Machine Learning-Based Mango Leaf Disease Detection: A Comparative Study of Random Forest, CNN, and Hybrid Approaches

Dr. Santosh Kumar Singh [1], Hrushi Bhola [2], Kunal Singh [3]

[1] H.O.D(IT), Department of Information Technology, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India

[2, 3] PG Student, Department of Information Technology, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India

Abstract: Mangoes hold nutritional, economic, and cultural value, offering diverse culinary options and environmental benefits. Monitoring leaf health aids in early disease detection and orchard management, minimizing crop loss and chemical usage. Using CNN, Random Forest, and Random Forest with CNN feature extraction, we achieved 98.12% accuracy in distinguishing healthy and unhealthy mango leaves. This method promises improved orchard management, resource optimization, and reduced environmental impact. Maintaining leaf health enhances fruit yield, quality, and economic viability, aligning with sustainable agriculture goals. Early disease detection facilitates timely intervention, reducing reliance on chemicals and preserving ecosystem balance. By revolutionizing orchard management, your research contributes to the long-term sustainability of mango cultivation and broader environmental stewardship efforts.

Keywords: CNN, Disease detection, Mango leaf, Random Forest.

I. Introduction

Mango farming is an essential part of agricultural economies around the world, giving farmers a sizable source of revenue and promoting global food security. However, the productivity and health of orchards are seriously threatened by the mango trees' vulnerability to several diseases. To minimize crop losses and conduct appropriate remedies, early detection of these diseases is critical. Recent developments in machine learning techniques have demonstrated the potential to transform agricultural practices, especially disease identification. In this research, we specifically emphasize utilizing the capabilities of Random Forest, Convolutional Neural Networks (CNN), and a hybrid strategy that incorporates both methodologies for the use of machine learning algorithms for the identification of illnesses in mango leaves.

The main goal of this research is to provide a trustworthy and effective method for dividing mango leaves into categories that are good and unhealthy. To attain this objective, strong algorithms that can precisely detect minute patterns and differentiate between various illness stages are needed. With its ensemble learning methodology, Random Forest provides a potent tool for classification tasks, whereas CNN is an excellent choice for image-based illness diagnosis since it can extract intricate characteristics from image data. The proposed hybrid model builds on the advantages of these separate techniques by utilizing CNN for feature extraction and Random Forest to take advantage of the complementing qualities of the two techniques.

II. Related work

We conduct this research to address the pressing need for early detection and management of mango leaf diseases, which significantly impact crop yield and quality. By leveraging machine learning techniques, we aim to provide farmers with efficient tools for disease identification, facilitating timely intervention and minimizing crop losses. Ultimately, this research contributes to sustainable agricultural practices, supporting food security and environmental stewardship efforts.

Saleem et al. [1] proposed Automated mango leaf disease recognition using novel segmentation and vein pattern. Achieved 95.5% accuracy with SVM for disease identification. Rajbongshi et al. [2] in their paper focuses on mango leaf disease recognition using CNN models. Utilizes DenseNet201, InceptionResNetV2, ResNet50, Xception, and more models. DenseNet201 achieved the highest accuracy of 98.00% among models. Veling et al. [3] focus on Mango Disease Detection using Image Processing techniques. Features like Contrast, Correlation, Energy, Entropy, and Homogeneity were utilized. Proposed system achieves 90% accuracy. Mia et al. [4] proposed neural network ensemble for mango leaf disease recognition with 80% accuracy. Arivazhagan et al. [5] proposed Deep learning-based approach for automating leaf disease identification. CNN model achieves 96.67% accuracy in identifying mango leaf diseases. Gining et al. [6] developed a mango leaf disease recognition system using image processing. System detects and classifies disease with 68.89% accuracy. Saleem et al. [7] developed FrCNnet model for mango leaf disease segmentation achieving 99.2% segmentation accuracy. Nikam et al. [8] proposed image processing measures mango leaf disease severity accurately utilizing Laplacian filter techniques for disease region segmentation.

In this research, machine learning algorithms are employed, including Convolutional Neural Networks (CNN) and Random Forest, to classify mango leaf images into healthy and unhealthy categories. By leveraging these algorithms, we aim to develop robust models capable of accurately identifying signs of disease in mango leaves. Additionally, we explore the effectiveness of a hybrid approach that combines CNN feature extraction with Random Forest classification. The aim of this research is to enhance disease detection methods in mango cultivation, ultimately improving orchard management and crop sustainability.

III. Methodology

1. Data Collection

In adapting the MangoLeafBD Dataset [9], significant modifications were made to better align with the research objectives. Notably, the dataset's original structure underwent reorganization, consolidating eight subdirectories into two distinct categories: 'healthy' and 'unhealthy' mango leaf samples. This restructuring aimed to streamline the dataset's organization, facilitating a more focused analysis of the binary classification of leaf health status. By condensing the subdirectories, usability and clarity were enhanced, enabling a more efficient exploration of the relationship between leaf characteristics and health status. These modifications were implemented with careful consideration of the research objectives, essential for tailoring the dataset to effectively address specific research questions.

2. Data Preprocessing

Image data preprocessing was performed using Keras ImageDataGenerator. Pixel values were normalized to a range of 0 to 1 using rescaling. This normalization step was achieved by setting the rescale parameter to 1/255 in the ImageDataGenerator, ensuring consistency in feature scales across the dataset. While data augmentation options were available, they were not utilized. These preprocessing steps ensure uniformity in feature scales and aid in model convergence during training.

3. Exploratory Data Analysis (EDA)

In the exploratory data analysis (EDA) of the mango image dataset, 3200 training samples and 800 validation samples were examined across 2 classes: 'Unhealthy' and 'Healthy' mango leaf conditions. The class distribution plot indicates 'Unhealthy' as the most prevalent class and 'Healthy' as the least represented. Sample images showcase a variety of mango leaf conditions, including signs of disease or damage. All images in the dataset are consistently sized at 224x224 pixels with 3 color channels (RGB), ensuring uniformity. This initial analysis establishes a foundation for further study, offering insights into the dataset's composition and characteristics.

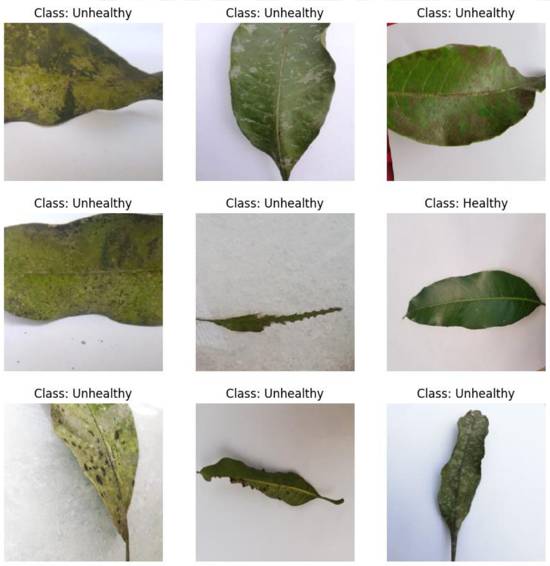

Figure 1. sample of dataset.

4. Models

A. Convolutional Neural Network (CNN) Model

The CNN model presented here is tailored for the classification of mango leaf images into two categories: healthy and unhealthy. It features a multi-layer architecture designed to process 224x224 RGB images efficiently. The model incorporates convolutional layers, followed by max-pooling layers, to extract and down sample features from the input images. These layers utilize rectified linear unit (ReLU) activation functions to introduce non-linearity into the model. Following the convolutional layers, a flatten layer reshapes the output into a one-dimensional vector, which is then fed into a dense layer comprising 128 neurons with ReLU activation. Finally, the output layer consists of a single neuron with sigmoid activation, facilitating binary classification. During training, the model is optimized using the Adam optimizer and trained with binary cross-entropy loss. Accuracy serves as the primary metric for evaluating the model's performance. The training process includes monitoring both training and validation accuracies across epochs. The model achieves its highest validation accuracy of 91.87% at epoch 10, indicating its effectiveness in distinguishing between healthy and unhealthy mango leaf images.

B. Random Forest (RF) Model

The Random Forest model is adept at categorizing mango leaf images into two classes: healthy and unhealthy. After preparing the data by loading and flattening the images, the model is constructed with 100 decision trees and a random state of 42. Following this, the model is trained using the provided training data. Upon completion of training, predictions are made on the validation set to evaluate the model's performance. The model achieves an accuracy rate of 93.75%, indicating its effectiveness in classifying mango leaf images. This assessment provides valuable insights into the model's capabilities, ensuring its suitability for distinguishing between healthy and unhealthy mango leaf images.

C. Random Forest with CNN feature extraction (RF_CNN) Model

The "Random Forest with CNN feature extraction" model is tailored for the classification of mango leaf images into healthy and unhealthy categories. Leveraging the VGG16 convolutional neural network (CNN) pre-trained on ImageNet, features are extracted from the images. These extracted features serve as input to a Random Forest classifier consisting of 100 decision trees. Following model training, predictions are made on a separate test set to evaluate performance. The model achieves an accuracy of 98.12%, demonstrating its effectiveness in distinguishing between healthy and unhealthy mango leaf images. This approach combines the strengths of both CNN feature extraction and Random Forest classification to provide a robust solution for image classification tasks.

5. Model Evaluation

Using a dedicated test set, the models are evaluated for accuracy. Accuracy measures the proportion of correctly classified mango leaf images. This requires the calculation of labels for test set images and comparing them to their real label. The accuracy is based on the model's ability to generalize and differentiate between healthy and unhealthy mango leaf images, demonstrating their effectiveness under different conditions.

IV. Results and Observation

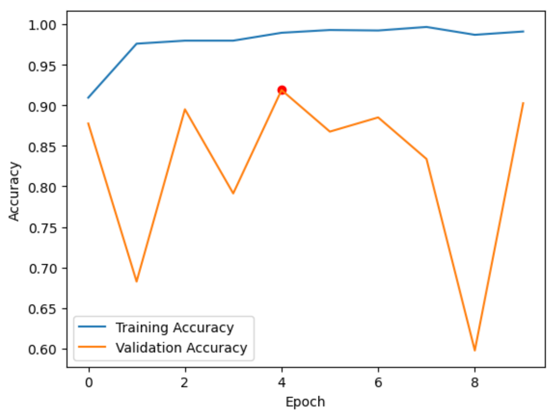

Figure 2. Graph represents comparison between Training and Validation Accuracy of CNN Model

Table 1. Accuracy of different models.

Algorithm	Accuracy (%)
CNN	91.87
RF	93.75
RF_CNN	98.12

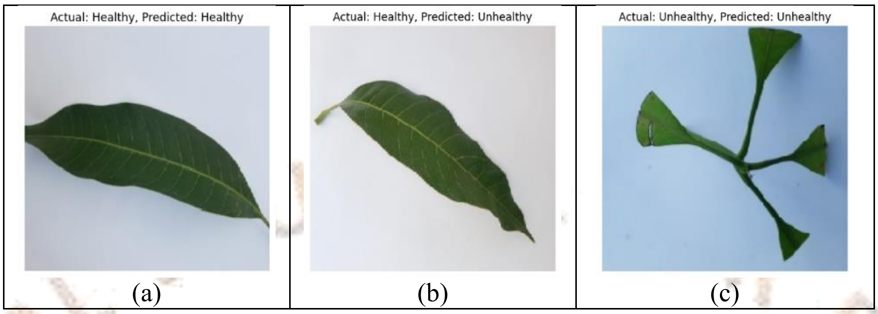

Figure 3. Actual versus Predicted Label
(a)CNN, (b)RF, (c)RF_CNN.

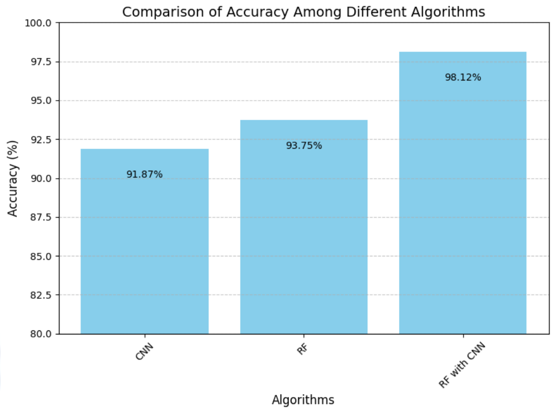

Figure 4. Graph represents comparison of Models. Here, Y-axis ranges from 80 to 100.

Observing the results alongside Table 1, significant differences emerge in the effectiveness of three distinct models utilized for mango leaf disease classification: Convolutional Neural Network (CNN), Random Forest (RF), and Random Forest with CNN feature extraction (RF_CNN). The Random Forest model achieved a superior accuracy rate of 93.75% compared to CNN's 91.87%. Moreover, the hybrid RF_CNN model, which combines CNN feature extraction with Random Forest classification, demonstrated the highest accuracy at 98.12%. These findings underscore the effectiveness of ensemble learning methodologies, particularly in agricultural classification tasks.

V. Conclusion and future work

The research aimed to detect and classify mango leaf diseases using machine learning techniques. Through meticulous data preprocessing and model development, including CNN, RF, and a hybrid RF_CNN approach, accurate classification of healthy and unhealthy mango leaves was achieved. Results showcased in Table 1 and Figure 4 the hybrid (RF_CNN) model's superior performance, with an accuracy of 98.12%. Proposed methodology promises efficient orchard management and supports sustainable agricultural practices, contributing to improved crop yield and reduced environmental impact.

References

[1] Saleem, Rabia, Jamal Hussain Shah, Muhammad Sharif, Mussarat Yasmin, Hwan-Seung Yong, and Jaehyuk Cha. "Mango leaf disease recognition and classification using novel segmentation and vein pattern technique." Applied Sciences 11, no. 24 (2021): 11901.

[2] Rajbongshi, Aditya, Thaharim Khan, M. M. R. A. Pramanik, Shah Md Tanvir, and Narayan Ranjan Chakraborty Siddiquee. "Recognition of mango leaf disease using convolutional neural network models: a transfer learning approach." Indonesian Journal of Electrical Engineering and Computer Science 23, no. 3 (2021): 1681-1688.

[3] Veling, P. S., Rohit S. Kalelkar, Likhita V. Ajgaonkar, Nivedita V. Mestry, and Nilesh N. Gawade. "Mango disease detection by using image processing." International journal for research in applied science and engineering technology 7, no. 4 (2019): 3717-3726.

[4] Mia, Md Rasel, Sujit Roy, Subrata Kumar Das, and Md Atikur Rahman. "Mango leaf disease recognition using neural network and support vector machine." Iran Journal of Computer Science 3, no. 3 (2020): 185-193.

[5] Arivazhagan, S., and S. Vineth Ligi. "Mango leaf diseases identification using convolutional neural network." International Journal of Pure and Applied Mathematics 120, no. 6 (2018): 11067-11079.

[6] Gining, R. A. J. M., S. S. M. Fauzi, N. M. Yusoff, T. R. Razak, M. H. Ismail, N. A. Zaki, and F. Abdullah. "Harumanis mango leaf disease recognition system using image processing technique." Indonesian Journal of Electrical Engineering and Computer Science 23, no. 1 (2021): 378-386.

[7] Saleem, Rabia, Jamal Hussain Shah, Muhammad Sharif, and Ghulam Jillani Ansari. "Mango Leaf Disease Identification Using Fully Resolution Convolutional Network." Computers, Materials & Continua 69, no. 3 (2021).

[8] Nikam, Ravindra, and Mahesh Sadavarte. "Application of Image Processing Technique in Mango Leaves Disease Severity Measurement." In National Conference on Emerging Trands in Computer, electrical and Electronics (ETCEE-2015), International Journal of Advance Engineering and Research Development (IJAERD). 2015.

[9] Ali, S., Ibrahim, M., Ahmed, S. I., Nadim, M., Mizanur, M. R., Shejunti, M. M., & Jabid, T. MangoLeafBD Dataset. Mendeley Data, 2022. doi: 10.17632/hxsnvwty3r.1