

Ensemble Learning Framework for Mango Plant Disease Detection and Classification

Dr. Meenakshi Thalor¹, Mr. Sanjay Mate², Dr. Ashpana Shiralkar³, Dr. Amita Shinde⁴

^{1,3,4}AISSMS Institute of Information Technology, Pune, India

²Department of Information Technology, Govt. Polytechnic, Daman, India

ARTICLE INFO

Received: 11 Oct 2024

Revised: 13 Dec 2024

Accepted: 23 Dec 2024

ABSTRACT

Agriculture sector play a vital role in economy of India where the crop of mangoes is also considered as major fruit crop as it contributes significantly to the country's agricultural economy. Mango cultivation provides livelihood to millions of farmers across the country. One of the main barriers to increased food production is the diseases of the plants. Mango trees are prone to a variety of diseases and addressing them effectively can be quite challenging. This paper presents an ensemble-based classification of mango tree leaf diseases. Ensemble based classification makes use of multiple classifiers in order to make an efficient decision about the crop disease. In this paper, homogeneous VGG-19 CNN architecture is employed in bagging manner which proves the validity of the system by providing the accuracy of 95%, precision of 97%, recall of 97% and F-score of 97%. This system will be useful for Ministry of agricultural and farmer welfare for taking preventive measures to make Mango trees disease free.

Keywords: Disease Detection, Deep Learning, Ensemble, Mango

INTRODUCTION

India cultivates a wide range of mango varieties, each with its distinct flavor, texture, and characteristics. As per the agricultural statistics, 20 million tons of Mango is only produced by India and significantly contribute 40% of global mango production. Some notable varieties of mangos produce in India are Alphonso (Hapus), Dasheri, Langra, Chausa, Kesar and Banganapalli mangoes etc. well known for its sweet taste, color, size, vitamins and minerals.

Mangoes are highly significant both economically and culturally around the world. Mangoes are a major fruit crop with a significant presence in international trade. Major exporters include countries like India, Mexico, Thailand, and the Philippines. This trade contributes significantly to the economies of these countries. In many developing countries, mango farming is a key source of income and employment. The sale of mangoes, especially premium varieties like Alphonso, plays an important role in their income. India is the largest producer of mangoes in the world, accounting for about 40% of global production.

There are various challenges associate with the crop of mangoes i.e. Mango tree is susceptible to a range of diseases and addressing them effectively can be quite challenging. Mango diseases significantly affects production and overall economic returns. Detecting mango diseases early is crucial to preventing a significant drop in yield and agricultural output. Traditional methods of disease detection and diagnosis are often labor-intensive, time-consuming, and prone to inaccuracies. This can result in considerable crop losses, as diseases can spread rapidly and harm mango trees. Table 1 shows some common mango leaf diseases with their symptoms and images.

Table1: Mango Tree Leaf Diseases

Sr. No.	Mango Leaf Diseases	Symptoms	Images	
1	Anthraxnose	Dark, sunken lesions on leaves, often surrounded by a yellow halo. Affected leaves may drop prematurely.		
2	Bacterial Canker	Small, dark, necrotic spots with yellow halos on leaves. Infected areas may also develop a water-soaked appearance		
3	Cutting Weevil	Weevils may cause holes or irregular notches in the leaves. Infested leaves often show signs of wilting or yellowing.		
4	Die Back	progressive death of twigs, branches, shoots, or roots, starting at the tips.		
5	Gall Midge	Round, blister-shaped galls, Approximately 2–3mm in diameter and 0.4–0.7mm high. Visible on both sides of the leaf.		

To effectively address these challenges and to curtail crop losses, farmers and local experts often rely on visual inspections. However, this manual approach can be impractical due to the time involved, the limited availability of experts, and the potential for human error. Consequently, an automated system for accurately detecting and classifying these diseases is essential for enhancing crop management and reducing production losses.

In this paper, our focus is on identification of six categories of Mango leave— Anthracnose, Bacterial Canker, Cutting Weevil, Die Back, Gall Midge and healthy leaf. The key main objective of this work is to introduce automated system which makes use of ensemble learning approach to train VGG19 model of CNN. In addition, this work validate ensemble based system on several evaluation parameters like accuracy, precision, recall and f measure.

RELATED WORK

In literature, several techniques like image processing, artificial intelligence, machine learning were introduced to combat the crop diseases. The research over crop disease is significantly improve in last one decade because of advancement of technology and AI. Machine learning and deep learning techniques have played an important role in agricultural sectors for the prediction, classification, and detection of plant disease and provide non-destructive, low-cost, fast, and reliable means of plant disease detection. Different researchers have researched the diagnosis and detection of plant disease specifically on mango disease. Mia et al. [1] and Saleem et al. [2] worked on small dataset of 20 and 100 images to classified the mango diseases respectively. Kumar et al. [3] implemented VGG-16 CNN architecture to detect diseases in plant village dataset as well as validated the system for anthracnose disease by self-captured images of mango leaves. Singh et al. [4] build the multilayer Alexnet architecture of CNN and tested the model on plant village and on some mango leaf images infected by Anthracnose disease. Merchant et al. [5] made use of clustering technique on mango leaf images to determine nutrient deficiencies. Too et al. [6] used different classifiers like VGG-16, ResNet, DenseNet and InceptionNet on plant village dataset. Gandhi et al. [7] tried CNN and GAN on Plant Village Dataset. Durmucs et al. [8] introduced Alexnet and Squeezenet whereas Picon et al. [9], Lu et al. [10] and Ma et al. [11] utilize DCNNs. Golhani et al. [12] employ various neural networks for disease detection. Joshi et al. [13] made use of ensemble based learning [14,15,16] for plant disease detection.

From literature, we can conclude that a quality work is carried out on plant village dataset which consist of leafs of potato rather than Mango. Limited research is carried out on Mango Leafs considering dedicated dataset of mangoes as well as considering several diseases associated with it. This paper focus on the multiclass classification of leaf disease detection of Mango trees by using dedicated Mango leaf dataset i.e. MangoLeafBD.

RESEARCH METHODOLOGY

This paper proposes, ensemble learning framework for detection and classification of Mango leaf diseases. Ensemble base classification make use of multiple classifiers rather than rely on the decision of one classifier. These multiple classifiers can be attached in sequence (Boosting approach) or in parallel(Bagging approach). This paper presents ensemble based learning using Bagging (Bootstrap aggregation) where multiple homogenous classifiers will be used together to detect and classify the type of disease in mango leaf. The training data is divided into chunk of data(bag) using random subsampling and provided to classifier for learning purpose. Final classification of instance/ sample is decided based on Majority voting technique.

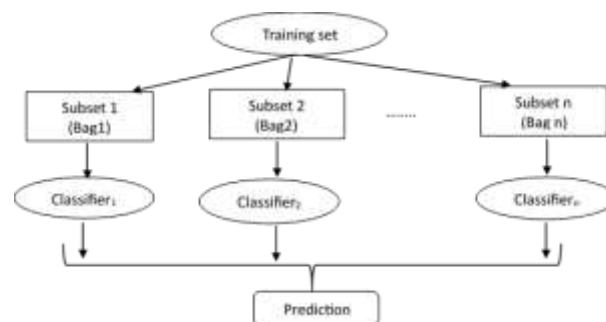


Fig.1: Ensemble based Learning

Fig. 2. Shows the system architecture where initially mango leaf images are downloaded from mangoleaf BD dataset available at Kaggle repository. The MangoLeaf BD dataset consist images of six categories named as Anthracnose(count=500), Bacterial Canker(count=500), Cutting Weevil(count=500), Die Back(count=500), Gall Midge(count=500) and Healthy leaves(count=29). After analysis of distribution of leaf categories, this has been observed that the no. of instances for healthy leaves are significantly low as compare to other categories. The balancing of dataset is carried out by using SMOTE (Synthetic Minority Oversampling Technique) algorithm [17]. The oversampling of healthy leaves is carried out at preprocessing stage so classification results will not be biased.

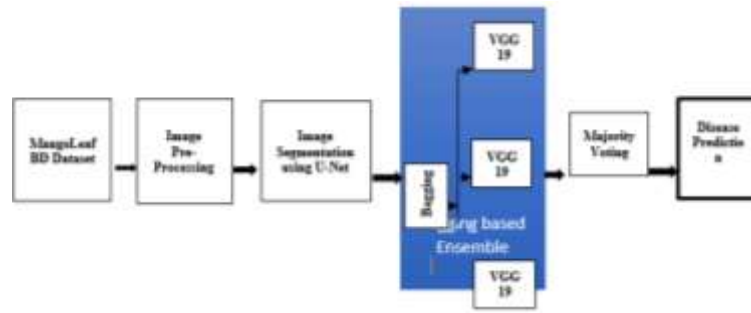


Fig.2: Ensemble Learning Framework for Mango Plant Disease Detection and Classification

Image preprocessing involves intensity normalization and resizing to improve prediction accuracy. During intensity normalization all pixels of leaf image are adjusted in range of $[0, 255]$ and leaf images are resized to a standard dimension of 256×256 pixels ensuring consistent parameter training and enhancing overall model performance for both segmentation and classification tasks.

After image Preprocessing, all the images of MagoLeafBD dataset are segmented using U-Net algorithm which is a deep learning CNN architecture. The U-Net architecture make use of convolutional layer, pooling and softmax in order to segment the image and generate the output image as shown in fig. 3:

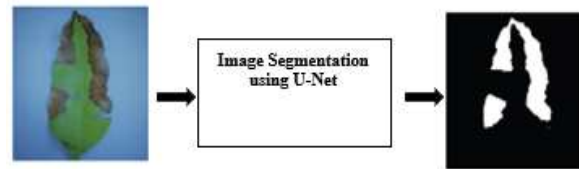


Fig. 3. Image Segmentation

After Image segmentation, to automatically identify diseases on mango leaf images, an ensemble based approach is used which makes use of three homogeneous classifiers. During bagging process, three different bag of dataset will be created using random subsampling and each classifier will be trained on its bag of dataset and each classifier will predict the category of leaf and at the end, based on majority voting the final category of unseen leaf will be provided

After Segmentation, VGG-19 architecture as shown in fig. 4 is used for classification of segmented image where multiclass classification is carried out by trained model.

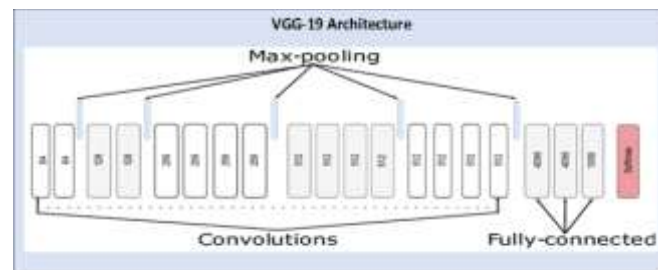


Fig. 4: VGG Architecture

The model is trained over 100 epochs with a minimum batch size of 32. The trained model is evaluated by using accuracy, precision, recall, and F1-score.

EXPERIMENTAL RESULTS

The implementation of this system is conducted in python using Kera library. The mango leaf images are downloaded from mangoleaf BD dataset available at Kaggle repository. The MangoLeaf BD dataset consist images of six categories named as Anthracnose, Bacterial Canker, Cutting Weevil, Die Back, Gall Midge and Healthy leaves. The distribution of Data is shown in Table 1 after applying SMOTE on health leaf images. This system is trained with 80 % of data while remaining 20 % data is used for testing of model. The trained model is evaluated by using accuracy, precision, recall,

and F1-score as shown in Table 2. Table 3 shows the evaluation of trained model by considering 20% samples as testing data.

Table1: Dataset details and Data Distribution

Class	Total Samples	Training Dataset	Test Dataset
Anthraco nose	500	400	100
Bacterial Canker	500	400	100
Cutting Weevil	500	400	100
Die Back	500	400	100
Gall Midge	500	400	100
Healthy(After SMOTE)	200	160	40

Table 2: Evaluation Measures

Measure	Derivations
Recall/Sensitivity	True Positive Rate = True Positive / (True Positive + False Negative)
Precision	PPV = True Positive / (True Positive + False Positive)
Accuracy	ACC = (True Positive + True Negative) / (Positive + Negative)
F1 Score	F1 = 2 True Positive / (2 True Positive + False Positive + False Negative)

Table 3: Evaluation of PotatoLeaf Insight

Class/Catego ry	Testing Dataset Instanc es	True Positiv e	True Negati ve	False Positiv e	False Negati ve	Accura cy	Precisio n	Recal l	F-score
Anthraco nose	100	95	1	3	1	96%	96.94%	98.96 %	97.94 %
Bacterial Canker	100	94	2	2	2	96%	97.92%	97.92 %	97.92 %
Cutting Weevil	100	95	0	3	2	95%	96.94%	97.94 %	97.44 %
Die Back	100	96	0	2	2	96%	97.96%	97.96 %	97.96 %
Gall Midge	100	95	1	1	3	96%	98.96%	96.94 %	97.94 %
Healthy	40	36	0	2	2	90%	94.74%	94.74 %	94.74 %

Experimental results show that the validity of the system as overall accuracy of proposed system is 95%, precision is 97%, recall is 97 and F-score is 97%.

ACKNOWLEDGEMENT

We extend our sincere thanks to everyone who directly or indirectly played a role in developing this application. A special appreciation goes to AISSMS Institute of Information Technology for providing experimental setup environment. This research is inline with the objectives of the Department of Agriculture and Farmers welfare, Ministry of Agriculture, New Delhi, India.

FUNDING SOURCE

As a software project, this research work implemented on open-source resources and contributions from the community. We are grateful to open-source ecosystem that has enabled us to bring this work into fruition.

CONFLICT OF INTEREST

There is no conflict of interest associated with the development of this work. Our team has accurately conducted the research, design, and implementation phases, ensuring the integrity and impartiality of the platform.

AUTHORS CONTRIBUTION

Author 1: Initiated the research work by outlining, mentioning objectives and prepared the system architecture.

Author 2: Contributed to customized data collection using camera and in documentation of paper.

Author 3: Implementation of system starting from data collection to classification task.

Author 4: Performed the validation of model by using different evaluation measures like precision, recall and f score.

DATA AVAILABILITY

The mango leaf images are downloaded from mangoleaf BD dataset available at Kaggle repository. The MangoLeaf BD dataset consist images of six categories named as Anthracnose, Bacterial Canker, Cutting Weevil, Die Back, Gall Midge and healthy leaves

INFORMED CONSENT STATEMENT

As part of our commitment to ethical research practices, all participants involved in data collection for this project provided informed consent.

CONCLUSION

This paper presents an ensemble-based approach to classify the Mango leaf diseases. In this work, U-Net architecture of CNN is employed for segmentation of mango leaf disease followed by ensemble based classification. Bagging based ensemble is implemented with homogeneous VGG-19 classifiers and finally multiclass classification is done using majority voting approach. Experimental results show that the validity of the system as overall accuracy of proposed system is 95%, precision is 97%, recall is 97 and F-score is 97%. This system is a prominent tool for agricultural departments and farmers for safeguarding their yield by taking necessary measures. In future, recommendation techniques can be integrated with this system so that what measures to be taken against a particular disease can be provided to farmers.

REFERENCES

- [1] Mia M., Roy S., Das S., and Rahman M. Mango leaf disease recognition using neural network and support vector machine. *Iran Journal of Computer Science*, 2020;3, 09.
- [2] Saleem R., Shah J.M., Sharif M., Yasmin M., Yong H.S., and Cha J. Mango leaf disease recognition and classification using novel segmentation and vein pattern technique. *Applied Sciences*. 2021;11(24):11901.
- [3] Kumar P., Ashtekar S., Jayakrishna S., Bharath K., Vanathi P., and Kumar R.. Classification of mango leaves infected by fungal disease anthracnose using deep learning. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021;1723–1729.
- [4] Singh Uday, Chouhan S., Jain S., and Jain S.. Multilayer convolution neural network for the classification of mango leaves infected by anthracnose disease. *IEEE Access*, 2019;7:43721–43729.
- [5] Merchant M., Paradkar V., Khanna M., and Gokhale S.. Mango leaf deficiency detection using digital image processing and machine learning. In 2018 3rd International Conference for Convergence in Technology (I2CT), pages ,2018,1–3.

-
- [6] Edna Chebet Too, Li Yujian, Sam Njuki, and Liu Yingchun. A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*, 161:272–279, 2019.
 - [7] Gandhi Rutu, Nimbalkar Shubham, Yelamanchili Nandita, and Ponkshe Surabhi. Plant disease detection using cnns and gans as an augmentative approach. In 2018 IEEE International Conference on Innovative Research and Development (ICIRD), pages 1–5. IEEE, 2018.
 - [8] Halil Durmuş, Ece Olcay Güneş, and Mürvet Kırıcı. Disease detection on the leaves of the tomato plants by using deep learning. In 2017 6th International Conference on Agro-Geoinformatics, 2017; 1–5.
 - [9] Artzai Picon, Aitor Alvarez-Gila, Maximilian Seitz, Amaia Ortiz-Barredo, Jone Echazarra, and Alexander Johannes. Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Computers and Electronics in Agriculture*, 2019;161:280–290.
 - [10] Lu Yang, Yi Shujuan, Zeng Nianyin, Liu Yurong, and Zhang Yong. Identification of rice diseases using deep convolutional neural networks. *Neurocomputing*, 2017;267:378–384
 - [11] Juncheng Ma, Keming Du, Feixiang Zheng, Lingxian Zhang, Zhihong Gong, and Zhongfu Sun. A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. *Computers and electronics in agriculture*, 2018;154:18–24.
 - [12] Golhani Kamleshv Balasundram Siva K, Vadamalai Ganesan, and Pradhan Biswajeet. A review of neural networks in plant disease detection using hyperspectral data. *Information Processing in Agriculture*, 2018;5(3):354–371.
 - [13] Joshi P., Dev A., Sharma A., and Jangra R., Plantbalance: An automated ensemble learning framework for plant disease detection, in 2022 IEEE Delhi Section Conference (DELCON). IEEE, 2022, pp. 1–6.
 - [14] R. Atallah and A. Al-Mousa, Heart disease detection using machine learning majority voting ensemble method, in 2019 2nd international conference on new trends in computing sciences (ICTCS). IEEE, 2019, pp. 1–6.
 - [15] Suk H.I., Lee S.W., Shen D., t al., Deep ensemble learning of sparse regression models for brain disease diagnosis, *Medical image analysis*, 2017;37:101–113.
 - [16] Muller D., Soto-Rey I., and Kramer F., “An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks,” *IEEE Access*, 2022;10:467–88.
 - [17] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 2002,16, 321–357.