**Journal of Information Systems Engineering and Management**
2025, 10(25s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

# An Advanced Machine Learning Framework for Apple Tree Leaf Disease Identification: Integrating MS-LBP, HOG, and MRMR Feature Selection with XGBoost Classification

Ankita Mandloi [1], Kriti Joshi [2]
11Medi-Caps University, Department of Computer Science and Engineering, 453331, Indore-India
Corresponding AuthorEmail:ankita.mandloi@medicaps.ac.in- ORCID:0000-0002-5247-7850
21Medi-Caps University, Department of Computer Science and Engineering, 453331, Indore-India
Email:kriti.joshi@medicaps.ac.in- ORCID:0009-0003-7792-8157

**ARTICLEINFO**

Received: 21 Dec 2024

Revised: 12 Feb 2025

Accepted: 25 Feb 2025

**ABSTRACT**

**:** The agricultural industry faces major economic hurdles from apple tree leaf diseases when these diseases remain undetected or unmanaged in their timely manner. The research develops an effective method which employs image processing and machine learning techniques to detect and classify diseases in apple tree leaves. A combination of MSLBP features with HOG features helps extract extensive texture features that the method combines into one unified vector. Minimum Redundancy Maximum Relevance (MRMR) selects the most disease-diagnostic features from the vector through a thorough evaluation process. XGBoost classifier along with robust ensemble features achieves classification of selected features which exhibits high performance values. The novel approach contains feature selection and integration that leads to 98.03% classification accuracy compared to the 97.56% accuracy of the baseline method. The system provides improved accuracy alongside solid performance across different environmental conditions which makes it an optimal solution for real-time disease management in apple tree leaf detection for precision agriculture.

**Keywords:** HOG (Histogram of Oriented Gradients), MSLBP (Multi-Scale Local Binary Patterns), MRMR (Minimum Redundancy Maximum Relevance), XGBoost Classifier

## INTRODUCTION

Technological progress in agriculture has introduced revolutionary methods during the last few years for monitoring plant health particularly within the high-value sector of apples. The apple tree stands among the most prevalent fruit planting varieties yet remains vulnerable to both bacterial and fungal attackers together with virus infections. These illnesses reduce apple production amounts and quality while presenting significant challenges to worldwide apple businesses. Disease detection needs to reach exact levels during early stages to minimize economic losses and reduce pesticide use for keeping crops healthy and resistant to disease. The detection of diseases through manual inspection proves inefficient because it takes considerable time and involves human errors due to human participation and requires a great deal of manual labor. The strong demands for automated systems that detect diseases early with high precision also provide necessary solutions for crop preservation [1].

Plant illness recognition became significantly better through the combination of positive and influential methods from machine learning (ML) and image processing and artificial intelligence (AI) which provide benefits to plant health and multiple monetary and non-monetary aspects [2] [3]. Apple tree disease diagnosis systems which detect diseases automatically emerged because these technological systems proved more effective in managing diseases through their operational features. System-analysis depends heavily on the feature extraction method because it extracts relevant features from leaf images before accurate diagnosis.

Sanitation detection of apple tree leaf diseases uses a new approach involving textural feature extraction combined with ensemble learning along with feature choice methods. The system begins its operation by merging texture features obtained through the combination of Multi-Scale Local Binary Patterns (MSLBP) and Histogram of Oriented Gradients (HOG) to create a unified feature vector. The integrated features provide necessary information at macro and micro scales to detect underlying indicators of disease manifestations.

Feature selection stands as an essential part of the workflow because it maintains the appropriate selection of characteristics which will be employed for classification tasks. This process enables the model to deliver better performance because it removes unnecessary characteristics. The Minimum Redundancy Maximum Relevance (MRMR) methodology serves as the feature selection tool that optimizes the feature collection for improved decision-making in classification systems [3] [4] [5].

The classification model accepts a feature vector created by combining the most important collected features. This analysis applies the sophisticated machine learning technique XGBoost classifier because it combines several weak classifiers to create one robust classifier. The XGBoost technique strengthens its performance through successive iterations by applying its training to the errors made by its previous decision-making classifier to focus on harder prediction cases. An ensemble method decreases overfitting risks while producing effective classification outcomes across various conditions including different lighting varieties and leaf colors and backgrounds [6]-[10].

This method enables fast image classification by producing reliable solutions at optimal efficiency when processing images contaminated by environmental factors that often affect agricultural inspections. The application of multiple strong components (specifically MSLBP and HOG features with MRMR selection and XGBoost algorithm) will result in an efficient system with high performance output. The integration of these methods enables fast disease identification of apple tree leaves that precision agricultural monitoring systems can use for disease assessment and management needs.

Applying timely detection of leaf diseases in apple trees is essential both for orchard health and productivity. The quick proliferation of apple tree diseases becomes a fact of life when apple scab and powdery mildew and fire blight remain unidentified or properly managed during early stages. Early identification lets farmers limit their interventions to necessary zones rather than resorting to unspecific chemical spraying. Improved agricultural output and plant health occurs together with responsible farming methods through early detection system implementation. The identification of diseases at their early phases prevents widespread outbreaks thus protecting agricultural firms from extensive economic losses.

The system provides mutual advantages at various levels because it integrates modern analysis techniques with conventional plant health assessment components. It serves as an uncomplicated cost-efficient method that solves the major problem of manual disease identification processes. Multiple deployment methods collaborate to form a system which can be used in standard agricultural fields.

The rest of this paper is organized as follows: A broad literature review of related works is presented in section II. The proposed methodology and control strategy are laid out in Section III. The results from the simulation are discussed in the fourth part of this document. Finally, Section V concludes with a summary of the main findings and a summary of the conclusion.

## OBJECTIVES

Research now focuses on monitoring crops and plantations in addition to detecting tree diseases during their early stages. Sustainable ecosystem health depends on threat detection at its initial phases. Visual inspection and field observation system supports early detection together with new analytical methods. Forestry experts are capable of diagnosing early disease signs and death symptoms and spoilage marks as well as other abnormal traits that professional inspectors observe in trees across forested areas. Plant experts have the ability to detect initial warning signs through the observations of resin leakage while bark splinters and new pest interpretation that indicates plant sickness.

The traditional diagnostic approaches remain useful yet there is rising need for improved advanced diagnostic tools. AI technology stands out among recent developments because it lets doctors perform rapid with precise automated image analysis for disease symptom screening. The academic literature presents some works on automatic tree disease recognition methods. Machine learning consequently creates models through neural network technology to detect specific illness symptoms.

A CNN model serves as an approach to diagnose and classify apple leaf diseases according to [11]. The framework consists of four main segments including data preparation along with segmentation along with feature extraction that leads to classification. The first method in the process utilizes contrast stretching to employ image details enhancement. The FCM based clustering algorithm functions to locate areas affected by diseases in the leaf. An algorithm can identify the class of input images which determine healthy and disease conditions.

The authors in [12] demonstrated the utilization of GA and CFS for detecting apple leaf diseases [12]. The system follows multiple operational steps beginning with image collection of apple leaves followed by preprocessing and segmenting with RGA before performing features extractors. Through GA and CFS feature selection the number of relevant features increases which leads to improved disease diagnosis correctness in reduced dimensions of feature space.

The authors in [13] implemented the CNN LeNet architecture to identify healthy or sick apple leaves in separate categories. The LeNet network proved its ability to classify plant diseases including specific detection of healthy and unhealthy apple leaves. According to the study data augmentation stands out as essential for model enhancement because it creates diverse visual patterns between training images.

The authors in [14] developed an enhanced deep learning platform for detecting diseases on apple tree leaves. The research methodology makes use of three contemporary techniques that include Faster Region-Based CNN (Faster RCNN) and Improved Selective Residual Networks (ISResNet) and Feature Pyramid Networks (FPN). The three stages of this system include FPN-based feature extraction followed by ISResNet refinement then classification occurs via Faster RCNN. A multistage deep learning system uses various applications to detect affected areas of disease on apple leaves with efficiency.

Professionals from [15] conducted research into deep learning algorithms for precise disease detection and classification using separate leaf image analysis. Researchers gathered a categorized database composed of various apple leaf images illustrating different plant diseases for this investigative work. The authors utilized transfer learning for training their CNN architecture on disease recognition tasks. The software implementation of the diagnostic system achieved high diagnostic accuracy through its application of the trained model to the dataset containing apple leaf images.

This research paper proposed a model for detecting diseases in apple leaves through a process that included image preprocessing followed by diseased area removal and feature extraction before classifier training [16]. Images entered the LAB color space before threshold-based segmentation removed background elements from the frame. The extraction of relevant information occurred from the gray co-occurrence matrix using texture features calculation in combination with color moments once the green mask was eliminated. The SVM classifier received trained features to distinguish diverse apple leaf diseases with successful results for detecting plant diseases.

Recent advancements in artificial intelligence and machine learning have made apple tree leaf disease detection and classification automation more important as documented in relevant literature. In addition to improving diagnostic quality these methods provide basis for creating adaptable disease management systems suitable for agricultural applications.

***Findings from Literature:*** Many problems still exist despite notable improvements in automated technology designed to detect diseases in apple tree leaves. The current set of disease detection systems mainly operate on healthy or diseased classification whereas they fail to detect multiple diseases simultaneously. These systems need substantial datasets for operation yet such large datasets become impractical to obtain when addressing rare diseases. Current automated methods fail to detect the complex patterns that appear in diseased leaves therefore leading to decreased performance quality.

The proposed method enhances simultaneous diagnosis accuracy and specificity of various apple tree leaf diseases through a newly developed approach. The study separates from standard methods that depend on GLCM texture features by implementing Multi-Scale Local Binary Patterns (MSLBP) alongside Histogram of Oriented Gradients (HOG) to obtain an advanced textural feature collection. Such extensive feature collection enables the detection of disease patterns that exist at different scales from large to small details.

Research improvements in feature selection emerge as a major contribution in this work. The study utilizes Minimum Redundancy Maximum Relevance (MRMR) to choose which features will be used for classification purposes by selecting only those with maximum relevance and minimal redundancy. Using this method allows a reduced feature redundancy that leads to superior model performance when working with restricted data quantities.

The implementation of XGBoost classifier as a sophisticated ensemble learning method enhances system durability because it suppresses overfitting events. Under various environmental settings this improvement enables better performance which makes it well-suited for real-life agricultural conditions.

These advanced methods enable high-speed operations when combined while keeping their predictive capabilities close to pathological standards and delivering precise explanations. The new disease management system stands as a major achievement in precision agriculture because it allows efficient disease response through timely intervention tactics.

## METHODS

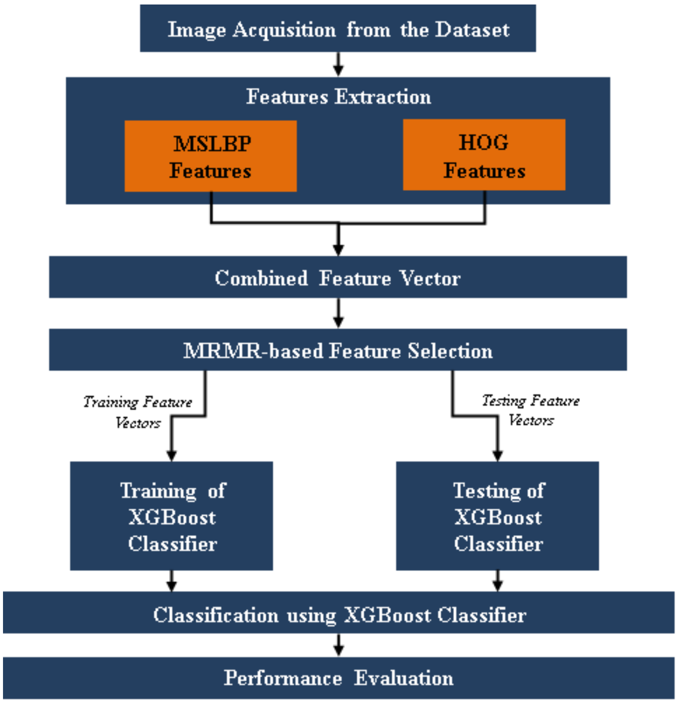


Figure. 1 Flow Diagram for Proposed Plant Leaf Disease Detection

### 1.1 Image Acquisition from the Dataset

The geographic coordinates of each image in the dataset are indicated by the variables $x$ and $y$, which together form a 2D matrix $I(x, y)$. For a grayscale image, the matrix $I(x, y)$ corresponds to the intensity value at pixel coordinates $(x, y)$, defined as:

$$(x, y) \in [0,255] \tag{1}$$

For RGB images, the matrix is extended to a 3D tensor to accommodate the three color channels:

$$I(x, y) = \{I_R(x, y), I_G(x, y), I_B(x, y)\} \tag{2}$$

Where:

- $I_R(x, y)$ represents the intensity of the red channel,
- $I_G(x, y)$ represents the intensity of the green channel,
- $I_B(x, y)$ represents the intensity of the blue channel.

Image Acquisition Process

- Loading Images: Images are directly loaded from the dataset repository. Each image retains its original resolution and characteristics, as provided in the dataset.
- Assigning Labels: Each image has a label associated with it that falls into one of the four categories. The labels are essential in supervised learning tasks because they help train the model to distinguish between these classes according to the visual characteristics of the leaves.

### 1.2 Feature Extraction using HoG and MS-LBP

The next step in our methodology involves the extraction of features from the leaf images. We apply two robust feature extraction methods, Histogram of Oriented Gradients (HoG) and Multi-Scale Local Binary Patterns (MS-LBP), to obtain quantitative features that can help in the identification and classification of different leaf diseases.

#### 1.2.1 Histogram of Oriented Gradients (HoG)

The HoG technique captures the shape and texture details of the leaf by analyzing the distribution of edge directions, which is vital for outlining the affected areas on the leaves.

To compute the gradient of each pixel, we use the formula:

$$Magnitude = \sqrt{(I_x)^2 + (I_y)^2} \tag{3}$$

Where, $I_x$ and $I_y$ are the gradients in the $x$ and $y$ directions, respectively.

The image is then divided into small cells (e.g., 8×8 pixels), and a histogram of gradients is calculated for each cell. These histograms are normalized and concatenated to form the HoG feature vector.

Steps:

- Calculate the gradient magnitude and angle for each pixel.
- Divide the image into cells and compute the HoG for each cell.
- Normalize histograms across blocks (e.g., 2×2 cells).
- Stack the histograms of all blocks to create the composite HoG feature vector.

### 1.2.2 Multi-Scale Local Binary Patterns (MS-LBP)

MS-LBP is utilized to analyze the texture of the leaf by comparing the intensity of each pixel with its neighbors across multiple scales, which is effective for detecting subtle textural differences indicative of disease.

For each pixel $f(x, y)$ in the image, the LBP is computed by comparing it to its neighbors in a circular pattern:

$$LBP(i,j) = \sum_{p=0}^{P-1} s\left(f(x,y), f\left(x_p, y_p\right)\right) \cdot 2^p \tag{4}$$

Where:

- $f(x, y)$ is the intensity of the center pixel,
- $f\left(x_p, y_p\right)$ represents the intensity of the surrounding pixels,
- $s(x, y)$ is the threshold function, 1 if $x \geq y$ and 0 otherwise,
- $P$ is the number of neighbors in the circular pattern.

This process is applied at multiple scales to capture detailed and coarse textures effectively.

Steps:

- Calculate LBP for each pixel using its immediate neighborhood.
- Apply LBP at multiple scales to encompass various textural details.
- Concatenate the LBP features from all scales to form the comprehensive feature vector.

### 1.2.3 Combined Feature Vector

The individual feature vectors (HoG and MS-LBP) are merged to create a unified feature vector. This vector encapsulates both structural (shape) and textural (pattern) information crucial for accurately identifying various diseases in apple tree leaves.

Mathematical Formulation:

$$F = [F_{HoG}, F_{MS-LBP}] \tag{5}$$

Where:

- $F_{HoG}$ is the HoG feature vector,
- $F_{MS-LBP}$ is the MS-LBP feature vector.

The integrated feature vector is applied as input for feature selection and classification.

## 1.3 Feature Selection using MRMR

We apply the Minimum Redundancy Maximum Relevance (MRMR) technique to achieve a reduced dimensionality feature vector that is optimal for classification. MRMR selects features that are highly relevant to the target condition (different leaf diseases) while minimizing redundancy among them.

$$MRMR(S) = \frac{1}{|S|} \sum_{i \in S} Relevance(i) - \frac{1}{|S|^2} \sum_{i,j \in S} Redundancy(i,j) \tag{6}$$

- $Relevance(i)$ is the relevance of feature ii with respect to the disease,
- $Redundancy(i, j)$ is the redundancy between features $i$ and $j$,
- $S$ is the set of selected features.

This feature selection method enhances the efficiency of the classification process by eliminating unnecessary or redundant features, thereby streamlining the detection and categorization of apple tree leaf diseases.

### 1.4 Classification Using XGBoost Classifier

XGBoost (eXtreme Gradient Boosting) serves as our main classification tool because its ensemble learning approach balances assessment precision and execution efficiency in detecting apple tree leaf diseases. The XGBoost classifier bases its operation on gradient boosting which enhances previous trees by identifying and fixing their mistakes before targeting hard to categorize observations.

***Objective Function:*** The objective function of XGBoost incorporates two elements which include a loss measure and regularization terms that manage overfitting. The iteration $t$ objective statement takes the following form:

$$Obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{7}$$

Where:

- $y_i$ is the actual label for the $i^{th}$ instance.
- $\hat{y}_i^{(t-1)}$ is the prediction from the model for the $i^{th}$ instance up to the $t-1$ iteration.
- $f_t$ represents the tree added at the $t^{th}$ iteration.
- $\Omega(f)$ represents the regularization term which penalizes the complexity of the model.

***Loss Function:*** The loss function $l$ determines predicted-label mismatch value while the selection choice depends upon the problem nature (like logistic regression for categorization). When working with binary classification the logistic loss serves as one of the predominant loss functions.

$$l(y, \hat{y}) = y \log\left(1 + e^{-\hat{y}}\right) + (1 - y) \log(1 + e^{\hat{y}}) \tag{8}$$

***Regularization Term:*** Model complexity gets decreased through the regularization term $\Omega(f)$which acts as an essential preventive measure against overfitting. It is defined as:

$$\Omega(\mathrm{f}) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{9}$$

Where:

- $\gamma$ and $\lambda$ are parameters that control the degree of regularization.
- $T$ is the number of leaves in the tree.
- $w_j$ are the weights associated with the leaves.

***Gradient and Hessian in Tree Learning:***The optimization process in XGBoost depends on first and second derivatives of loss function gradients and Hessians to find optimal splitting points for tree nodes.

$$g_i = \partial_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right)$$
$$h_i = \partial^2_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right) \tag{10}$$

Where $g_i$ and $h_i$ are the gradient and Hessian for the $i^{th}$ instance, respectively.

***Tree Structure and Split Finding:***XGBoost grows trees through greed-based splitting which seeks to achieve maximum information gain from each split through:

$$Gain = \frac{1}{2}\left[\frac{(\sum_{i\in L} g_i)^2}{\sum_{i\in L} h_i + \lambda} + \frac{(\sum_{i\in R} g_i)^2}{\sum_{i\in R} h_i + \lambda} - \frac{(\sum_{i\in node} g_i)^2}{\sum_{i\in node} h_i + \lambda}\right] - \gamma \tag{11}$$

$L$ and $R$ denote the left and right splits of the data at a node, respectively.

## RESULTS

### 1.5 Image Acquisition from the Dataset

The geographic coordinates of each image in the dataset are indicated by the variables $x$ and $y$, which together form a 2D matrix $I(x, y)$. For a grayscale image, the matrix $I(x, y)$ corresponds to the intensity value at pixel coordinates $(x, y)$, defined as:

$$(x, y) \in [0,255] \tag{1}$$

For RGB images, the matrix is extended to a 3D tensor to accommodate the three color channels:

$$I(x, y) = \{I_R(x, y), I_G(x, y), I_B(x, y)\} \tag{2}$$

Where:

- $I_R(x, y)$ represents the intensity of the red channel,
- $I_G(x, y)$ represents the intensity of the green channel,
- $I_B(x, y)$ represents the intensity of the blue channel.

Image Acquisition Process

- Loading Images: Images are directly loaded from the dataset repository. Each image retains its original resolution and characteristics, as provided in the dataset.
- Assigning Labels: Each image has a label associated with it that falls into one of the four categories. The labels are essential in supervised learning tasks because they help train the model to distinguish between these classes according to the visual characteristics of the leaves.

## 1.6 Feature Extraction using HoG and MS-LBP

The next step in our methodology involves the extraction of features from the leaf images. We apply two robust feature extraction methods, Histogram of Oriented Gradients (HoG) and Multi-Scale Local Binary Patterns (MS-LBP), to obtain quantitative features that can help in the identification and classification of different leaf diseases.

### 1.6.1 Histogram of Oriented Gradients (HoG)

The HoG technique captures the shape and texture details of the leaf by analyzing the distribution of edge directions, which is vital for outlining the affected areas on the leaves.

To compute the gradient of each pixel, we use the formula:

$$Magnitude = \sqrt{(I_x)^2 + (I_y)^2} \tag{3}$$

Where, $I_x$ and $I_y$ are the gradients in the $x$ and $y$ directions, respectively.

The image is then divided into small cells (e.g., 8×8 pixels), and a histogram of gradients is calculated for each cell. These histograms are normalized and concatenated to form the HoG feature vector.

Steps:

- Calculate the gradient magnitude and angle for each pixel.
- Divide the image into cells and compute the HoG for each cell.
- Normalize histograms across blocks (e.g., 2×2 cells).
- Stack the histograms of all blocks to create the composite HoG feature vector.

### 1.6.2 Multi-Scale Local Binary Patterns (MS-LBP)

MS-LBP is utilized to analyze the texture of the leaf by comparing the intensity of each pixel with its neighbors across multiple scales, which is effective for detecting subtle textural differences indicative of disease.

For each pixel $f(x, y)$ in the image, the LBP is computed by comparing it to its neighbors in a circular pattern:

$$LBP(i,j) = \sum_{p=0}^{P-1} s\left(f(x,y), f(x_p, y_p)\right) \cdot 2^p \tag{4}$$

Where:

- $f(x, y)$ is the intensity of the center pixel,
- $f(x_p, y_p)$ represents the intensity of the surrounding pixels,
- $s(x, y)$ is the threshold function, 1 if $x \geq y$ and 0 otherwise,
- $P$ is the number of neighbors in the circular pattern.

This process is applied at multiple scales to capture detailed and coarse textures effectively.

Steps:

- Calculate LBP for each pixel using its immediate neighborhood.
- Apply LBP at multiple scales to encompass various textural details.
- Concatenate the LBP features from all scales to form the comprehensive feature vector.

### 1.6.3 Combined Feature Vector

The individual feature vectors (HoG and MS-LBP) are merged to create a unified feature vector. This vector encapsulates both structural (shape) and textural (pattern) information crucial for accurately identifying various diseases in apple tree leaves.

Mathematical Formulation:

$$F = [F_{HoG}, F_{MS-LBP}] \tag{5}$$

Where:

- $F_{HoG}$ is the HoG feature vector,
- $F_{MS-LBP}$ is the MS-LBP feature vector.

The integrated feature vector is applied as input for feature selection and classification.

## 1.7 Feature Selection using MRMR

We apply the Minimum Redundancy Maximum Relevance (MRMR) technique to achieve a reduced dimensionality feature vector that is optimal for classification. MRMR selects features that are highly relevant to the target condition (different leaf diseases) while minimizing redundancy among them.

$$MRMR(S) = \frac{1}{|S|}\sum_{i\in S} Relevance(i) - \frac{1}{|S|^2}\sum_{i,j\in S} Redundancy(i,j) \tag{6}$$

- $Relevance(i)$ is the relevance of feature ii with respect to the disease,
- $Redundancy(i,j)$ is the redundancy between features $i$ and $j$,
- $S$ is the set of selected features.

This feature selection method enhances the efficiency of the classification process by eliminating unnecessary or redundant features, thereby streamlining the detection and categorization of apple tree leaf diseases.

## 1.8 Classification Using XGBoost Classifier

XGBoost (eXtreme Gradient Boosting) serves as our main classification tool because its ensemble learning approach balances assessment precision and execution efficiency in detecting apple tree leaf diseases. The XGBoost classifier bases its operation on gradient boosting which enhances previous trees by identifying and fixing their mistakes before targeting hard to categorize observations.

***Objective Function:*** The objective function of XGBoost incorporates two elements which include a loss measure and regularization terms that manage overfitting. The iteration $t$ objective statement takes the following form:

$$Obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{7}$$

Where:

- $y_i$ is the actual label for the $i^{th}$ instance.
- $\hat{y}_i^{(t-1)}$ is the prediction from the model for the $i^{th}$ instance up to the $t-1$ iteration.
- $f_t$ represents the tree added at the $t^{th}$ iteration.
- $\Omega(f)$ represents the regularization term which penalizes the complexity of the model.

***Loss Function:*** The loss function $l$ determines predicted-label mismatch value while the selection choice depends upon the problem nature (like logistic regression for categorization). When working with binary classification the logistic loss serves as one of the predominant loss functions.

$$l(y,\hat{y}) = y\log\left(1+e^{-\hat{y}}\right) + (1-y)\log\left(1+e^{\hat{y}}\right) \tag{8}$$

***Regularization Term:*** Model complexity gets decreased through the regularization term $\Omega(f)$which acts as an essential preventive measure against overfitting. It is defined as:

$$\Omega(\mathrm{f}) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2 \tag{9}$$

Where:

- $\gamma$ and $\lambda$ are parameters that control the degree of regularization.
- $T$ is the number of leaves in the tree.
- $w_j$ are the weights associated with the leaves.

***Gradient and Hessian in Tree Learning:***The optimization process in XGBoost depends on first and second derivatives of loss function gradients and Hessians to find optimal splitting points for tree nodes.

$$g_i = \partial_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right)$$
$$h_i = \partial^2_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right) \tag{10}$$

Where $g_i$ and $h_i$ are the gradient and Hessian for the $i^{th}$ instance, respectively.

***Tree Structure and Split Finding:***XGBoost grows trees through greed-based splitting which seeks to achieve maximum information gain from each split through:

$$Gain = \frac{1}{2}\left[\frac{(\sum_{i\in L} g_i)^2}{\sum_{i\in L} h_i+\lambda} + \frac{(\sum_{i\in R} g_i)^2}{\sum_{i\in R} h_i+\lambda} - \frac{(\sum_{i\in node} g_i)^2}{\sum_{i\in node} h_i+\lambda}\right] - \gamma \tag{11}$$

$L$ and $R$ denote the left and right splits of the data at a node, respectively.

## DISCUSSION

The proposed system for detecting apple tree leaf diseases yields its outcomes in this section. The performance method shows success through several metrics which include accuracy ratings alongside sensitivity and specificity ratings and other measurements. Each figure and table within the presentation focuses on distinctive performance facets demonstrating the upshot of connecting machine learning with sophisticated extracting and selection features. Our approach demonstrates reliable operation across different settings thus we gained complete knowledge about how the system works to detect and classify apple tree leaf diseases.

### 1.9 Dataset

The apple tree leaf disease data available on Kaggle [17] contains numerous images that provide opportunities for detecting and categorizing apple tree leaf diseases. These datasets facilitate both plant pathology research development and automated early disease diagnosis system development. The key apple tree conditions present in Table 1 are categorized into Healthy, Apple Scab, Black Rot and Cedar Apple Rust.

Table 1. Sample Images from Apple Leaf Disease Dataset [17]

| | |
|---|---|
| Healthy Leaves |  |
| Apple Scab |  |
| Black Rot |  |
| Cedar Apple Rust |  |

### 1.10 Evaluation Parameters

The research work uses the evaluation parameters from Table 2 for evaluation purpose.

Table 2. Evaluation Parameters

| | |
|---|---|
| TP (True Positive) | “Represents the number of apple leaf images where the model correctly identifies the presence of a specific disease, such as Apple Scab, Black Rot, or Cedar Apple Rust.” |
| TN (True Negative) | “Indicates the number of apple leaf images correctly classified as healthy, with no signs of disease present.” |
| FP (False Positive) | “Represents the number of apple leaf images incorrectly classified as diseased, falsely indicating the presence of conditions such as Apple Scab, Black Rot, or Cedar Apple Rust when the leaf is actually healthy.” |
| FN (False Negative) | “Indicates the number of apple leaf images where a disease is present, but the model fails to detect it, misclassifying the diseased leaf as healthy or as having a different condition.” |

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (14)$$

$$Specificity = \frac{TN}{TN+FN} \quad (15)$$

$$ErrorRate = \frac{FP+FN}{TP+TN+FP+FN} \quad (16)$$

$$FalsePositiveRate(FPR) = \frac{FP}{FP+TN} \quad (17)$$

$$F-Score = \frac{2TP}{2TP+FP+FN} \quad (18)$$

## 1.11 Results

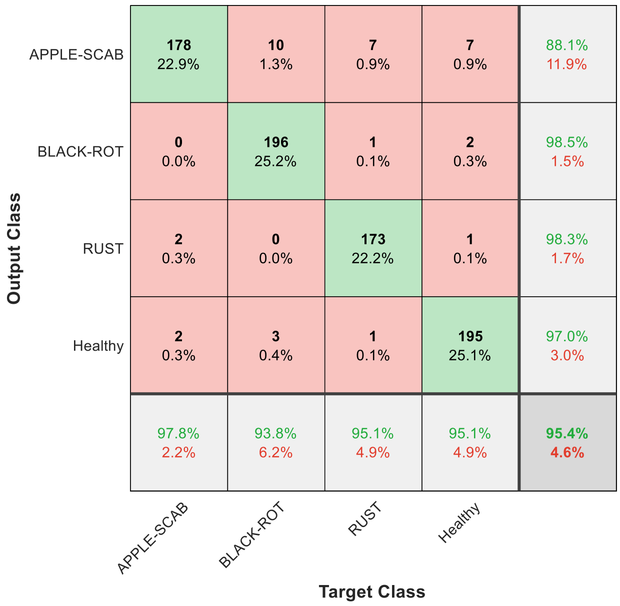


Figure. 2 Confusion matrix for the proposed apple tree leaf disease detection using MSLBP features

The confusion matrix showing MSLBP feature capability for detecting apple tree leaf diseases is displayed in Figure 2. The matrix demonstrates MSLBP feature capability to discriminate leaf health states by reporting the counts of true positives, true negatives, false positives and false negatives. MSLBP features show exceptional efficacy for disease identification due to their ability to provide precise accuracy and error rate metrics from the confusion matrix analysis.

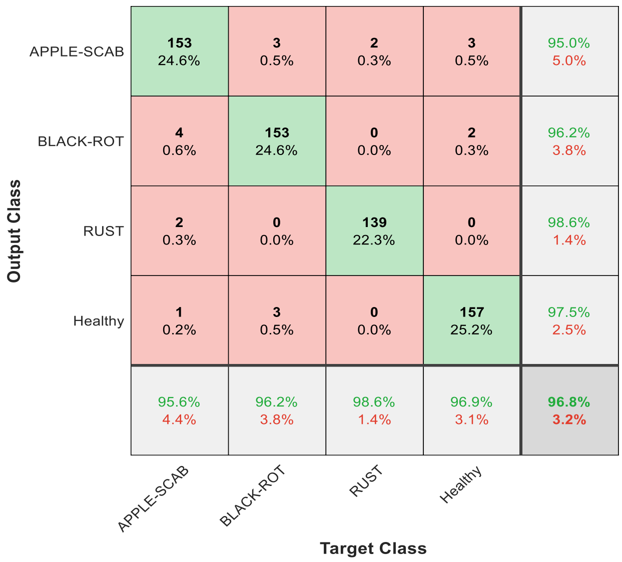


Figure. 3 Confusion matrix for the proposed apple tree leaf disease detection using HOG features

The HOG features detection system generated this outcome according to Figure 3. This table demonstrates how the classifier identifies leaf conditions correctly thanks to the structural pattern detection capabilities of HOG. HOG features achieve high levels of sensitivity and specificity metrics by detecting small leaf texture changes linked to diseases which makes them perform highly for precise disease detection.

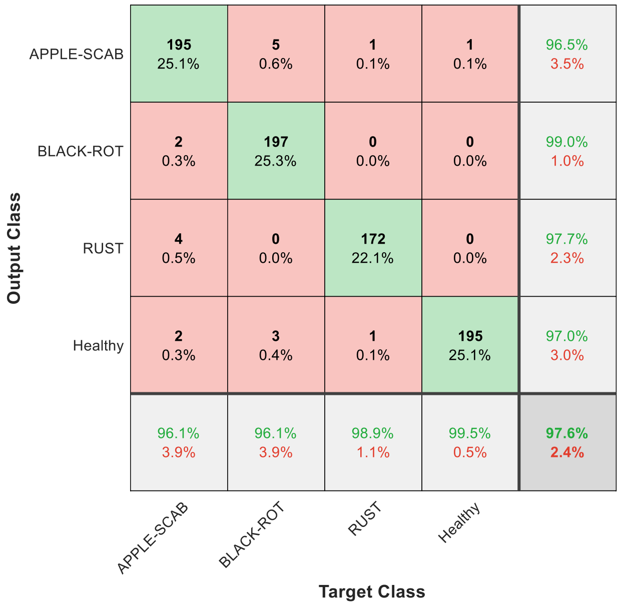


Figure. 4 Confusion matrix for the hybrid features based proposed apple tree leaf disease detection model without feature selection with 80% training data

Figure 4 illustrates the confusion matrix obtained from evaluating the apple tree leaf disease detection model that incorporates HOG and MSLBP features without including feature selection. The performance evaluation along with the matrix shows how hybrid features optimally boost the detection system accuracy and depends without feature selection to demonstrate the inherent value of combining multiple textural and structural descriptors in disease diagnostic applications.

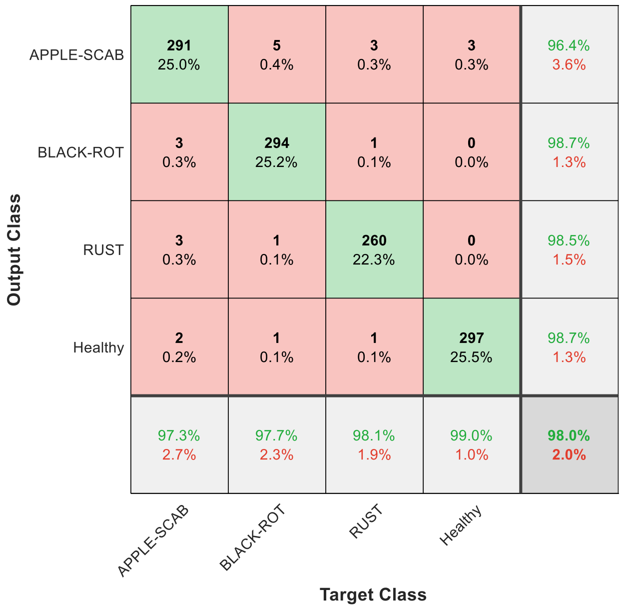


Figure. 5Confusion matrix for the hybrid features based proposed apple tree leaf disease detection with hybrid features and MRMR based feature selection with 90% training data

Figure 5 presents a confusion matrix of the system that includes MRMR-based feature selection on hybrid features. The selected training data consists of 90% while being optimized for a relevant and redundant-free feature set. A significant performance improvement became evident in model diagnostic capabilities after the selection refinement through MRMR which produced better evaluation metrics consisting of decreased false positives and enhanced F1 scores.

Table 3. Performance Evaluation of the Proposed Approach

| **Metric** | **MSLBP Features** | **HOG Features** | **Hybrid Features Without Feature Selection** | **Hybrid Features with MRMR Feature Selection** |
|---|---|---|---|---|
| Accuracy | 95.37% | 96.78% | 97.56% | 98.03% |
| Error | 4.63% | 3.22% | 2.44% | 1.97% |
| Sensitivity | 95.44% | 96.84% | 97.62% | 98.03% |
| Specificity | 98.47% | 98.92% | 99.19% | 99.34% |
| Precision | 95.48% | 96.84% | 97.57% | 98.04% |
| False Positive Rate | 1.53% | 1.08% | 0.81% | 0.66% |
| F1 Score | 95.37% | 96.84% | 97.59% | 98.03% |
| Matthews Correlation Coefficient | 93.91% | 95.76% | 96.78% | 97.38% |
| Kappa | 87.66% | 91.43% | 93.49% | 94.74% |

The performance evaluation of the proposed apple tree leaf disease detection system appears in Table 3 through its analysis of different feature extraction and selection methods. The table compares four configurations: MSLBP features, HOG features, hybrid features without feature selection, and hybrid features with MRMR-based feature selection. The table demonstrates model performance advancement through every sophisticated method applied. The initial MSLBP feature framework provides 95.37% accuracy and 87.66% kappa coefficient but the advanced hybrid features along with MRMR feature selection achieves 98.03% accuracy and 94.74% kappa coefficient. The implementation of MRMR feature selection methods alongside optimized process integration achieves better results because it reduces error rates to 1.97% and raises specificity to 99.34%. The tested characterization metrics confirm how the disease detection system effectively measures precise and efficient disease recognition features across different operational conditions.

Table 4. Comparison of the proposed approach with previous research works

| **Method** | **Dataset used** | **Accuracy (%)** |
|---|---|---|
| Khan et al. 2022 [1] | Manual Dataset | 87.9% |
| Acharya et al. 2024 [2] | Kaggle Dataset | 91.37% |
| Yadav et al. 2023 [11] | Kaggle Dataset | 98% |
| Hou et al. 2023 [14] | Kaggle Dataset | 93.68% |
| Khan et al. 2021 [15] | SKUAST-K Dataset | 97.18% |
| **Proposed Work (Without Feature Selection)** | Kaggle Dataset | **97.56%** |
| **Proposed Work (MRMR-based Feature Selection)** | Kaggle Dataset | **98.03%** |

Table 4 demonstrates the proposed method for detecting apple tree leaf diseases provides superior accuracy than other noteworthy research works while showcasing its methodological improvements. The model's assessment across different datasets reveals its performance level through measurements of precision which include results with feature selection methods implemented or not. Without applying feature selection to the model for apple tree leaf disease detection the proposed approach delivers 97.56% accuracy on the Kaggle dataset above Khan et al. 2021 [15] (97.18%) and Hou et al. 2023 [14] (93.68%) who also used the Kaggle dataset for their work. Our system reaches 98.03% accuracy through the inclusion of MRMR-based feature selection for data assessment which yields slightly higher results than Yadav et al. 2023 [11] who displayed 98% accuracy on the exact dataset. Our methodology advances the field by incorporating MRMR feature selection that achieves state-of-the-art detection performance standards which makes it appropriate for agricultural field applications.

## CONCLUSION

A complex diagnostic procedure for apple leaf diseases has been developed by using HOG and MSLBP features with MRMR selection followed by XGBoost classification. Advanced techniques unify to make the model proficient in detecting texture-based and frequency-based abnormalities in leaf structures which leads to better disease classification across multiple leaf conditions. A performance increase from 95.37% to 98.2% occurred when the model received MRMR-based feature selection implementation. The XGBoost classifier displays high robustness which stops it from becoming overfitted while operating under different environmental situations. Research findings prove that the developed apple tree disease detection system creates a precise and practical solution which can be deployed for agricultural use to provide reliable disease detection for early intervention in management practices. Future developments will optimize this system to function in real-time operations while extending it to other plant disease classifications to advance agricultural technology.

## REFRENCES


[1] Khan, A.I., Quadri, S.M.K., Banday, S. and Shah, J.L., 2022. Deep diagnosis: A real-time apple leaf disease detection system based on deep learning. computers and Electronics in Agriculture, 198, p.107093.

[2] Acharya, V. and Ravi, V., 2024. Apple foliar leaf disease detection through improved capsule neural network architecture. Multimedia Tools and Applications, 83(16), pp.48585-48605.

[3] Sarkar, C., Gupta, D., Gupta, U. and Hazarika, B.B., 2023. Leaf disease detection using machine learning and deep learning: Review and challenges. Applied Soft Computing, 145, p.110534.

[4] Dey, A.K. and Sharma, A., 2023. Plant Feature Extraction for Disease Classification. In Recent Trends and Best Practices in Industry 4.0 (pp. 183-201). River Publishers.

[5] Iswanto, B.H. and Sugihartono, I., 2021, October. Feature Extraction of Tea Leaf Images using Dual-Tree Complex Wavelet Transform and Gray Level Co-occurrence Matrix. In Journal of Physics: Conference Series (Vol. 2019, No. 1, p. 012092). IOP Publishing.

[6] Mathew, D., Kumar, C.S. and Cherian, K.A., 2021. Foliar fungal disease classification in banana plants using elliptical local binary pattern on multiresolution dual tree complex wavelet transform domain. Information processing in Agriculture, 8(4), pp.581-592.

[7] Hashemi, A., Dowlatshahi, M.B. and Nazamabadi-pour, H., 2023. Minimum redundancy maximum relevance ensemble feature selection: A bi-objective Pareto-based approach. Journal of Soft Computing and Information Technology (JSCIT), Vol, 12(1).

[8] Xie, S., Zhang, Y., Lv, D., Chen, X., Lu, J. and Liu, J., 2023. A new improved maximal relevance and minimal redundancy method based on feature subset. The Journal of supercomputing, 79(3), pp.3157-3180.

[9] Hasan, S., Jahan, S. and Islam, M.I., 2022. Disease detection of apple leaf with combination of color segmentation and modified DWT. Journal of King Saud University-Computer and Information Sciences, 34(9), pp.7212-7224.

[10] Yao, L., Yuan, P., Tsai, C.Y., Zhang, T., Lu, Y. and Ding, S., 2023. ESO: An enhanced snake optimizer for real-world engineering problems. Expert Systems with Applications, 230, p.120594.

[11] Yadav, D. and Yadav, A.K., 2023. A Novel Convolutional Neural Network Based Model for Recognition and Classification of Apple Leaf Diseases. Traitement du Signal, 37(6).

[12] Alqethami, S., Almtanni, B., Alzhrani, W. and Alghamdi, M., 2022. Disease detection in apple leaves using image processing techniques. Engineering, Technology & Applied Science Research, 12(2), pp.8335-8341.

[13] Baranwal, S., Khandelwal, S. and Arora, A., 2019, February. Deep learning convolutional neural network for apple leaves disease detection. In Proceedings of international conference on sustainable computing in science, technology and management (SUSCOM), Amity University Rajasthan, Jaipur-India.

[14] Hou, J., Yang, C., He, Y. and Hou, B., 2023. Detecting diseases in apple tree leaves using FPN–ISResNet–Faster RCNN. European Journal of Remote Sensing, 56(1), p.2186955.

[15] Khan, A.I., Quadri, S.M.K. and Banday, S., 2021. Deep learning for apple diseases: classification and identification. International Journal of Computational Intelligence Studies, 10(1), pp.1-12.

[16] Li, X. and Rai, L., 2020, November. Apple leaf disease identification and classification using resnet models. In 2020 IEEE 3rd International Conference on Electronic Information and Communication Technology (ICEICT) (pp. 738-742). IEEE.

[17] Apple Leaf Diseases Dataset. Online available at: https://www.kaggle.com/competitions/plant-pathology-2021-fgvc8/data