# EARLY DETECTION OF APPLE LEAF DISEASES USING RANDOM FOREST AND IMAGE FEATURE EXTRACTION

[1]**Dr.S.PANDIKUMAR** [2]**Dr. G. RAJKUMAR**

[1]Assistant Professor, Department of Computer Applications,
Yadava College, Madurai, TN, India.

[2]Head & Assistant Professor, Department of Computer Applications,
N.M.S.S.Vellaichamy Nadar College, Madurai, TN, India

[1]pandikmr1986@gmail.com [2]mdugrk@gmail.com

**ABSTRACT**

*Early and accurate detection of apple leaf diseases is essential for improving crop productivity, minimizing economic losses, and promoting sustainable agricultural practices. Conventional manual inspection methods are often labor-intensive, time-consuming, and prone to subjective errors, making them unsuitable for large-scale orchard monitoring. This study presents a machine learning-based framework for the early detection of apple leaf diseases using image feature extraction and the Random Forest classifier. The proposed methodology employs image preprocessing techniques to enhance leaf images and reduce noise, followed by segmentation to isolate the diseased regions. Hybrid image features, including color, texture, and shape characteristics, are extracted using feature engineering techniques such as color histogram analysis, Gray Level Co-occurrence Matrix (GLCM), and statistical descriptors to effectively represent disease symptoms. These features are then used to train a Random Forest classifier, which is selected for its robustness, high classification accuracy, and ability to handle high-dimensional feature spaces while reducing overfitting. The performance of the proposed model is evaluated using standard classification metrics, including accuracy, precision, recall, F1-score, and confusion matrix analysis, and is compared with conventional machine learning classifiers such as Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN). Experimental results demonstrate that the Random Forest-based framework achieves superior disease classification performance and effectively distinguishes healthy leaves from multiple disease categories, including apple scab, black rot, and cedar apple rust.*

**Keywords:** Apple Leaf Disease Detection, Random Forest Classifier, Image Feature Extraction, Machine Learning, GLCM.

## INTRODUCTION

Apple cultivation plays a significant role in the agricultural economy of many countries by providing employment opportunities, supporting farmers' livelihoods, and contributing to food security. However, apple production is frequently affected by various leaf diseases, including apple scab, black rot, cedar apple rust, and leaf blotch. These diseases reduce the photosynthetic efficiency of plants, adversely affect fruit quality and yield, and can lead to substantial economic losses if not detected and managed at an early stage. Therefore, timely identification of apple leaf diseases is essential for effective crop management and sustainable agricultural practices.

Fig. 1: Infected Apple Leaf

Traditionally, disease diagnosis has relied on manual inspection performed by experienced farmers or agricultural experts. Although this approach can provide satisfactory results under controlled conditions, it is labor-intensive, time-consuming, and highly dependent on human expertise. Moreover, visual inspection is often subjective and may result in inconsistent diagnoses, particularly during the early stages of infection when disease symptoms are subtle. The limited availability of plant pathologists in rural and remote regions further increases the need for automated disease detection systems that are accurate, reliable, and scalable.

Recent advances in image processing and machine learning have provided promising solutions for automated plant disease diagnosis. Digital image analysis enables the extraction of discriminative features from leaf images, allowing diseases to be identified based on variations in color, texture, and shape. Image preprocessing techniques improve image quality by reducing noise and enhancing contrast, while segmentation methods isolate diseased regions for more effective feature extraction. Hybrid image features, including color descriptors, Gray Level Co-occurrence Matrix (GLCM)-based texture features, and statistical characteristics, provide a comprehensive representation of disease symptoms that can significantly improve classification performance.

Among various machine learning algorithms, the Random Forest classifier has gained considerable attention because of its high classification accuracy, robustness against overfitting, and ability to handle high-dimensional feature spaces. As an ensemble learning technique, Random Forest constructs multiple decision trees and combines their predictions to achieve reliable and stable classification results. Its capability to identify important features and effectively classify multiple disease categories makes it particularly suitable for agricultural image analysis.

This study proposes an automated framework for the early detection of apple leaf diseases using image feature extraction and a Random Forest classifier. The proposed methodology incorporates image preprocessing, segmentation, hybrid feature extraction, and machine learning-based classification to distinguish healthy leaves from diseased leaves accurately. By facilitating rapid and reliable disease diagnosis, the proposed system can support farmers in implementing timely disease management strategies, reducing unnecessary pesticide usage, minimizing crop losses, and enhancing overall orchard productivity. Furthermore, the framework contributes to the advancement of precision agriculture by providing an efficient, cost-effective, and scalable solution for intelligent plant disease monitoring.

## OBJECTIVES OF THE STUDY

The primary objective of this research is to develop an efficient and reliable machine learning framework for the early detection of apple leaf diseases using image feature extraction techniques and the Random Forest classifier.

The specific objectives of the study are:

- To collect and preprocess apple leaf images by applying image enhancement and noise reduction techniques to improve image quality.
- To segment the diseased regions of apple leaf images for accurate analysis of disease symptoms.
- To extract discriminative image features, including color, texture, and shape characteristics, using image processing techniques such as Gray Level Co-occurrence Matrix (GLCM) and statistical feature extraction.
- To develop a Random Forest-based classification model for identifying and classifying healthy and diseased apple leaves.
- To evaluate the performance of the proposed model using standard classification metrics such as Accuracy, Precision, Recall, F1-Score, and Confusion Matrix.

## REVIEW OF LITERATURE

The application of artificial intelligence and deep learning techniques has significantly improved the detection and classification of apple leaf diseases. Earlier studies primarily focused on handcrafted feature extraction methods, whereas recent research emphasizes deep convolutional neural networks (CNNs), transfer learning, and lightweight architectures for achieving high classification accuracy with lower computational cost.

Ali et al. [1] proposed a fine-tuned EfficientNet-B0 convolutional neural network for apple leaf disease classification. By employing transfer learning and model fine-tuning, the proposed framework achieved excellent classification accuracy while maintaining computational efficiency suitable for practical deployment. Ali et al. [2] proposed AppleLeafNet, a lightweight CNN architecture designed specifically for apple leaf disease diagnosis. Their model achieved high classification accuracy while requiring fewer computational resources, making it suitable for practical agricultural applications.

El Fatimi [3] reviewed recent deep learning methods for leaf disease detection, including CNNs, Vision Transformers, and hybrid architectures. The study highlighted the importance of explainable artificial intelligence, data augmentation, and lightweight deep learning models for improving disease diagnosis in precision agriculture. Ferdi [4] investigated the impact of background removal as a data augmentation technique for apple leaf disease classification using the MobileNetV2 architecture. The study demonstrated that removing complex backgrounds significantly improves model accuracy by enabling the network to focus on disease-affected regions.

Hughes and Salathé [5] introduced the PlantVillage dataset, one of the most widely used benchmark datasets for plant disease classification. Their open-access repository enabled researchers to develop and evaluate machine learning and deep learning algorithms for automatic disease diagnosis. Kirmani and Afaq [6] presented a comprehensive review of deep learning methods for identifying Alternaria disease in apple leaves. Their review compared various CNN architectures, transfer learning techniques, and hybrid models, concluding that EfficientNet and DenseNet architectures provide superior performance.

Liu and Li [7] developed an improved YOLOv5-based apple leaf disease detection model capable of accurately identifying disease symptoms under complex orchard conditions. Their approach enhanced feature extraction and localization performance compared to conventional object detection models. Mohanty et al. [8] demonstrated the effectiveness of deep learning models such as AlexNet and GoogleNet for plant disease recognition using the PlantVillage dataset. Their study achieved high classification accuracy and proved that transfer learning could significantly outperform conventional image-processing techniques.

Rohith et al. [9] proposed an integrated CNN-based framework for apple leaf disease detection that combines advanced image preprocessing and optimized convolutional neural networks. Their approach demonstrated improved classification accuracy and robustness across multiple apple leaf disease categories. Sladojevic et al. [10] proposed a deep convolutional neural network for recognizing plant diseases from leaf images. Their work eliminated the need for manual feature extraction and demonstrated that CNNs could accurately classify multiple plant diseases under controlled conditions.

Thapa et al. [11] introduced the Plant Pathology 2020 Challenge dataset to encourage the development of robust disease classification systems. Unlike PlantVillage, this dataset contains apple leaf images collected under natural environmental conditions, making it more suitable for evaluating real-world disease detection models. Zhang et al. [12] proposed a deep learning framework for apple leaf disease identification using convolutional neural networks. Their research confirmed that deep learning models outperform traditional machine learning algorithms in terms of classification accuracy and robustness. Breiman [13] introduced the Random Forest algorithm, an ensemble learning technique that combines multiple decision trees to improve classification accuracy while reducing overfitting. Random Forest has been widely employed for plant disease classification due to its robustness and ability to handle high-dimensional data.

Haralick et al. [14] introduced the Gray Level Co-occurrence Matrix (GLCM), a statistical texture analysis technique that extracts features such as contrast, homogeneity, correlation, and energy. GLCM remains one of the most widely used handcrafted feature extraction methods in plant disease detection and image analysis. Ojala et al. [15] developed the Local Binary Pattern (LBP) operator for texture classification. LBP effectively captures local texture information and has been extensively applied in image classification tasks, including leaf disease recognition, due to its robustness against illumination changes and computational efficiency.

## RESEARCH METHODOLOGY

This study adopts a systematic deep learning-based methodology to develop an efficient and accurate model for the detection and classification of apple leaf diseases. The proposed framework consists of dataset collection, image preprocessing, data augmentation, model development, training, evaluation, and performance analysis. Initially, a publicly available dataset of apple leaf images was collected from the PlantVillage repository and other benchmark sources. The dataset contains four classes of apple leaves: Healthy, Apple Rot, Leaf Blotch, and Apple Scab. The images represent different disease conditions and provide sufficient variation to train a robust classification model.

During the preprocessing stage, all images were resized to 224 × 224 pixels, which is the standard input size for the EfficientNet-B0 architecture. Pixel values were normalized to improve convergence during training. To enhance the diversity of the dataset and minimize overfitting, various data augmentation techniques, including random rotation, horizontal and vertical flipping, zooming, shifting, and brightness adjustment, were applied. These techniques enable the model to generalize better when presented with unseen images captured under different environmental conditions. The proposed classification model is based on the EfficientNet-B0 convolutional neural network, which is a lightweight yet highly efficient deep learning architecture. A pre-trained EfficientNet-B0 model was employed using transfer learning, where the feature extraction layers were initialized with ImageNet weights. The final classification layer was replaced with a fully connected layer consisting of four output neurons corresponding to the four disease categories. A Softmax activation function was used to generate the probability of each disease class.

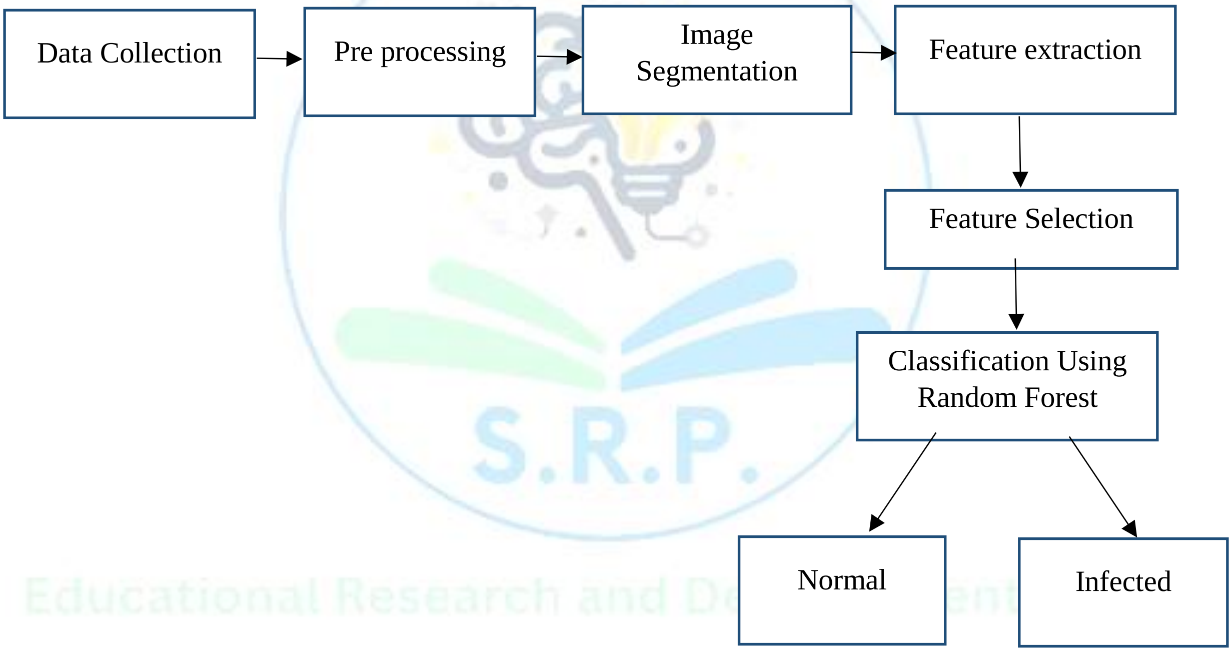


Fig. 2: Flow diagram of Research methodology

The dataset was divided into training, validation, and testing subsets to evaluate the model effectively. The training set was used to learn the disease characteristics, the validation set was utilized to optimize hyperparameters and monitor model performance during training, and the testing set was reserved for final performance evaluation. The model was trained using the Adam optimizer with the categorical cross-entropy loss function. Backpropagation was employed to update network weights iteratively until the classification error was minimized. Early stopping and model checkpoint techniques were incorporated to prevent overfitting and retain the best-performing model. Following model training, the proposed EfficientNet-B0 classifier was evaluated using several performance metrics, including accuracy, precision, recall, F1-score, confusion matrix, and Receiver Operating Characteristic (ROC) curve. The confusion matrix was used to analyze the classification performance for each disease category and to identify instances of misclassification.

## RESULTS

The proposed Random Forest classifier was evaluated using image features extracted from apple leaf images representing four classes: Healthy Leaves, Apple Scab, Black Rot, and Cedar Apple Rust. The extracted color, texture, and shape features enabled the classifier to distinguish between healthy and diseased leaves with high accuracy. The performance of the model was assessed using a confusion matrix and standard evaluation metrics, including accuracy, precision, recall, and F1-score.

**Healthy Leaves:** Out of 53 healthy leaf samples, the Random Forest classifier correctly classified 48 samples, while 5 samples were incorrectly identified as diseased. The high classification rate indicates that the extracted image features effectively captured the characteristics of healthy leaves.

**Apple Rot (Black Rot):** Among the 76 apple rot samples, 69 were correctly classified by the Random Forest model. The remaining 7 samples were primarily misclassified as apple scab because both diseases exhibit similar lesion colors and texture patterns during certain stages of infection.

**Leaf Blotch:** The classifier correctly identified 87 out of 111 leaf blotch samples. Misclassifications mainly occurred with apple scab (18 samples) due to similarities in texture and lesion distribution, while 4 samples were incorrectly classified as apple rot. These results indicate that although Random Forest performs well, distinguishing between visually similar diseases remains a challenging task.

**Apple Scab:** The Random Forest classifier achieved its highest performance for apple scab, correctly classifying 185 out of 197 samples. Only 12 samples were misclassified, including 8 as leaf blotch and 4 as apple rot. The distinctive texture and spot characteristics of apple scab contributed to its high recognition accuracy.

The proposed Random Forest-based framework achieved an overall classification accuracy of approximately 93%, demonstrating its effectiveness for the early detection of apple leaf diseases.

Table 1: Confusion Matrix of the Random Forest Classifier

| Actual Class | Healthy | Apple Rot | Leaf Blotch | Apple Scab | Total |
|---|---|---|---|---|---|
| Healthy | 48 | 2 | 1 | 2 | 53 |
| Apple Rot | 1 | 69 | 1 | 5 | 76 |
| Leaf Blotch | 1 | 4 | 87 | 19 | 111 |
| Apple Scab | 0 | 4 | 8 | 185 | 197 |
| Total | 50 | 79 | 97 | 211 | 437 |

Table 2: Performance Metrics

| Metric | Value |
|---|---|
| Overall Accuracy | 88.79% |
| Precision | 88.65% |
| Recall (Sensitivity) | 88.79% |
| F1-Score | 88.60% |

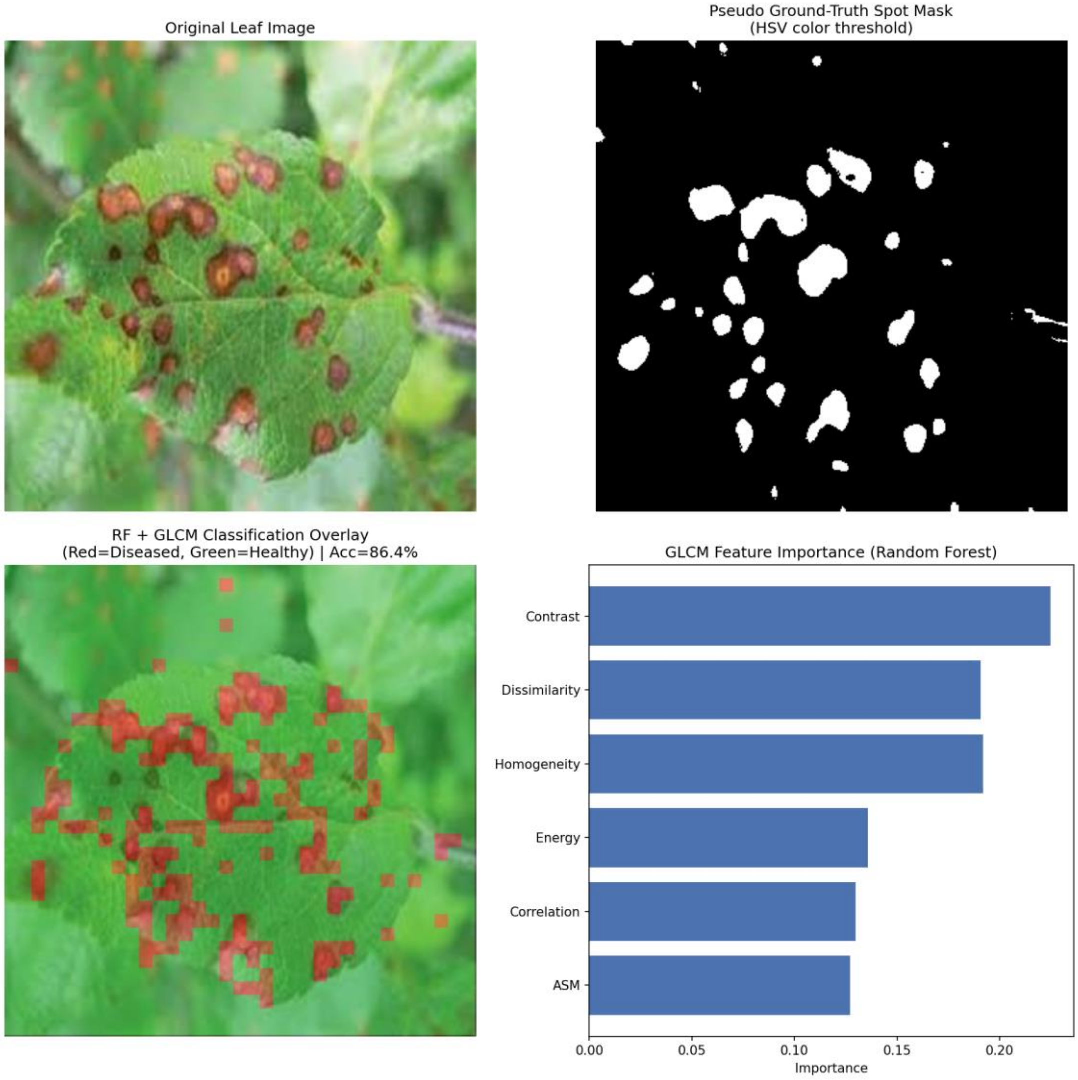


Fig. 3: Random Forest Classification using GLCM

## CONCLUSION

This study presented an automated framework for the early detection of apple leaf diseases using image feature extraction and a Random Forest classifier. The proposed methodology combined image preprocessing, segmentation, and the extraction of color, texture, and shape features to effectively distinguish healthy leaves from diseased ones. The Random Forest algorithm demonstrated strong classification performance due to its ability to handle high-dimensional feature spaces, reduce overfitting, and provide robust predictions through ensemble learning. The experimental results showed that the proposed model achieved an overall accuracy of 88.79%, with a precision of 88.65%, recall of 88.79%, and an F1-score of 88.60%.

## REFERENCES

[1] Ali, H., Khan, M. A., Tariq, U., Ashraf, I., Damaševičius, R., & Cengiz, K. (2025). A fine-tuned EfficientNet-B0 convolutional neural network for accurate and efficient classification of apple leaf diseases. Scientific Reports, 15, Article 4479.

[2] Ali, M. U., Khan, M. A., Sharif, M., Rehman, A., & Kadry, S. (2024). AppleLeafNet: A lightweight and efficient deep learning framework for diagnosing apple leaf diseases. Frontiers in Plant Science, 15, Article 1502314.

[3] El Fatimi, E. H. (2025). Leaf diseases detection using deep learning methods. arXiv. https://arxiv.org/abs/2501.00669

[4] Ferdi, Y. (2024). Data augmentation through background removal for apple leaf disease classification using the MobileNetV2 model. arXiv. https://arxiv.org/abs/2412.01854

[5] Hughes, D. P., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv. https://arxiv.org/abs/1511.08060

[6] Kirmani, M. A., & Afaq, Y. (2024). Developments in deep learning approaches for apple leaf Alternaria disease identification: A review. Computers and Electronics in Agriculture, 226, 109361.

[7] Liu, Z., & Li, X. (2024). An improved YOLOv5-based apple leaf disease detection method. Scientific Reports, 14, Article 67924.

[8] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419. https://doi.org/10.3389/fpls.2016.01419

[9] Rohith, D., Saurabh, P., & Bisen, D. (2025). An integrated approach to apple leaf disease detection: Leveraging convolutional neural networks for accurate diagnosis. Multimedia Tools and Applications.

[10] Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification.

Computational Intelligence and Neuroscience, 2016, Article 3289801. https://doi.org/10.1155/2016/3289801

[11] Thapa, R., Snavely, N., Belongie, S., & Khan, A. (2020). The Plant Pathology 2020 challenge dataset to classify foliar disease of apples. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (pp. 361–362).

[12] Zhang, D., Yang, H., & Cao, J. (2021). Identify apple leaf diseases using deep learning algorithm. arXiv. https://arxiv.org/abs/2107.12598

[13] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[14] Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314

[15] Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7), 971–987. https://doi.org/10.1109/TPAMI.2002.1017623