International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v10i6.5555
***Article Received**: 10 April 2022 **Revised**: 18 May 2022 **Accepted**: 02 June 2022 **Publication**: 30 June 2022*

Deep Learning and Computer Vision based Model for Detection of Diseased Mango Leaves

Sandhya S
School of Computer Science and Engineering,
Vellore Institute of Technology (VIT),
Chennai, India
e-mail: ssandhyarmk@gmail.com

Balasundaram A
Assistant Professor (Senior Grade), School of Computer Science and Engineering,
Centre for Cyber Physical Systems, Vellore Institute of Technology (VIT),
Chennai, India
Email: balasundaram.a@vit.ac.in

Arunkumar Sivaraman
Assistant Professor (Senior Grade), School of Computer Science and Engineering,
Vellore Institute of Technology (VIT),
Chennai, India
Email: dr.arunsivaraman@gmail.com

Abstract— Mangifera Indica, commonly known as mangoes, is the most commercialized export fruit crop in India, accounting for about 40% of the total global production. Due to its widespread production, it is vulnerable to a variety of diseases that affect its yield and resulting in loss. These diseases like Anthracnose, Powdery Mildew, Leaf blights, etc., occur primarily on leaves. As a result, there is a great need for a system that helps in the detection of diseased mango leaves. In this paper, we propose a system that makes use of pre-trained Convolutional Neural Network architecture, the ResNet-50 for the detection of infected mango leaves. The dataset contains 435 images of mango leaves with binary classification as healthy and diseased. These images are pre-processed by resizing them and applying CLAHE. After applying in-place data augmentation on the dataset, the features are extracted using the ResNet-50 model. For the classification process, we make use of fine-tuned head and Machine Learning classifiers such as Support Vector Machine, Gradient Boosting, Logistic Regression, XGBoost, Decision Tree, and K Nearest Neighbour. Among them, the fine-tuned head classifier achieved an accuracy of 97.7%, and Machine Learning classifiers such as SVM, Logistic Regression obtained an accuracy of 100%. The experimental results obtained validate that the system is efficient in its performance of detecting the two classes of mango leaves.

Keywords- Deep Learning, Computer Vision, Convolutional Neural Network, ResNet-50, Machine Learning Classifiers.

I. INTRODUCTION

Agriculture is the foundation of the global economic system and plays a pivotal role in the overall functioning of a given economy [1]. The global need for food security is growing rapidly as the world's population increases exponentially each day. The primary goal for achieving food security is the cultivation of good crop varieties along with early disease detection. Mangifera Indica, commonly known as Mangoes is the most commercialized export crop in India [2] and ranks first among the world's mango producers. Since this crop is easily adaptable and cultivable to a variety of factors like location, climate, it is also vulnerable to a variety of disease-causing agents. These agents, such as pests and other microorganisms that cause diseases, damage nearly 30-40% of crop yield. These diseases like anthracnose, leaf spots, gall infestation cause bacterial, viral, fungal, or parasitic infections [3]. The presence of disease in leaves disrupts their natural processes eventually leading to their death. Some ailments are also caused by some external factors, resulting in physical damage to the surface of the fruits. This will further reduce its cost and declines its quality. Figure 1 shows the list of diseases that affects the mango trees.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v10i6.5555
***Article Received**: 10 April 2022 **Revised**: 18 May 2022 **Accepted**: 02 June 2022 **Publication**: 30 June 2022*

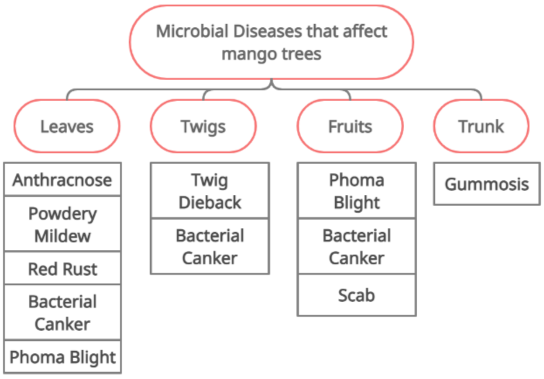

Figure 1. Microbial Diseases that affect different parts in the mango trees.

The quality and quantity losses caused by the plant diseases have had a negative impact on the crop production cost, along with the revenues of the agriculture stakeholders. Traditional disease detection methodologies used by experts and farmers are expensive, laborious, and highly infeasible in scenarios where the disease is advanced. This necessitates the development of an early disease detection system [4] to avoid further devastating losses. Advances in computer vision [5] have paved the way for early disease diagnosis using cutting-edge "Deep Learning" or "Machine Learning" algorithms. Because of significant advancements in computer vision models, there has been an exponential rise in the development of automated systems for detecting plant diseases [6] based on observable symptoms. These would result in fully automatic systems which would need no human intervention for the disease diagnosis. Despite the availability of many methodologies for detecting diseased plants, the need for a generalized system is high. Figure 2 shows the various purposes where machine learning is integrated in agriculture paving way for a new domain called smart agriculture.

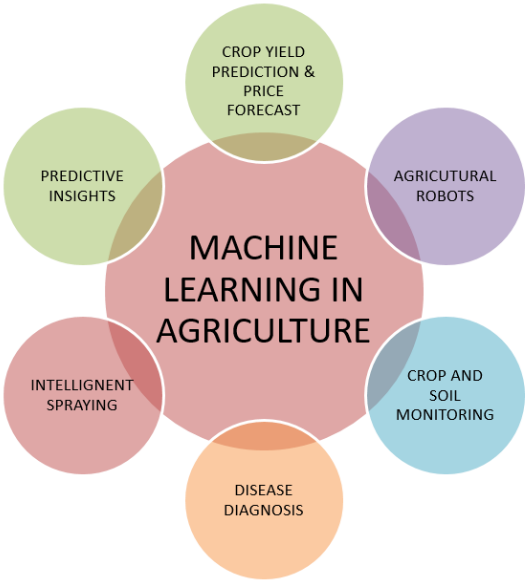

Figure 2. Machine Learning in Agriculture

In this research paper, we propose a transfer learning methodology that would help in detecting diseased mango leaves from healthy ones using pre-trained Resnet-50 CNN model. The dataset consists of labelled images of both infected and healthy leaves, and on which contrast enhancement is performed for better clarity, followed by pre-processing techniques like image resizing. In-place data augmentation is performed to combat the issue of a small dataset. The images are then fed into a pre-trained residual neural network followed by extraction of features. Classification is performed on the mango leaves dataset using a fine-tuned model and various machine learning (ML) algorithms, and good results are achieved.

II. Related Works

Md. Rasel Mia et al. [7] presents an ensemble neural network system for disease recognition in mango leaves, followed by classification using a Support Vector Machine classifier, where the model is trained and evaluated using the color and shape features of the Mango leaves dataset. This system was able to classify five types of mango leaf images accordingly, with the initial classification as disease-free or not. Then if the leaf image is found to be infected with diseases like Golmachi, Moricha, Dag, and Shutimold, the process is repeated using SVM. Due to the limitation of overfitting for higher-dimensional problems, the usage of Kernel Support Vector Machine is replaced with the classic SVM for classification as it works the best for finding an effective linear separator with maximum margin.

Uday Pratap Singh et al. [8] proposed a methodology called "Multilayer Convolutional Neural Network" for classifying the leaves of mango that is affected by the disease anthracnose. They make use of a real-time mango leaves dataset which consists of diseased and healthy mango leaves. The total of 2200 images are then preprocessed by rescaling, followed by contrast enhancement for obtaining a uniform contrast enhancement on the images. These images after split into 80:20 train test ratios are fed to the customized CNN model which initially classifies whether it is a mango leaf or not. Then it is further classified into diseased or healthy leaves. The proposed method achieved good performance with an accuracy of 97.13%, whereas Radial Basis Function, SVM, and Particle Swarm Optimization algorithms achieved an accuracy of 94.2%, 92.75%, and 88.39% respectively.

In this research paper, S Arivazhagan et al. [9] proposed a novel convolutional neural network model to classify six classes of mango leaves as healthy and five diseased leaf classification. The proposed methodology inspired by pre-trained CNN models like VGG-16, AlexNet performs feature extraction, and by selecting the features with the highest probability values. Among them, classification of leaf gall, leaf

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v10i6.5555
***Article Received**: 10 April 2022 **Revised**: 18 May 2022 **Accepted**: 02 June 2022 **Publication**: 30 June 2022*

burn, leaf Webber and Alternaria leaf spot diseases achieved 100% accuracy. Whereas for the classes healthy, and anthracnose, it achieved an accuracy of 86% and 94% respectively. The overall average testing accuracy of the obtained results is 96.67%.

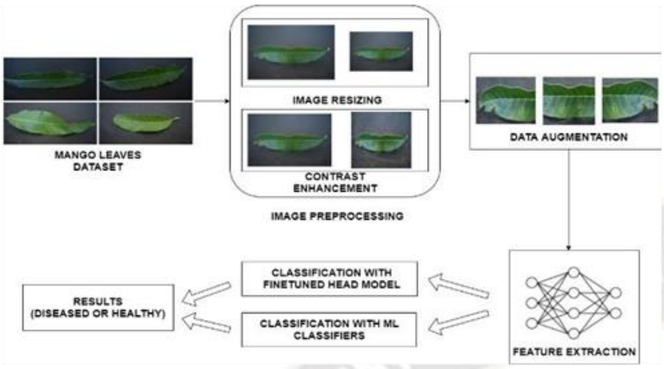

Figure 3. Machine Learning in Agriculture Distribution of healthy and diseased mango leaves in the dataset.

The methodology proposed by Anjna et al. [10] identifies the various fungal, viral, bacterial diseases that occur in capsicum using a hybrid system. By using k means clustering, the disease-infected areas on the leaves of capsicum are extracted. After extraction of GLCM features, classification is performed by classifiers such as Decision Tree, K Nearest Neighbor (KNN), Support Vector Machine (SVM). Among these, SVM and KNN gave the highest accuracy of 100%, followed by 98.4 % for both Linear Discriminant and Decision Tree.

By employing transfer learning, Kien Trang et al. [11] proposed a deep residual neural network approach for detection of diseases in mango leaves using Plant Village Dataset. Various image preprocessing techniques like contrast enhancement, resizing is performed on the image dataset. The model after being trained on the Plant Village Dataset, is then trained later on the Mango leaves dataset by using transfer learning. Results show that the presented system achieved good performance with an accuracy of 88.76% when compared with pre-trained CNN models such as AlexNet, Mobilenetv2, and Inceptionv3.

K Srunitha et al. [12] presented a methodology for the identification of disease-infected regions using k means clustering and classification by multiclass SVM. Color-based segmentation is performed to obtain RGB segments of the input image and is followed by feature extraction. Features like Entropy, Contrast, Smoothness, Correlation, and other metrics which belong to Gray Level Occurrence matrix [13] are extracted and are fed to the classification. algorithms. Among the two classifiers, Multiclass Support Vector Machine achieved higher accuracy of 96% when compared to ANN.

In the research paper, Bed Prakash et al. [14] presented a system for segmenting regions of interest on mango leaves using k means clustering, and classification using BPNN methodology. After obtaining an optimal k value for segmentation, features such as orientation, solidity, area, extent are calculated and are then extracted. The system is tested with various combinations of k values combined with classification using BPNN and performed the best with five clusters achieving an accuracy of 94%.

Mustafa Merchant et al. [15] proposed a methodology to identify the nutrient deficiencies in the mango leaves based on the RGB intensities of the mango leaf image. It works on the idea that the mango leaves that are deficient in nutrients are of a different color when compared to the healthy leaves. The RGB values and texture data of the leaves are extracted, and then clustering is performed to cluster the leaves suffering from various deficiencies like copper, nitrogen, potassium, and iron.

Bhagyashri et al. [16] presented a system for detection of diseases affecting the rice leaves using color and textural features [17]. The key features are selected by identifying the infected from the images of rice leaves using the Gray Level Occurrence matrix. Additionally, color and area features are also extracted, followed by a selection of important features [18] by using a genetic algorithm. The classification is performed using SVM and an ANN classifier, among which GLCM features with SVM classification gave the highest accuracy of 92.5%.

These methodologies make use of self-captured mango leaves dataset for training the model. Hence, the performance of each of the systems varies and a common conclusion cannot be obtained from them. This results in a need for a generalized system that is trained on a publicly available dataset that can be accessed by everyone. PlantVillage dataset [19] is a generalized dataset that is used in many methodologies for testing the performance of a plant disease prediction system. Due to the absence of mango leaf images in this dataset, there is no generalized dataset which can be used for testing the mango leaves disease prediction system. Our proposed methodology makes use of a public dataset of mango leaves for training the model. Feature extraction by incorporating transfer learning would reduce the training time significantly. Furthermore, it would contribute to developing an efficient system with a good fit.

III. PROPOSED METHODOLOGY

The images of the mango leaves are pre-processed by resizing them into a smaller and equal dimension as all the images are of different sizes. It is followed by contrast enhancement of the leaf images so that the main areas of the leaves are enhanced for better results. Data augmentation is incorporated for combating the issues of a smaller number of images for training, and testing combined. The dataset is then split into an 80:20 ratio, resulting in 80% of the total images allocated to the training set, and the remaining 20% of the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v10i6.5555
***Article Received**: 10 April 2022 **Revised**: 18 May 2022 **Accepted**: 02 June 2022 **Publication**: 30 June 2022*

images are allocated to the testing set. This allocation is performed randomly and also stratified to obtain images of equal classes ratio in both training and the testing set. Feature extraction is performed using pre-trained ResNet-50. This transfer learning methodology will help to retain the weights of the pre-trained ResNet-50 model, which is already trained on the ImageNet dataset. In addition, this will reduce the training time of the model achieving a good fit model.

A. Dataset Description

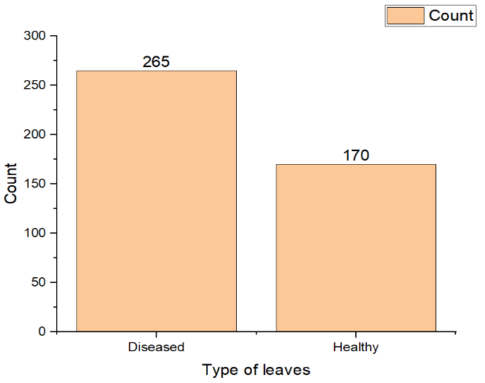

Figure 4. Distribution of healthy and Diseased Mango leaves in the dataset.

The dataset is from Data Mendeley [20] that contains images of leaves belonging to different plants like Pomegranate, Lemon, Mango, Guava. Among all these varieties of leaves, we have considered the images of mango leaves for our proposed system. The dataset contains around 435 images of mango leaf out of which 265 are diseased mango leaves, and the remaining 170 images are healthy mango leaves. Figure 4 shows the distribution of the two classes of mango leaves. The images of the mango leaves that are considered for our methodology are placed on a dark blue background. The contrasting color of the mango leaves appears distinct from the dark background, which aids in the easy segregation of the leaf. The diseased leaf images consist of blobs or tiny spots of disease on its surface, whereas the healthy leaf image has a uniform surface with no deformities. Figure 5(A) shows a sample of healthy mango leaves, and 5(B) depicts the diseased mango leaves.

(A) Healthy Mango leaves

(B) Diseased Mango leaves
Figure 5. Sample images from the mango leaves dataset.

B. Image Preprocessing

The images of the mango leaf from the dataset are available in different sizes. Therefore, before feeding them into the system, the images undergo preprocessing such as resizing all the images to convert them into the same dimension. Resizing is done to reduce the dimension of the image to the target size of 224 × 224. The image is further preprocessed by performing contrast enhancement. There are numerous contrast enhancement methodologies [21] that can be used for increasing the intensities of the image pixels.

Figure 6 Contrast enhancement on images using CLAHE

Generally, contrast enhancement results in an ununiformed increase in contrast resulting in a useless image. Adaptive Histogram Equalization still faces some drawbacks as it is capable of over amplifying the noise in certain regions of the image. By employing the Contrast Limited Adaptive Histogram Equalization method, the intensities are amplified uniformly resulting in an image that could prove useful for our system. By comparing both the images (A) and (B) in figure 6, we can see that the image on which CLAHE is applied is sharper and has a better resolution of the areas infected by the disease. This improvement helps in the easy segregation of the leaf from the background. Additionally, the venation pattern on the leaves is visible for easy identification of areas infected by the disease. As the dataset consists of a total of only 435 images, data augmentation is performed. Data augmentation also aids in preventing overfitting during the training process. It is performed by augmenting the datasets with images that go through transformation techniques like image rotation, intensity variation, affine transformation on the original images. We have made use of in-place data augmentation which works in such a way that augmented images sent to the model are previously never seen by the model.

C. Methodology

Our proposed methodology makes use of ResNet-50 Convolutional Neural Network architecture [22] consisting of

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v10i6.5555
***Article Received**: 10 April 2022 **Revised**: 18 May 2022 **Accepted**: 02 June 2022 **Publication**: 30 June 2022*

50 layers in which 48 are convolution layers followed by one MaxPool and Average Pool layers. The typical architecture of ResNet-50 is shown in Figure 7. The layers in ResNet are represented as serially linked residual blocks that construct skip connections which ignores the initial network layers and helps in resolving the vanishing gradient problems that occur in deep networks that are usually hard to train. ReLU activation functions are used in the intermediate layers which aid in faster convergence and are proved to improve the model's efficiency during training. By using a pre-trained ResNet-50 model, the training process is initiated [23] with the weights of the ImageNet dataset, followed by training on the mango leaves dataset with a decaying learning rate, employing Binary Cross-Entropy loss and Adam optimizer for 20 epochs. After performing the necessary image preprocessing techniques, feature extraction is performed to extract key features in the form of feature vectors from the mango leaf image.

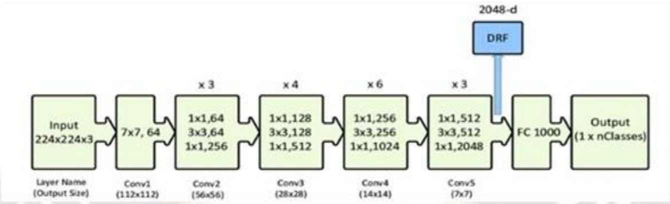

Figure 7 ResNet-50 architecture [22]

D. Classifiers

The final step is the classification of images which is performed using a fine-tuned model and various machine learning classifiers to categorize them into healthy or diseased leaves. Figure 6 shows the taxonomy of various classifiers. When the common behavioral pattern is unknown, training-based approaches are preferred for detecting anomalies. Training-based approaches involve the usage of some set of data to train the system to understand the common behavioral pattern and thereby classify any abnormal activities as anomalies. Standard classifier-based approaches such as Random Forest, SVM, and other classification mechanisms are used to classify anomalies. When the training data is unbalanced, an ensemble of classifiers are deployed to balance the training data. When the data is auto-correlated, time-series-based approaches or Recurrent Neural Network-based approaches are used. Figure 8 describes the classification approaches that are available for different types of problem statements. However, the training data may not be available at all times. In such cases, anomaly detection can be accomplished using semi-supervised or unsupervised learning. It may be applying some point-based anomaly approaches such as percentiles and histograms or applying some collective anomaly approaches. If the data is univariate in nature, Markov chain-based approach or any model-based approach can be deployed to detect anomalies. When the data is multivariate and ordered, a combination of clustering and Markov chain-based approaches can be used. If the data is multivariate and un-ordered, any of the clustering-based approaches or K-nearest neighbor-based approaches can be used.

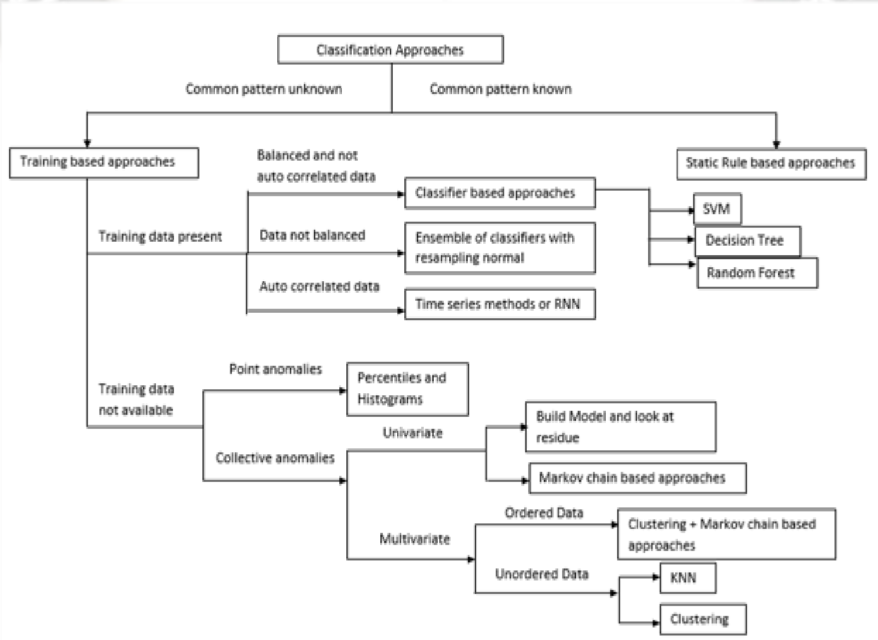

Figure 8 Taxonomy of different classifiers

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v10i6.5555
***Article Received**: 10 April 2022 **Revised**: 18 May 2022 **Accepted**: 02 June 2022 **Publication**: 30 June 2022*

IV. Experimental Results

A. *Performance of the model*

Table 1 lists the training loss and validation loss values obtained at each epoch during the training of the model. Initially, the training and validation loss is high, and as the training of the model continues, the loss values decrease steadily in the upcoming epochs. This can be observed in figure 9, which depicts the plot of training loss and validation loss for each epoch number during the training of the model where the red line denotes the training loss, and the blue line represents the validation loss. Moreover, the training and validation loss plots are overlapping which indicates a point of stability.

Table 2 lists the training accuracy and the validation accuracy values at each epoch. As the number of epochs increases, the accuracy of the model is increasing steadily. Figure 10 depicts the training accuracy and validation accuracy plot where the red line denotes the training accuracy, and the blue line represents the validation accuracy.

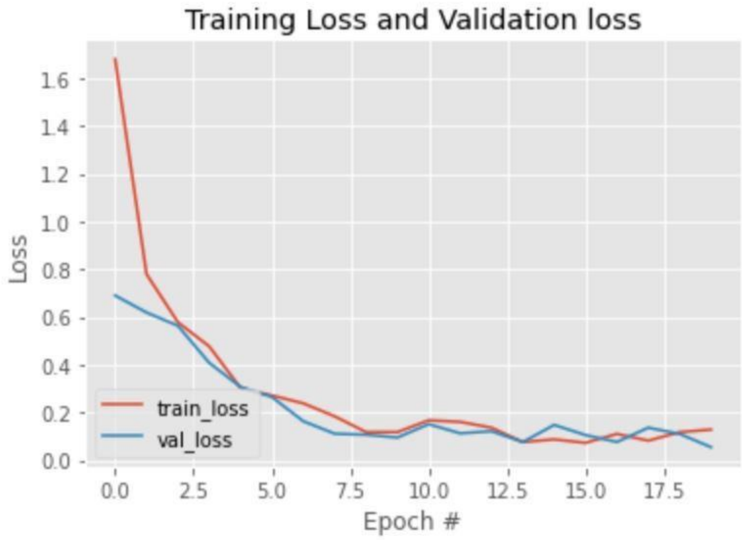

Figure 9 Training and Validation loss plot

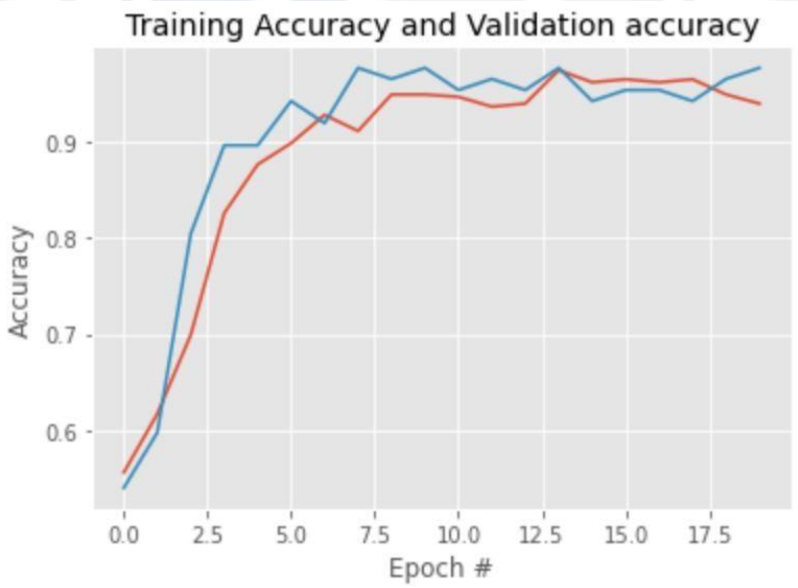

Figure 10 Training and Validation Accuracy plot

From table 2 and figure 10, we can infer that the training and validation accuracy is gradually increasing, indicating a good fit of the model. The model has achieved an accuracy of 97.7% for 20 epochs. From the plots, we can deduce further that the proposed model is not overfitted or underfitted. This is achieved by employing a decaying learning rate for training the model. Moreover, the usage of Adam optimizer and the addition of dropout layer in the finetuned-head layer has further contributed to achieving a good fit model.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v10i6.5555
***Article Received**: 10 April 2022 **Revised**: 18 May 2022 **Accepted**: 02 June 2022 **Publication**: 30 June 2022*

TABLE I. TRAINING AND VALIDATION LOSS AT EACH EPOCH

Epoch Number	Training Loss	Validation Loss
1	0.9391	0.6218
2	0.5983	0.5832
3	0.4892	0.4064
4	0.3101	0.3841
5	0.2952	0.2897
6	0.2463	0.1782
7	0.1913	0.1299
8	0.1204	0.1189
9	0.1627	0.1231
10	0.1873	0.1416
11	0.1785	0.1258
12	0.1547	0.1473
13	0.0933	0.1061
14	0.0952	0.1606
15	0.0903	0.1174
16	0.1271	0.0925
17	0.1002	0.1324
18	0.1127	0.1178
19	0.1216	0.0589
20	0.1219	0.0506

TABLE II. TRAINING AND VALIDATION ACCURACY AT EACH EPOCH

Epoch Number	Training Accuracy	Validation Accuracy
1	0.6093	0.5102
2	0.7002	0.8133
3	0.8325	0.8998
4	0.8854	0.9002
5	0.9001	0.9470
6	0.9306	0.9273
7	0.9213	0.9796
8	0.9599	0.9699
9	0.9565	0.9708
10	0.9502	0.9624
11	0.9464	0.9655
12	0.9488	0.9585
13	0.9702	0.9702
14	0.9695	0.9499
15	0.9703	0.9579
16	0.9635	0.9604
17	0.9610	0.9490
18	0.9521	0.9701
19	0.9486	0.9768
20	0.9398	0.9770

B. Performance of the classifiers

The output of the classifiers predicts whether the leaf is diseased or a healthy one. To evaluate the performance of the model with respect to each machine learning classifier, The following evaluation metrics such as the Precision, Recall, F1-Score values and Accuracy is calculated using equations (1), (2), (3), (4) respectively.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

$$Accuracy = \frac{TP + TN}{(TP + FP) + (TN + FN)} \quad (4)$$

where, TP, TN, FP, and FN, stands for True Positive, True Negative, False Positive, and False Negative, respectively.

Classification report and confusion matrix best describe the performance of the classifier model that is created and whose true values are known. The confusion matrix and the classification report denoting the precision, recall, F1-score and support pertaining to the various classifiers used is shown in table 3 to table 9 respectively.

TABLE III. CLASSIFICATION REPORT OF KNN ALGORITHM

	PRECISION	RECALL	F1-SCORE	SUPPORT
Diseased	0.84	0.89	0.86	53
Healthy	0.81	0.74	0.77	34
Accuracy			0.83	87
Macro Average	1.00	1.00	1.00	87
Weighted Average	1.00	1.00	1.00	87
Confusion Matrix	[[53 0] [0 34]]			

TABLE IV. CLASSIFICATION REPORT OF GAUSSIAN NAÏVE BAYES ALGORITHM

	PRECISION	RECALL	F1-SCORE	SUPPORT
Diseased	0.83	0.98	0.90	53
Healthy	0.96	0.68	0.79	34
Accuracy			0.86	87
Macro Average	0.89	0.83	0.84	87
Weighted Average	0.88	0.86	0.86	87
Confusion Matrix	[[52 1] [11 23]]			

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v10i6.5555
***Article Received**: 10 April 2022 **Revised**: 18 May 2022 **Accepted**: 02 June 2022 **Publication**: 30 June 2022*

TABLE V. CLASSIFICATION REPORT OF LOGISTIC REGRESSION ALGORITHM

	PRECISION	RECALL	F1-SCORE	SUPPORT
Diseased	1.00	1.00	1.00	53
Healthy	1.00	1.00	1.00	34
Accuracy			1.00	87
Macro Average	1.00	1.00	1.00	87
Weighted Average	1.00	1.00	1.00	87
Confusion Matrix	[[53 0] [0 34]]			

TABLE VI. CLASSIFICATION REPORT OF DECISION TREE ALGORITHM

	PRECISION	RECALL	F1-SCORE	SUPPORT
Diseased	0.91	0.98	0.95	53
Healthy	0.97	0.85	0.91	34
Accuracy			0.93	87
Macro Average	0.94	0.92	0.93	87
Weighted Average	0.93	0.93	0.93	87
Confusion Matrix	[[52 1] [5 29]]			

TABLE VII. CLASSIFICATION REPORT OF SVM ALGORITHM

	PRECISION	RECALL	F1-SCORE	SUPPORT
Diseased	1.00	1.00	1.00	53
Healthy	1.00	1.00	1.00	34
Accuracy			1.00	87
Macro Average	1.00	1.00	1.00	87
Weighted Average	1.00	1.00	1.00	87
Confusion Matrix	[[53 0] [0 34]]			

TABLE VIII. CLASSIFICATION REPORT OF GRADIENT BOOST ALGORITHM

	PRECISION	RECALL	F1-SCORE	SUPPORT
Diseased	0.90	1.00	0.95	53
Healthy	1.00	0.82	0.90	34
Accuracy			0.93	87
Macro Average	0.95	0.91	0.92	87
Weighted Average	0.94	0.93	0.93	87
Confusion Matrix	[[53 0] [6 28]]			

TABLE IX. CLASSIFICATION REPORT OF XGBOOST ALGORITHM

	PRECISION	RECALL	F1-SCORE	SUPPORT
Diseased	0.95	1.00	0.97	53
Healthy	1.00	0.91	0.95	34
Accuracy			0.97	87
Macro Average	0.97	0.96	0.96	87
Weighted Average	0.97	0.97	0.97	87
Confusion Matrix	[[53 0] [3 31]]			

Table 10 compares the performance metrics of the classification algorithms. From the results, we can conclude that Support Vector Machine (SVM), and Logistic Regression (LR) algorithms achieved the highest performance with an accuracy rate of 100%. The fine-tuned model was able to achieve an accuracy of 97.7%, followed by XG Boost achieving an accuracy of 96.55 %. Decision Tree (DT) and Gradient Boosting achieved 93.1% accuracy, with Gaussian Naïve Bayes and K Nearest Neighbor (KNN) resulting in an accuracy of 86.21% and 82.76% respectively. Figure 11 compares the accuracy rates of the classification algorithms.

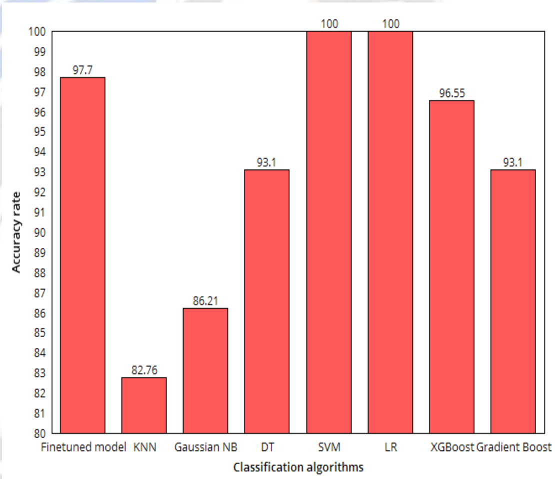

Figure 11 Accuracy of classification algorithms

From table 11, we can conclude that our proposed system has excelled in terms of performance with the other methodologies. Md. Rasel Mia et al. [7] have made use of a neural network ensemble with SVM that achieved an accuracy of 80%, whereas our system when combined with the SVM classifier achieved an accuracy of 100%. MCNN, which is a customized CNN model proposed by Uday Pratap Singh et al. [8] achieved an accuracy of 97.14%, whereas the fine-tuned head model has achieved an accuracy of 97.7%. Feature extraction using GLCM incorporated in [10], [12], [16], but our proposed methodology has superseded in its performance by employing feature extraction using pre-trained ResNet-50. The methodologies presented in [9], [11], [14], [15] are customized CNN models, trained on respective datasets, but our pre-trained model by incorporating transfer learning and trained on a publicly available dataset has achieved the highest performance in terms of accuracy.

TABLE X. PERFORMANCE COMPARISON OF VARIOUS ALGORITHMS

S.NO	ALGORITHM	PRECISION	RECALL	F1-SCORE	ACCURACY
1	Fine-tuned model	0.97	0.97	0.97	97.7%
2	KNN Classifier	0.82	0.81	0.81	82.76%
3	Naive Bayes	0.90	0.83	0.84	86.21%
4	Decision Tree	0.94	0.91	0.93	93.1%
5	Support Vector Machine	1.00	1.00	1.00	100%
6	Logistic Regression	1.00	1.00	1.00	100%
7	XGBoost	0.97	0.95	0.96	96.55%
8	Gradient Boost	0.95	0.91	0.92	93.1%

TABLE XI. PERFORMANCE COMPARISON OF VARIOUS ALGORITHMS

Reference	Proposed Methodology	Performance
Md. Rasel Mia et al. [7]	Neural Network Ensemble with SVM	80%
Uday Pratap Singh et al. [8]	Multilayer Convolutional Neural Network	97.13%
S Arivazhagan et al. [9]	Convolutional Neural Network	96.67%
Anjna et al. [10]	GLCM features with SVM	100%
Kien Trang et al. [11]	Residual Neural Network (RNN)	88.46%
K Srunitha et al. [12]	GLCM features with SVM	96%
Bed Prakash et al. [14]	Back Propagation Neural Network	94%
Mustafa Merchant et al. [15]	K-Means Clustering	-
Bhagyashri et al. [16]	Area, GLCM and Color Moment features with SVM Classifier	92.5%
Proposed Methodology	Pre-trained ResNet-50 with fine-tuned head model, Pre-trained Reset-50 model with classifiers such as SVM, Logistic Regression	Fine tuned head model – 97.7%. SVM and LR classifier – 100%

V. CONCLUSION AND FUTURE WORK

The agriculture domain has always foreseen losses caused by plant diseases and has resulted in disastrous effects on food security. Therefore, the advancement of Computer Vision technologies in the last decade resulted in the easy development of systems that help in the early diagnosis and identification of plant diseases. In this research paper, we have presented a deep learning approach based on a pre-trained ResNet-50 CNN model. Image preprocessing is performed on all images by employing contrast enhancement, and also resizing them accordingly, to feed them to the deep neural network. The features are extracted using Res-Net 50 having ImageNet weights and are sent as inputs to the fine-tuned model and various other ML Classifiers. Furthermore, the use of the pre-trained ResNet50 helps in better feature extraction to obtain the best results for the final classification into two classes. This is extremely preferable as we need not train the model with other datasets, but with the help of the ImageNet weights available in the pre-trained ResNet50 model, training becomes much easier, and faster.

This proposed methodology can be incorporated into a system that has a camera. The camera can be used for capturing the images of the leaves and can be sent to the system for differentiating between healthy and diseased mango leaves. It can be further improved by detecting the type of disease infecting the leaf such as anthracnose, powdery mildew, scab, etc. A recommender system can be created to suggest the right treatment for the respective disease. This automatic system can help the farmers to prevent any further damage caused by diseases.

ACKNOWLEDGMENT

The authors would like to express their profound gratitude to the VIT management for their unwavering support and encouragement.

REFERENCES

[1] Neha Khanna, Praveen Solanki. Role of agriculture in the global economy. https://www.longdom.org/proceedings/role-of-agriculture-in-the-global-economy-15024.html. Accessed 20 March 2021.

[2] MANGO - National Horticulture Board. http://nhb.gov.in/report_files/mango/mango.htm . Accessed 19 March 2021.

[3] Prakash, O., 2004. Diseases and disorders of mango and their management. In *Diseases of Fruits and Vegetables Volume I*. Springer, Dordrecht. (pp. 511-619)

[4] Saleem, M.H., Potgieter, J. and Arif, K.M., 2020. Plant Disease Classification: A Comparative Evaluation of Convolutional Neural Networks and Deep Learning Optimizers. *Plants*, *9*(10), p.1319

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 6
DOI: https://doi.org/10.17762/ijritcc.v10i6.5555
***Article Received**: 10 April 2022 **Revised**: 18 May 2022 **Accepted**: 02 June 2022 **Publication**: 30 June 2022*

[5] Wani, J.A., Sharma, S., Muzamil, M., Ahmed, S., Sharma, S. and Singh, S., 2021. Machine Learning and Deep Learning Based Computational Techniques in Automatic Agricultural Diseases Detection: Methodologies, Applications, and Challenges. *Archives of Computational Methods in Engineering*, pp.1-37.

[6] Balasundaram, A., Pradeep, K.V. and Sandhya, S., 2021. An Extensive study on disease prediction in mango trees using computer vision. *Annals of the Romanian Society for Cell Biology*, pp.1895-1905.Electronic Publication: Digital Object Identifiers (DOIs): Article in a journal:

[7] Vanitha, D. D. . (2022). Comparative Analysis of Power switches MOFET and IGBT Used in Power Applications. International Journal on Recent Technologies in Mechanical and Electrical Engineering, 9(5), 01–09. https://doi.org/10.17762/ijrmee.v9i5.368

[8] Mia, M.R., Roy, S., Das, S.K. and Rahman, M.A., 2020. Mango leaf disease recognition using neural network and support vector machine. *Iran Journal of Computer Science*, *3*(3), pp.185-193.

[9] D Kothandaraman, A Balasundaram, SeenaNaik Korra, E Sudarshan and B Vijaykumar "Enhancing dull images using discrete wavelet families and fuzzy" 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022020

[10] Arivazhagan S. and Ligi, S.V., 2018. Mango leaf diseases identification using convolutional neural network. *Int. J. Pure Appl. Math*, *120*(6), pp.11067-11079.

[11] Gupta, D. J. . (2022). A Study on Various Cloud Computing Technologies, Implementation Process, Categories and Application Use in Organisation. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(1), 09–12. https://doi.org/10.17762/ijfrcsce.v8i1.2064

[12] Sood, M. and Singh, P.K., 2020. Hybrid System for Detection and Classification of Plant Disease Using Qualitative Texture Features Analysis. *Procedia Computer Science*, *167*, pp.1056-1065.

[13] Trang, K., TonThat, L., Thao, N.G.M. and Thi, N.T.T., 2019, November. Mango Diseases Identification by a Deep Residual Network with Contrast Enhancement and Transfer Learning. In *2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET)* (pp. 138-142). IEEE.

[14] Srunitha, K. and Bharathi, D., 2018. Mango leaf unhealthy region detection and classification. In *Computational Vision and Bio Inspired Computing* (pp. 422-436). Springer, Cham.

[15] Sebastian V, B., Unnikrishnan, A. and Balakrishnan, K., 2012. Gray level co-occurrence matrices: generalisation and some new features. *arXiv preprint arXiv:1205.4831*.

[16] Prakash, B. and Yerpude, A., 2015. 'Identification of Mango Leaf Disease and Control Prediction using Image Processing and Neural Network'. *Internationl Journal for Scientific Research & Development*, *3*(5).

[17] Balasundaram, A., Ashok Kumar, S., & Magesh Kumar, S. (2019). Optical flow based object movement tracking. International Journal of Engineering and Advanced Technology, 9(1), 3913–3916.

[18] Ghyar, B.S. and Birajdar, G.K., 2017, November. Computer vision based approach to detect rice leaf diseases using texture and color descriptors. In *2017 International Conference on Inventive Computing and Informatics (ICICI)* (pp. 1074-1078). IEEE.

[19] A. Balasundaram, C. Chellappan, "Vision Based Gesture Recognition: A Comprehensive Study", The IIOAB Journal, Vol.8, Issue.4, pp.20-28, 2017.

[20] Lu, Y., Yi, S., Zeng, N., Liu, Y. and Zhang, Y., 2017. Identification of rice diseases using deep convolutional neural networks. *Neurocomputing*, *267*, pp.378-384.

[21] PlantVillageDataset. https://www.kaggle.com/abdallahalidev/plantvillage-dataset. Accessed 15 March 2021.

[22] CHOUHAN, Siddharth Singh; Kaul, Ajay; SINGH, UDAY PRATAP; & Science, Madhav Institute of Technology (2019), "A Database of Leaf Images: Practice towards Plant Conservation with Plant Pathology", Mendeley Data, V1, doi: 10.17632/hb74ynkjcn.

[23] Maragatham, G. and Roomi, S.M.M., 2015. A review of image contrast enhancement methods and techniques. *Research Journal of Applied Sciences, Engineering and Technology*, *9*(5), pp.309-326.

[24] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[25] V. N. Patil and D. R. Ingle, "A Novel Approach for ABO Blood Group Prediction using Fingerprint through Optimized Convolutional Neural Network", Int J Intell Syst Appl Eng, vol. 10, no. 1, pp. 60–68, Mar. 2022.

[26] Chen, J., Chen, J., Zhang, D., Sun, Y. and Nanehkaran, Y.A., 2020. Using deep transfer learning for image-based plant disease identification. *Computers and Electronics in Agriculture*, *173*, p.105393.