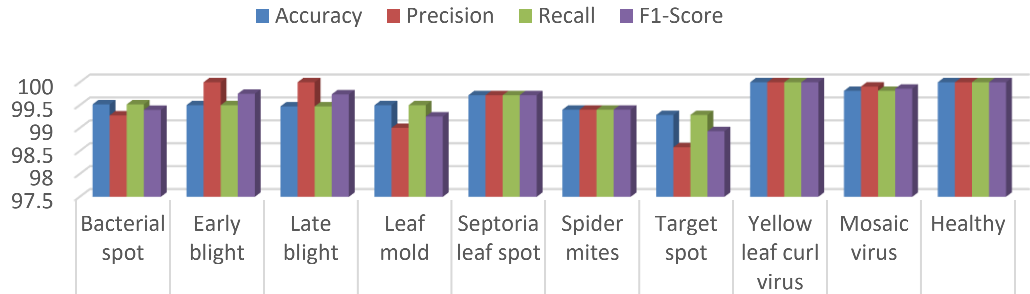
The confusion matrix is N X N in size, and it is used to evaluate the performance measurement for the multiclass classification problem, where N is the number of classes. In this matrix, the diagonal entry represents the accuracy of the relevant class, and the corresponding row represents the falsely classified value. From Fig. 8, It is observed that almost all the classes have performance greater than 98%, with mosaic and healthy classes recognized with 100% performance. In the case of bacterial spot classification, the model falsely classified yellow leaf curl and septoria leaf spot due to their close similarity in nature. Some other conflicts in the prediction of diseases are described in the confusion matrix as affecting the performance of a model. The suggested model is evaluated using a total of 3760 samples from the tomato village dataset, and as a result, it offers an accuracy rating of 99.65%.

4.4 Performance comparison with different dataset categories

As per the literature survey described in section 2, most of the models are tested on the openly accessible plant village color database without modification of the original images.

We have tested our model on various types of datasets, like 1. color, 2. grayscale, and 3. color + grayscale. We have also compared the performance of all the categories by calculating precision, recall, and F1-score, as shown in Table 4. The result indicates a significant improvement in all the performance matrices in the color and grayscale images. The improvement is from 97.83% to

Received: August 24, 2023. Revised: November 17, 2023.



	Bacterial spot	Early blight	Late blight	Leaf mold	Septoria leaf spot	Spider mites	Target spot	Yellow leaf curl virus	Mosaic virus	Healthy
Accuracy	99.5169	99.5	99.4737	99.5	99.7175	99.4012	99.2857	100	99.8131	100
Precision	99.2771	100	100	99.005	99.7175	99.4012	98.5816	100	99.9065	100
Recall	99.5169	99.5	99.4737	99.5	99.7175	99.4012	99.2857	100	99.8131	100
F1-Score	99.3969	99.7494	99.7361	99.2519	99.7175	99.4012	98.9324	100	99.8597	100

Figure. 7 Performance comparison with different categories of classes

Training Set											
TARGET / OUTPUT	Bacterial spot	Early blight	Late blight	Leaf mold	Septoria leaf spot	Two spotted spider mite	Target spot	Mosaic virus	Yellow leaf curl virus	Healthy	SUM
Bacterial spot	412 10.99%	0 0.00%	0 0.00%	0 0.00%	1 0.03%	0 0.00%	0 0.00%	0 0.00%	1 0.03%	0 0.00%	414 99.52% 0.48%
Early blight	0 0.00%	199 5.31%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	1 0.03%	0 0.00%	0 0.00%	0 0.00%	200 99.50% 0.50%
Late blight	0 0.00%	0 0.00%	378 10.08%	0 0.00%	0 0.00%	2 0.05%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	380 99.47% 0.53%
Leaf mold	0 0.00%	0 0.00%	0 0.00%	199 5.31%	0 0.00%	0 0.00%	1 0.03%	0 0.00%	0 0.00%	0 0.00%	200 99.50% 0.50%
Septoria leaf spo	1 0.03%	0 0.00%	0 0.00%	0 0.00%	353 9.41%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	354 99.72% 0.28%
spotted spider	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	332 8.85%	2 0.05%	0 0.00%	0 0.00%	0 0.00%	334 99.40% 0.60%
Target spot	0 0.00%	0 0.00%	0 0.00%	2 0.05%	0 0.00%	0 0.00%	278 7.41%	0 0.00%	0 0.00%	0 0.00%	280 99.29% 0.71%
Mosaic virus	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	200 5.33%	0 0.00%	0 0.00%	200 100.00% 0.00%
llow leaf curl vir	2 0.05%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	1068 28.48%	0 0.00%	1070 99.81% 0.19%
Healthy	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	318 8.48%	318 100.00% 0.00%
SUM	415 99.28% 0.72%	199 100.00% 0.00%	378 100.00% 0.00%	201 99.00% 1.00%	354 99.72% 0.28%	334 99.40% 0.60%	282 98.58% 1.42%	200 100.00% 0.00%	1069 99.91% 0.09%	318 100.00% 0.00%	3737 / 3750 99.65% 0.35%

Figure. 8 Confusion matrix for the proposed model

99.16% to 99.83% in precision, 97.67% to 99.16% to 99.68% in recall, 97.75% to 99.16% to 99.75% in f1-score, and 97.3% to 98.99% to 99.65% in accuracy in the color and grayscale images. The main reason for improvement is that the parallel model interchangeably extracts color and grayscale features after proper segmentation from the images without any bias. Moreover, this suggested model shows the best performance for all the performance indicators as compared to the existing model available in the literature for different categories of images in color and grayscale. We took into account all three kinds of images, and as can be seen from the Table 4, the suggested model performs better than all three for different categories of images in color and grayscale.

Received: August 24, 2023. Revised: November 17, 2023.

Table 4. Performance comparison with different dataset categories

Model	Grayscale				Color				Color + Grayscale			
	Precision	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Accuracy**
LeNet	96.00	95.84	95.92	95.14	97.67	97.66	97.66	97.20	98.33	98.50	98.41	98.12
AlexNet	96.67	96.67	96.67	95.99	98.33	98.32	98.32	97.99	99.17	99.00	99.08	98.89
VGG16	97.17	97.82	97.49	96.97	97.51	97.61	97.56	97.10	98.33	99.33	98.83	98.65
VGG19	97.50	97.60	97.55	97.10	98.50	98.66	98.58	98.30	98.16	98.14	98.15	98.98
GoogleNet	96.50	96.34	96.42	95.78	98.00	98.33	98.16	97.80	99.41	99.25	99.33	99.25
ResNet	96.50	97.47	96.98	96.31	97.55	97.58	97.56	97.15	99.00	99.01	99.00	98.65
DenseNet	96.66	96.66	96.66	95.99	98.01	98.67	98.34	98.31	99.17	99.17	99.17	99.01
MobileNet	96.67	97.48	97.07	96.54	98.17	98.33	98.25	97.89	99.18	99.05	99.11	98.99
Proposed Model	97.83	97.67	97.75	**97.30**	99.16	99.16	99.16	**98.99**	99.83	99.68	99.75	**99.65**

Table 5. Performance comparison with existing methods

Paper ID	Dataset	Precision	Recall	F1-score	Accuracy
[10]	Tea	93.20%	92.8%	92.99%	93%
[15]	Cucumber	90.70%	89.80%	90.29%	89.19%
[19]	Tomato	99.50%	99.50%	98.80%	99.40%
[5]	Tomato	99.50%	99.50%	98.08%	99.09%
[13]	Tomato	95.35%	95.10%	95.22%	95.50%
[14]	Tomato	99.13%	99.23%	99.17%	99.17%
[12]	Maize	-	-	-	96.70%
Proposed Method	**Tomato**	**99.60%**	**99.62%**	**99.60%**	**99.65%**

4.5 Comparison with existing methods

The suggested parallel custom deep CNN classification method is tested with existing methodologies such as statistical univariate features and SVM [10], K-mean clustering, GLCM, and SVM [15], enhance fast learning [19], DWT, PCA, GLCM, and CNN [5], transfer learning with VGG16 [13], hybrid CNN and inception module [14] and three stage CNN [12] as described in Table 5. All the existing models were trained and tested on various kinds of leaf datasets, like tea, maize, cucumber, wheat, and tomato, as described in section 2. We have also compared the performance of the existing model with different performance parameters like precision, recall, f1-score, and accuracy with the suggested model. As per Table 5, the suggested model outperforms in all the metrics. Rather than comparing on a single dataset, we also compared the performance on various datasets. On different kinds of datasets, this proposed model shows significant improvement by extracting blended parallel features from the leaf on grayscale and color images after several convolution-pooling steps. It is achieved due to proper segmentation using k-means clustering on color images and thresholding operations on grayscale images. It identifies the infected part from the images during feature extraction from the leaf. This model can accurately distinguish between healthy leaves and other leaves with a success rate of up to 99.65%; therefore, it works well on unseen samples to identify diseased and healthy leaf plants. Finally, we can conclude that the suggested approach is better than currently available models in the literature by comparing the aforementioned results in different categories.

5. Conclusion

There are numerous deep learning and handcrafted feature extraction-based models available in the literature for leaf disease classification. Still, this subject required more exploration and improvement. Therefore, we have tried to overcome some challenges found in the existing classification method. In this proposed model, we use computer vision techniques with convolutional neural networks for leaf disease prediction. Resizing, cropping, k-means clustering, thresholding, and normalization are applied to extract informative features from the leaf image. We have achieved the best solution by applying multiple experiments to this research methodology on color and grayscale images. Our proposed model achieves 99.65% accuracy, which is remarkable given the recent methods available in the research.

In future work, this suggested method can be applied to real-time application design with the use of drone technology. The drone will collect real-time data from the farm through an inserted in-built camera, and it will be checked by an application that is running on a mobile device or computer. A fully automatic, most efficient system can be created for decision-making, and it will help the farmer take further action. Furthermore, this research work used only 10 diseases for recognition in the leaf plant. By contacting the farmer, we can add other diseases that are not included in this research. And last, this research should also work toward creating an application that allows the current solution to be applied to any kind of leaf plant disease, like fruits and vegetables.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Both the authors' have equally contributed in this research work in Conceptualization, methodology, validation, resources, writing original draft preparation, writing review and editing.

References

[1] L. Ngugi, M. Abelwahab, and M. Zahhad, “Recent advances in image processing techniques for automated leaf pest and disease recognition – a review”, *Information Processeing in Agriculture*, Vol. 8, No. 1, pp. 27-51, 2011.

[2] Food and agriculture organization of united nations, The state of food and agriculture, 2019.

[3] Food and agriculture organization of united nations, 2019, https://www.fao.org/home/en.2019.

[4] A. Chaudhury and J. Barron, “Plant species identification from occluded leaf images”, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 17, No. 3, pp. 1042-1055, 2018.

[5] S. Harakannanavar, J. Rudagi, V. Puranikmath, A. Siddiqua, and R. Pramodhini, “Plant leaf disease detection using computer vision and machine learning algorithms”, *Global Transitions Proceedings*, Vol. 3, No. 1, pp. 305-310, 2022.

[6] S. Patel and R. Makwana, “Connectionist Temporal Classification Model for Dynamic Hand Gesture Recognition using RGB and Optical flow Data”, *The International Arab Journal of Information Technology (IAJIT)*, Vol. 17, No. 4, pp. 65-74, 2022.

[7] S. Abed and A. Esmaeel, “A novel approach to classify and detect bean diseases based on image processing”, In: *Proc. of IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Penang, Malaysia, pp. 297-302, 2018.

[8] V. Kanabur, S. Harakannanavar, V. Puranikmath, P. Hullol, and D. Torse, “Detection of Leaf Disease Using Hybrid Feature Extraction Techniques and CNN Classifier”, In: *Proc. of International Conference on Computational Vision and Bio Inspired Computing*, Coimbatore, India, pp. 1213-1220, 2019.

[9] K. Lin, L. Gong, Y. Huang, C. Liu, and J. Pan, “Deep learning-based segmentation and quantification of cucumber Powdery Mildew using convolutional neural network”, *Fronter in Plant Science*, Vol. 10, pp. 1-10, 2019.

[10] S. Hossain, R. Mou, M. Hasan, S. Chakraborty, and M. Razzak, “Recognition and detection of tea leaf's diseases using support vector machine”, In: *Proc. of IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*, Penang, Malaysia, pp. 150-154, 2018.

[11] S. Nema and A. Dixit, “Wheat Leaf Detection and Prevention Using Support Vector Machine”, In: *Proc. of International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)*, Kottayam, India, pp. 1-5, 2018.

[12] C. D. Chant, T. W. Hanks, S. Chen, E. Stewart, J. Yosinski, M. Gore, and R. Nelson, “Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning”, *Phytopathology*, Vol. 107, No. 11, pp. 1426-1432, 2017.

[13] S. Habiba and M. Islam, “Tomato Plant Diseases Classification Using Deep Learning Based Classifier from Leaves Images”, In: *Proc.*

of International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, pp. 82-86, 2021.

[14] M. Sanida, T. Sanida, A. Sideris, M. Dasygenis, "An Efficient Hybrid CNN Classification Model for Tomato Crop Disease", *Technologies*, Vol. 11, No. 1, pp. 1-10, 2023.

[15] A. Bhatia, A. Chug, and A. Singh, "Application of extreme learning machine in plant disease prediction for highly imbalanced dataset", *Journal of Statistics and Management Systems*, Vol. 23, No. 6, pp. 1059-1068, 2020.

[16] D. Wang, R. Vinson, M. Holmes, G. Seibel, A. Bechar, S. Nof, and Y. Tao, "Early detection of tomato spotted wilt virus by hyperspectral imaging and outlier removal auxiliary classifier generative adversarial nets (OR-AC-GAN)", *Scientific Reposit*, Vol. 9, pp. 1-14, 2019.

[17] H. Durmuş, E. Gunes, and M. Kirci, "Disease detection on the leaves of the tomato plants by using deep learning", In: *Proc. of International Conference on Agro-Geoinformatics*, Fairfax, VA, USA, pp. 1-5, 2017.

[18] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep Learning for Tomato Diseases: Classification and Symptoms Visualization", *Applied Artificial Intelligence*, Vol. 31, No. 4, pp. 299-315, 2017.

[19] T. Vadivel and R. Suguna, "Automatic recognition of tomato leaf disease using fast enhanced learning with image processing", *Acta Agriculture Scandinavica*, Vol. 72, No. 1, pp. 1-13, 2022.

[20] A. Pandian and G. Gopal, https://data.mendeley.com/datasets/tywbtsjrjv/1, 2019.

[21] X. Jin and J. Han, "K-Means Clustering", *Encyclopedia of Machine Learning*, Springer. 2011.

[22] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance", *Applied Soft Computing*, Vol. 97, pp. 1-7, 2020.