Maize Plant Leaf Disease Classification Using Supervised Machine Learning Algorithms

Ashish Patel[1], Richa Mishra [1], Aditi Sharma[1,2*]

[1] Department of Computer Science and Engineering,
Parul University, Vadodara, Gujarat, India.
[2] IEEE Senior Member, Astana IT University, Astana, Kazakhstan
Emails: ashish.patel28275@paruluniversity.ac.in; richa.mishra21150@paruluniversity.ac.in;
aditi.sharma@ieee.org

Abstract

Maize is an important staple crop all over the world, and its health is very important for food security. It is important for crop management and yield to find diseases that affect maize plants as soon as possible. In this study, we suggest a new way to classify diseases on maize plant leaves by using supervised machine learning algorithms. Our method uses the power of texture analysis with Gray-Level Co-occurrence Matrix (GLCM) and Gabor feature extraction techniques on the Plant-Village dataset, which has images of both healthy and unhealthy maize leaves. This method uses four supervised machine learning algorithms, called Decision Tree, Gradient Boosting, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN), to sort the extracted features into healthy and diseased groups. By doing a lot of tests, we show that our way of finding maize leaf diseases works well. The results show that these techniques have the potential to quickly and non-invasively diagnose diseases, giving farmers important information for acting quickly. We talk about the pros and cons of each algorithm and suggest ways to make them even better. This research contributes to the advancement of automated plant disease detection systems, fostering sustainable agriculture practices and aiding in crop management decisions. The proposed approach holds promise for real-world application, enabling farmers to mitigate disease-related losses and secure global food supplies.

Keywords: Plant leaf disease detection; Machine learning; Decision tree; Gradient boosting

1. Introduction

Agriculture forms the bedrock of global food security, with maize (Zea mays) being one of its fundamental staples. The cultivation of maize spans continents, sustaining billions of people and serving as a crucial source of nourishment and livelihood. However, the proliferation of diseases among maize crops poses a significant threat to agricultural productivity and food sustainability. The timely identification and management of these diseases are paramount to ensure healthy crop yields and secure food supplies for a burgeoning global population.

Over 70% of the workforce in the developed nation of India works in agriculture. Farmers have many alternatives when choosing the finest plants, crops, and pesticides. Plant disease diagnosis is a difficult task that must be completed as soon as possible. First, a field expert manually observed and examined plant illness. This requires considerable effort and processing time. Visually assessing plant diseases is an inherently subjectivist activity that is subject to psychological and cognitive processes that can lead to bias, optical illusions, and, ultimately, error[1]. Despite the remarkable ability of human vision and cognition to detect and interpret patterns, this activity is subject to psychological and cognitive processes that can lead to error.

The most common method for identifying plant diseases is skilled naked-eye inspection]. However, this necessitates ongoing expert surveillance, which may be unaffordable on large farms. Because it can assist in monitoring vast fields of crops and, as a result, early detection of plant disease saves the crop from more damage. Hence, automatic plant disease detection is a crucial research topic. Therefore, researchers are searching for a quick, automated, less costly, and precise method to diagnose disease by computing leaf area using pixel number

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023

data [3]. Monitoring leaf area helps with the physiological research of transpiration, photosynthesis, and other aspects of plant growth. Additionally, it assists in quantifying the strain put on the environment and water, the need for fertiliser for maintenance and treatment, and the harm caused by leaf diseases and pastes.

According to statistics compiled by the FAO from the world's agricultural organisations in 2014, the agricultural sector in India is the leading producer of a wide range of fresh fruits and vegetables. In 2010, India was responsible for producing more than 80 percent of all agricultural goods globally. These items included major commodities such as cotton and coffee. In terms of agricultural output, India was consistently ranked among the top five countries worldwide. [4] India was one of the top five producers of meat from animals and poultry in the world in 2011, having seen one of the highest growth rates in the industry.

Maize, like many other crops, can get disease from fungi, bacteria, and viruses, which are called pathogens. Symptoms of these diseases include discoloured leaves, necrosis, wilting, and deformities. All of these things can be very bad for the maize plant's overall health and yield. Traditional ways to keep track of diseases and figure out what's wrong involve a lot of manual inspection by agricultural experts, which is hard work, takes a long time, and leaves room for mistakes. Because of this, we need new and effective ways to find diseases right away to deal with these problems.

Recent improvements in computer vision and machine learning have made it possible to think of exciting ways to automate the process of finding diseases in agricultural settings. Researchers have made systems that use these technologies to quickly analyse and classify images of plants to look for signs of diseases. This paper shows a new way to classify maize plant leaf diseases by using supervised machine learning algorithms.

The primary objective of this study is to harness the potential of machine learning, specifically supervised learning, for the automated classification of maize leaves as healthy or diseased. To achieve this, we employ advanced feature extraction techniques, including Gray-Level Co-occurrence Matrix (GLCM) and Gabor filters, which enable the capture of essential textural information from leaf images. Our main source of data is the PlantVillage dataset, which has a wide range of high-resolution images that show the intricate details of maize leaf diseases.

This study compares four well-known supervised machine learning algorithms: Decision Tree, Gradient Boosting, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). Each algorithm has to figure out complex patterns and relationships in the extracted features in order to correctly classify maize leaves as healthy or diseased. By comparing how well these algorithms work, we hope to find the best way to automate the detection of maize diseases, which will help with sustainable farming practises and world food security.

This paper addresses the pressing need for reliable and efficient methods of maize disease detection. Our approach showcases the potential of combining advanced machine learning techniques with texture-based feature extraction from leaf images. The outcomes of this research hold the promise of transforming agricultural practices by enabling timely disease diagnosis, ultimately bolstering crop management and food production for a growing world population.

2. Related Work

Savita N. Ghaiwat et al. [5] describe the research of several classification methods for categorizing plant leaf diseases. In the current test setting and class prediction, the k-nearest-neighbors approach seems appropriate. It is challenging to use SVM for classification if the training data cannot be separated linearly.

Sanjay B. Dhaygude et al. [6] present the leaf segmentation approach in which the RGB image is converted into HIS color format. This approach has four steps. The RGB is first transformed into HIS format. The second stage is masking green pixels using a threshold value. The segmented picture is retrieved in the third stage, once the original image has been given a mask. The fourth and last main stage is when segmentation is finished.

Mrunalini R. Badnakhe et al. describe a method for categorising and identifying different plant diseases in their article [7]. A machine learning-based identification system will benefit the Indian economy, saving time, money, and effort. The article in question uses the colour co-occurrence approach to extract feature sets. To automatically find disease in leaves, neural networks are used. While needing less computing labour, the suggested method can significantly help in accurately detecting the leaf and seems to be a crucial step in the event of steam and root infections.

According to S. Arivazhagan et al. [8], the disease detection process consists of the four most important steps. The thresholding segmentation technique is used to mask the segmented green pixels once the input image has been converted to RGB format. The features gathered to categorize the disease are then sent to the classifier. The proposed methodology correctly diagnoses and categorises the ailments with a 94 percent accuracy rate, proving

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023

its effectiveness. Experimental evidence of the suggested method's robustness is provided using information from a database of about 500 plant leaves..

Anand H. Kulkarni et al. [9] propose artificial neural networks (ANN) and a range of image processing techniques to enable early and precise identification of plant diseases. This system's accuracy can be increased to 91% by utilising the Gabor feature extraction approach and the ANN classifier.

K-mean clustering, texture analysis, and colour analysis are the three methods that Sabah Bashir and colleagues [10] use to identify Malus domestica disease. It employs texture and colour features in healthy and diseased areas to detect and differentiate between leaf features. Future applications might use data categorization techniques like Bayes, principal components, and K-means clustering.

Histogram matching is a technique that can be utilised, as stated by Smita Naikwadi et al. [11], in order to identify plant diseases. Due to the fact that the disease primarily manifests itself in the leaves of plants, histogram matching is predicated on the identification of edges and colour features. The layers of an RGB picture are separated into red, green, and blue channels using the training approach, which combines an edge detection algorithm with a layers separation methodology. The colour co-occurrence texture analysis method is built on top of the spatial grey level dependence matrices as its underlying structure.

The fundamental threshold technique and the Triangle threshold approach are provided by Sanjay B. Patil et al. [12]. These methods enable the segmentation of the leaf area and the lesion region. The next step is to classify the illness based on the proportion of the leaf and lesion areas. The technique for diagnosing leaf illness is said to be quick and accurate, and leaf area is determined using threshold segmentation.

In order to distinguish disease spots on plant leaves using image processing methods, an algorithm is applied, according to Piyush Chaudhary et al. [13]. The methods for identifying disease spots utilising the HSI, CIELAB, and YCbCr colour spaces are contrasted in this study. The image is made easier with the Median filter. In the last stage, the threshold for identifying the hazardous site may be determined using the Otsu approach on the colour component.

Image processing techniques were utilised in Arti N. Rathod et al. [14] to conduct an analysis of a variety of methodologies for identifying leaf diseases. The research of the present techniques attempts to improve throughput and decrease subjectivities that arise from direct visual observation to detect and diagnose plant diseases.

Utilizing machine learning methods, Himanshu Das et al. [15] demonstrate the detection and categorization of maize leaf disease. For disease diagnosis in maize plants using plant photos, this method employs machine learning techniques such as Naive Bayes (NB), Decision Trees (DT), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Random Forests (RF). The RF algorithm gets the highest level of accuracy, 79.23%, when compared to other classification methods.

3. Proposed system

Fig. 1 depicts the system's block diagram

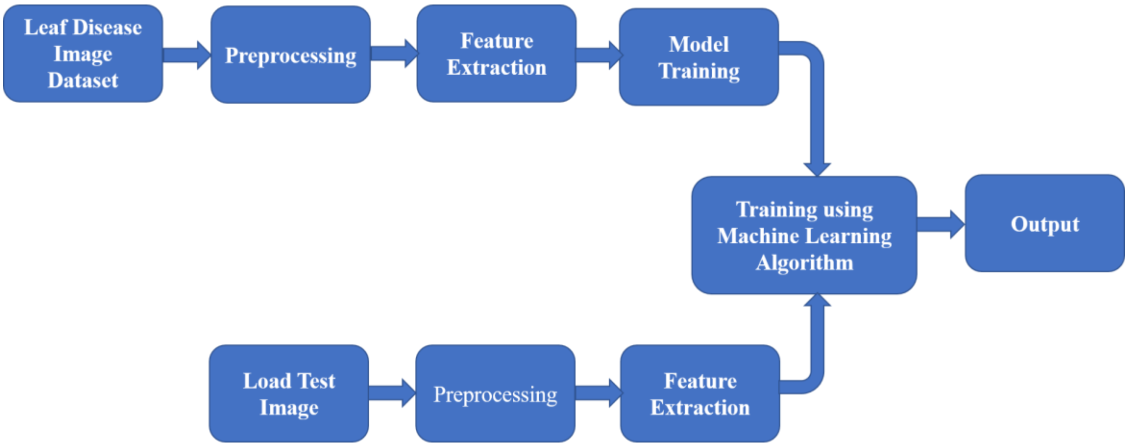


Figure 1: Block diagram of the Plant Leaf Disease Detection

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023

A. Leaf Disease Image Database

Images from the PlantVillage dataset of healthy and unhealthy apple, grape, tomato, and corn-leaf specimens are included in the collection [16]. Data is used for testing 20% of the time and training 80% of the time. After that, the image is scaled down to 256 by 256 pixels. Because it is responsible for determining both the efficacy of the classifier and the operation of the solution that is being provided, this database is designed with extreme care. The breakdown of the distribution of training and testing datasets may be found in Table 1.

Table 1: Dataset distribution

Dataset Plants	Total Images	Training Images	Testing Images
Corn(maize) Grayleafspot	*513*	*411*	*102*
Corn(maize) Commonrust	*1192*	*954*	*238*
Corn(maize) healthy	*1157*	*925*	*232*

B. Image Pre-processing

Prior to future processing and analysis, image pre-processing aims to improve image quality. The given images are initially in RGB format but are then transformed to grayscale. The resulting images exhibit some noise. Color transformation is employed to ascertain an image's color and luminance characteristics. To enhance image quality, a median filter is applied.

C. Feature Extraction

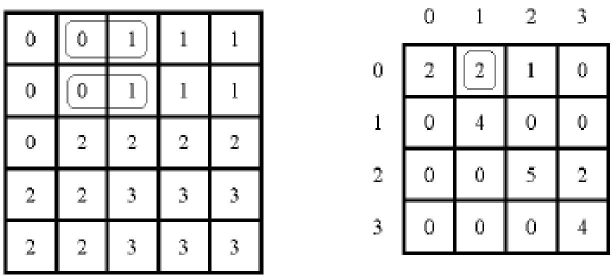
The feature assesses some of the essential characteristics of the object. It is calculated as a consequence of one or more measurements, each of which identifies a measurable attribute of an item. Both low-level and high-level qualities can be utilised to categorise all features. It is still possible to extract low-level features directly from the photos that are being entered, even though the extraction of low-level features is required for the extraction of high-level features. The texture of the surface can be used as a barometer to rate its quality. The limits of a neighbourhood are set by the distribution of shades of grey across a certain region. The texture of an image is affected by the size or resolution at which it is displayed. A texture having different characteristics on a small scale may turn into a uniform texture when shown at a larger size [17].

a) Texture feature

The surface's quality may be gauged by its texture. A neighbourhood is defined by the distribution of grey levels in its geographic area. Since texture discloses its features through pixel positions and values, several methods exist for categorizing textures. The size of an image as well as the resolution at which it is displayed both have an effect on the image's texture. It's possible that a texture that has different features on a small size can become homogenous when viewed on a larger scale [18].

In statistical texture analysis, pixel intensity distribution at a specific location represents texture properties. It generates first-, second-, and higher-order statistics by basing each combination on the number of pixels or dots that make up that combination. An image may be analysed as a texture using second-order statistics for extracting features based on GLCM [19].

The frequency of a specific combination of pixel brightness values appearing in a picture is displayed in the GLCM table. The GLCM of a four-level image is created at a distance of 1 and a direction of 0°, as illustrated in Fig. 2.



0	0	1	1	1
0	0	1	1	1
0	2	2	2	2
2	2	3	3	3
2	2	3	3	3

	0	1	2	3
0	2	2	1	0
1	0	4	0	0
2	0	0	5	2
3	0	0	0	4

(a) (b)

Figure 2: Example of (a) image with 4 grey level image (b) GLCM for distance 1 and direction 0°.

The image's statistical data are referred to as features. GLCM is a technique for extracting distinct characteristics from grayscale and binary images. The following GLCM characteristics are retrieved using the suggested method.

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023

- **Contrast:** The local differences in the grey level co-occurrence matrix are measured using contrast.

$$Contrast = \sum_{i,j} |i-j|^2 p(i,j) \qquad (1)$$

- **Homogeneity:** The degree to which the element distribution in GLCM is aligned with the diagonals of GLCM is one way to measure homogeneity.

$$Homogeneity = \sum_{i,j} \frac{1}{1+(i-j)^2} p(i,j) \qquad (2)$$

- **Energy:** It measures the uniformity among the pixels.

$$Energy = \sum_{i,j} p(i,j)^2 \qquad (3)$$

- **Entropy:** It measures the statistical measurement of the randomness of each pixel.

$$Entropy = -\sum_{i,j} p(i,j)\, log\big(p(i,j)\big) \qquad (4)$$

- **Dissimilarity:** Dissimilarity is a metric that describes how different grey level pairings in an image vary.

$$Dissimilarity = \sum_{i,j} |i-j| p(i,j) \qquad (5)$$

Where, $p(i,j)$ = image pixel to be processed

b) Discrete Wavelet Transform

The presented technique extracts texture characteristics from the region of interest-containing picture segment using 2-D discrete wavelet transformations. Discrete wavelet transformations account for a signal's frequency and any unique information. For accurate texture feature extraction, this attribute is helpful. A signal's information may be stored via 2-D DWT using fewer coefficients. DWT maintains the image as a two-dimensional signal with columns and rows when processing images. Wavelet transformations investigate an image's more minute characteristics, including its horizontal, vertical, and diagonal subbands. Fig. 3 shows how 2-D DWT is used for image decomposition.

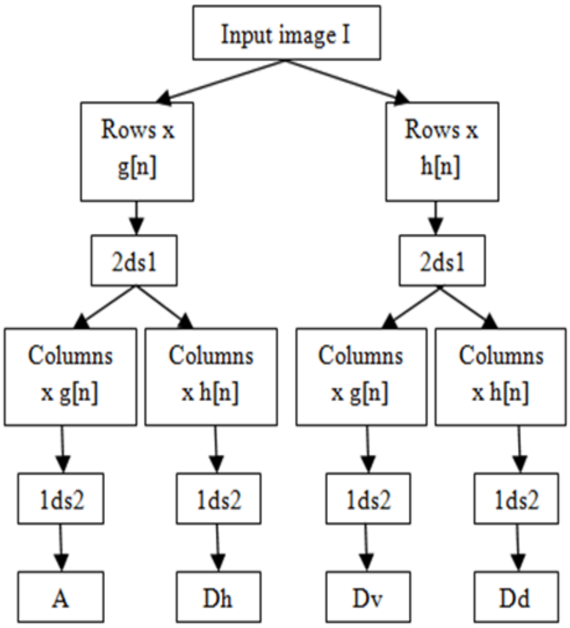


Figure 3: 2-D DWT Decomposition

Fig. 3 displays the input image, a high pass filter (h[n]), and a low pass filter (g[n]). When rows are downsampled by 1 and columns by 2, rows are downsampled by 2 and columns by 1. (or 2ds1). This work's key findings are

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023

immediately at level one. An approximation coefficient is represented by A. Dh, a vertical coefficient by Dv, and a diagonal coefficient by Dd denotes a horizontal coefficient.

c) Gabor filter

A complex Gabor results from multiplying a complex sinusoid by a Gaussian kernel. The complex sinusoid is referred to as the carrier, whereas the Gaussian function is referred to as the envelope. The diagram that follows depicts a two-dimensional Gaussian curve that has dispersion in both the x and y dimensions simultaneously.:

$$g(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \qquad (6)$$

The complex sinusoid is defined as follows, where the symbols μ, θ *and* φ stand for the spatial frequency, direction, and phase shift ($j = \sqrt{-1}$), respectively.

$$s(x, y, \mu, \theta, \varphi) = exp\{j2\pi(x.\mu cos\theta + ysin\theta) + \varphi\} \qquad (7)$$

Thus, the following representation of the complex Gabor function can be;

$$h(x, y, \mu, \theta, \varphi) = g(x, y, \sigma)s(x, y, \mu, \theta, \varphi) \qquad (8)$$

D. Classification

Images of maize leaves are classified in this study using machine learning. The classifiers are then used on the testing set after being trained on the training set. After that, the performance is evaluated by comparing the predicted labels to the actual labels, and a judgement is made based on the results of that comparison. This method trains and tests leaf images for categorizing healthy and relevant diseases using SVM, KNN, DT, and GB algorithms.

a) Support vector machine

In a high-dimensional feature space that might be used for classification, the ideal use of an SVM is to find the hyperplane between two unique classes. There is the use of supervised machine learning methods such as SVM [13]. The two stages of the supervised learning approach are training and testing.

The SVM classifier uses a linear function to categorize the image as

$$f(x) = W^T X + b \qquad (9)$$

Where X is the input training samples, W is the weight assigned to the network, b is bias or offset

In order to distinguish between linear and non-linear categories, SVM was employed for classification. To convert the input pattern into a higher dimensional feature space, the non-linear classifier is abandoned in favour of the linear SVM classifier. Hyperplanes may be used to study data that can be separated linearly. In contrast, kernel functions like higher-order polynomials can be used to investigate data that cannot be separated linearly carefully. The SVM classification technique uses a variety of kernel strategies, including the Radial Basic Function (RBF), Linear Kernel Function, and Quadratic Kernel Function. Two samples, x and x', are applied as feature vectors in some input space using the RBF kernel.

$$K(x, x') = \exp\left(\frac{||x-x'\ \ ||^2}{2\sigma^2}\right) \qquad (10)$$

As the distance increases, the kernel function's value drops and fluctuates between zero (at the limit) and one (when x = x').

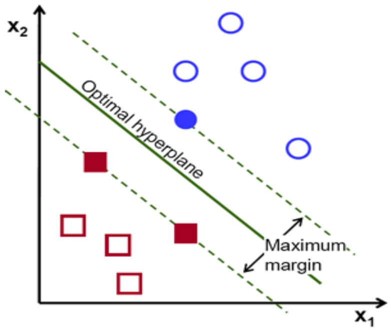


Figure 4: Optimal hyperplane margin

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023

b) K-nearest Neighbor

KNN is an easy-to-use and trustworthy classification technique. This classifier categorizes the testing feature vector by identifying the k-nearest training neighbor vectors. Different distance measuring methods, such as Euclidean, city-block, Chebychev, etc., are used to determine the separation between the training and testing vectors. This approach separates the training and testing data vectors by Euclidean, cityblock, cosine, and correlation distances. Euclidean distance is given by the

$$d(a,b) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2} \tag{11}$$

As a result of a feature being extracted from a training and testing image set, various dimensions in space and the values of the retrieved features are then used to calculate the coordinates of a characteristic in each dimension, resulting in a set of points in the space. Then, by employing the appropriate metric, we may relate the similarity of two distinct points to their spatial separation.

The k nearest data points in the training set are picked. The most common class is used to determine whether points are substantially similar to the point being investigated when deciding which class to predict for a new observation. The KNN algorithm worked in this manner.

The following is the KNN algorithm:

- The new sample is defined together with a positive integer value of k.
- We chose the k database records most similar to the new testing sample.
- We determine which category best describes these entries.
- Using the value of k, we categorize the new sample as follows.
- If the outcome is unsatisfactory, adjust k's value until a good degree of accuracy is reached.

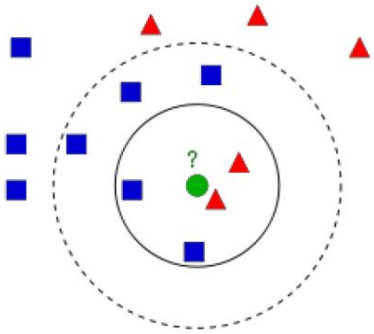


Figure 5: Classification of query image through KNN classifier

Two classes are depicted in Fig. 3 by Blue Squares and Red Triangles. The feature set now includes the new testing functionality shown in green. Using KNN, the testing feature was categorized into the appropriate class by applying the label of more excellent majority neighbors.

c) Decision Tree Algorithm

A DT is a key framework for scenario classification. Because data is continually divided based on a parameter, it is comparable to supervised machine learning. A DT is a tree structure with nodes such as the root, middle, and leaf nodes. Both classification and regression are performed using a DT.

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023

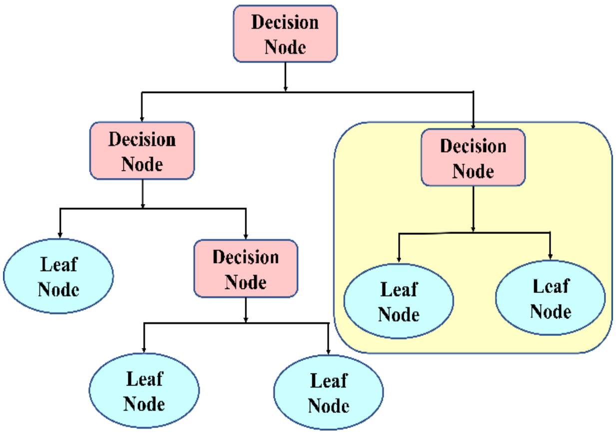


Figure 6: Decision Tree

The decision Tree (DT) algorithm is given below.

Input: S, where S= set of classification instances

Output: Decision Tree

```
Procedure Build Tree
    repeat
        max gain ← 0
        Split A ← null
        e ← Entropy (Attributes)
        For all Attributes α in S, do
            gain← InformationGain(a, e)
            if gain>maxGain, then
                max gain ← gain
                splitA← α
            end if
        end for
        Partition (S, splitA)
    Until all partitions are processed
end procedure
```

d) Gradient Boosting (GB) Algorithm

GB is a classification and regression machine learning approach that creates a prediction model using an ensemble of generally binary trees [20]. As shown in Fig. 4, the GB method is the central component of the XGBoost ensemble method, which is a method that integrates a variety of key machine learning algorithms into a more comprehensive framework.

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023

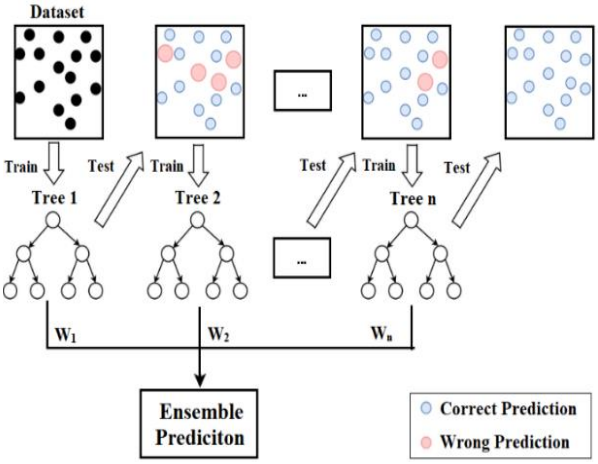


Figure 7: Flow diagram of GB machine learning method

The GB Algorithm may classify data in binary or many classes. The recommended strategy offers more accuracy than current classification methods [21-26]. When the findings were evaluated, the recommended strategy had a greater accuracy of 80.02%. Either feature engineering or the use of the boosting method right away can increase the prediction model's accuracy.

```
Input: Training Images

Initialize $f_0(x) argmin_p \sum_{i=1}^{N} L(y_i, p)$
For $m=1$ to $M$, do
    Step1: Compute the negative gradient
        $y_i = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F_{x_i}}\right]$
    Step 2: Fit the model
        $\alpha_m = argmin_{\alpha,\beta} \sum_{i=1}^{N} [\bar{y} - \beta h(x_i; \alpha_m)]^2$
    Step 3: Choose a gradient descent step size as
        $\rho_m = argmin_p \sum_{i=1}^{N} L(y_i, F_m - 1(x_i) + \rho h((x_i; \alpha_m)))$
    Step 4: Update the estimation of $F(x)$
        $F_m(x) = F_{m-1}(x) + \rho_m h(x, \alpha_m)$
end for
Output: the final regression function $F_m(x)$
```

4. Results

Python is utilized in the construction of the proposed system, with OpenCV and Scikit-Learn being utilized for image processing and classification, respectively. Windows 10 was used for the training and testing of this system, and it comes equipped with an Intel i5 processor and 8 gigabytes of RAM. Assessments both qualitative and quantitative are used to determine how well the system is working. Classification accuracy is used in quantitative analysis. Table II displays the findings from a qualitative study utilizing several machine learning classifiers.

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023

Table 2: Performance of the proposed system For GLCM feature extraction technique with 0°

	Precision	Recall	F-measure	Accuracy
SVM_ Rbf	0.7239	0.6914	0.6805	0.6914
SVM_Poly	0.5457	0.4381	0.3628	0.4381
SVM_sigmoid	0.5046	0.50073	0.49678	0.5007
KNN (k=3)	0.6544	0.6492	0.6511	0.6492
KNN (k=5)	0.6821	0.6798	0.6795	0.6797
KNN (k=7)	0.7023	0.7001	0.6976	0.7001
DT	0.9147	0.9141	0.9144	0.9141
GB	0.9426	0.9432	0.9426	0.9432

From Table 2 it is observed that both Decision Trees (DT) and Gradient Boosting (GB) classifiers, showing the highest values across all metrics. GB, in particular, emerges as the top performer with the highest Precision (0.9426), Recall (0.9432), F-measure (0.9426), and Accuracy (0.9432), highlighting its suitability for this specific maize leaf classification task. This analysis indicates that the Decision Tree and Gradient Boosting classifiers, when combined with the GLCM feature extraction technique at a 0-degree orientation angle, outperform other models for the classification of maize leaves as healthy or diseased.

Table 3: Performance of the proposed system For GLCM feature extraction technique with 45°

	Precision	Recall	F-measure	Accuracy
SVM_ Rbf	0.6302	0.6317	0.5989	0.6317
SVM_Poly	0.597	0.5793	0.5304	0.5793
SVM_sigmoid	0.4688	0.4614	0.4583	0.4614
KNN (k=3)	0.6361	0.6244	0.6292	0.6244
KNN (k=5)	0.6425	0.6375	0.6389	0.6375
KNN (k=7)	0.6549	0.6579	0.6546	0.6579
DT	0.9071	0.9068	0.9069	0.9068
GB	0.9487	0.949	0.9484	0.949

Table 3 presents a comprehensive analysis of the performance of various machine learning classifiers when applied to the task of classifying healthy and diseased maize leaves using the GLCM feature extraction technique at a 45-degree orientation angle. From Table II, we can see that both Decision Trees (DT) and Gradient Boosting (GB) classifiers have the highest values for all metrics. With the highest Precision (0.9426), Recall (0.9432), F-measure (0.9426), and Accuracy (0.9432), GB stands out as the best performer. This shows that it is a good choice for classifying maize leaves. This analysis shows that the Decision Tree and Gradient Boosting classifiers work better than other models for figuring out whether maize leaves are healthy or sick when used with the GLCM feature extraction method at an orientation angle of 45 degrees.

Table 4: Performance of the proposed system For GLCM feature extraction technique with 90°

	Precision	Recall	F-measure	Accuracy
SVM_ Rbf	0.7129	0.7016	0.6847	0.7016
SVM_Poly	0.6577	0.5371	0.5067	0.5371
SVM_sigmoid	0.5125	0.4963	0.4953	0.4963
KNN (k=3)	0.7151	0.7176	0.7156	0.7176
KNN (k=5)	0.7137	0.7176	0.7141	0.7176
KNN (k=7)	0.7148	0.7191	0.7145	0.7191
DT	0.9155	0.9127	0.9136	0.9127
GB	0.9501	0.9505	0.9502	0.9505

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023

Table 4 gives a detailed look at how well different machine learning classifiers worked when they were used to tell the difference between healthy and sick maize leaves using the GLCM feature extraction technique at a 90-degree orientation angle. DT and GB classifiers perform better than other models. GB being the best with the highest Precision, recall, F-measure and accuracy of 0.9501, 0.9505, 0.9502, and 0.9501 respectively. This analysis shows that GB and DT are the best choices for classifying maize leaf diseases at 90-degree angle, while the performance of SVM depends on the type of kernel and KNN is strong no matter what k value use.

Table 5: Performance of the proposed system For GLCM feature extraction technique with 135°

	Precision	Recall	F-measure	Accuracy
SVM_ Rbf	0.7129	0.7016	0.6847	0.7016
SVM_Poly	0.6577	0.5371	0.5067	0.5371
SVM_sigmoid	0.5125	0.4963	0.4953	0.4964
KNN (k=3)	0.7151	0.7176	0.7155	0.7176
KNN (k=5)	0.7136	0.7176	0.714	0.7176
KNN (k=7)	0.7148	0.719	0.7145	0.719
DT	0.9069	0.9039	0.9051	0.9039
GB	0.9441	0.9446	0.9443	0.9446

Table 5 gives a detailed analysis of how well different machine learning classifiers work at separating healthy maize leaves from diseased ones using the GLCM feature extraction technique at a 135-degree orientation angle. Both DT and GB classifiers perform better, but GB is the best because it has the highest Precision, recall, F1 measure and accuracy of 0.9441, 0.9446, 0.9443, and 0.9443 respectively. According to this investigation, GB and DT are the best options for this specific maize leaf disease classification task when oriented at a 135-degree angle, whereas the Rbf kernel for SVM and KNN remains effective over a range of k values.

Table 6: Performance of the proposed system For the DWT feature extraction technique

	Precision	Recall	F-measure	Accuracy
SVM_ Rbf	0.7309	0.7751	0.7031	0.7751
SVM_Poly	0.151	0.3886	0.2175	0.3886
SVM_sigmoid	0.6149	0.596	0.5549	0.596
KNN (k=3)	0.151	0.3886	0.2175	0.3886
KNN (k=5)	0.151	0.3886	0.2175	0.3886
KNN (k=7)	0.151	0.3886	0.2175	0.3886
DT	0.7559	0.7663	0.7604	0.7663
GB	0.9172	0.9104	0.9036	0.9104

Table 6 gives an in-depth look at how well different machine learning classifiers worked when they were used to tell the difference between healthy and sick maize leaves using the Discrete Wavelet Transform (DWT) feature extraction method. Both Decision Trees (DT) and Gradient Boosting (GB) work well, but GB is the best with high Precision (0.9172), Recall (0.9104), F-measure (0.9036), and Accuracy (0.9104). This analysis shows that GB, along with DT and SVM with the Rbf kernel, is a good choice for classifying maize leaf diseases using DWT feature extraction.

Table 7: Performance of the proposed system For the GABOR feature extraction technique

	Precision	Recall	F-measure	Accuracy
SVM_ Rbf	0.1757	0.4192	0.2476	0.4192
SVM_Poly	0.1757	0.4192	0.2476	0.4192
SVM_sigmoid	0.1757	0.4192	0.2476	0.4192
KNN (k=3)	0.0328	0.1812	0.0556	0.18122
KNN (k=5)	0.1757	0.4192	0.2476	0.4192
KNN (k=7)	0.1757	0.4192	0.2476	0.4192
DT	0.17574	0.4192	0.2476	0.4192
GB	0.1757	0.4192	0.2476	0.4192

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023

Table 7 presents an analysis of the performance of various machine learning classifiers applied to the classification of healthy and diseased maize leaves using the Gabor feature extraction technique. Surprisingly, all classifiers, including SVM with Rbf, Poly, and Sigmoid kernel, K-Nearest Neighbors (KNN) with different values of k (3, 5, and 7), Decision Trees (DT), and Gradient Boosting (GB), exhibit very low Precision, Recall, F-measure, and Accuracy values. These consistently low scores across all classifiers suggest that the Gabor feature extraction technique may not be suitable for this particular maize leaf disease classification task. It indicates that other feature extraction techniques or model configurations may need to be explored to improve classification accuracy for this specific problem. This analysis highlights the importance of selecting the right feature extraction method in the context of machine learning tasks, as different techniques may yield vastly different results.

Table 8 and Fig. 8 compare the proposed system with the current system.

Table 8: Performance of the proposed system with state of the art methods for maize plant leaf disease detection

Classifier	Accuracy (%)
SVM with GLCM 0^o (Proposed)	70.16
KNN with GLCM 45^o (Proposed)	71.91
DT with GLCM 90^o (Proposed)	91.27
GB with GLCM 135^o (Proposed)	95.05
SVM [15]	77.56
NB [15]	77.46
KNN [15]	76.16
DT [15]	74.35
RF [15]	79.23

Table VIII and Fig. 8 offers a comparative analysis of the proposed system's performance, which utilizes different classifiers with specific feature extraction techniques at varying orientations, against state-of-the-art methods for maize plant leaf disease detection, as measured by Accuracy (%). The classifiers in the proposed system include SVM with GLCM at 0 degrees, KNN with GLCM at 45 degrees, DT with GLCM at 90 degrees, and GB with GLCM at 135 degrees. Notably, GB with GLCM at 135 degrees achieves the highest accuracy of 95.05%, surpassing all other proposed system configurations.

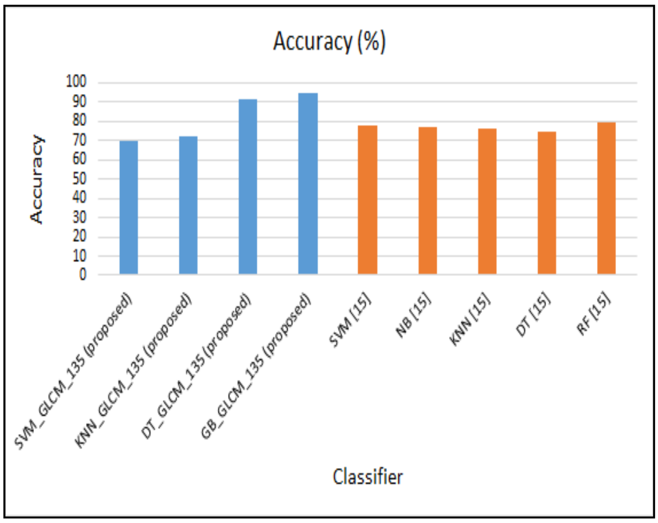


Figure 8: Analysis of the proposed system in comparison to existing techniques

When compared to existing method proposed by Himanshu Das et al. [15], which include SVM, Naive Bayes (NB), K-Nearest Neighbors (KNN), Decision Trees (DT), and Random Forest (RF), the proposed system demonstrates competitive performance. While some state-of-the-art methods have slightly higher accuracies, it is essential to consider the trade-offs between different classifiers and feature extraction techniques. This analysis underscores the effectiveness of the proposed system in maize leaf disease detection and its competitiveness with existing approaches while offering insights into the most promising configurations for accurate classification.

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023

5. Conclusion

In this paper, we have presented a novel approach for Maize Plant Leaf Disease Classification using Supervised Machine Learning Algorithms. Our research aimed to address the critical issue of timely disease detection in maize plants, a factor that is pivotal for global food security. Leveraging the power of texture analysis through Gray-Level Co-occurrence Matrix (GLCM) and Gabor feature extraction techniques, we achieved promising results in accurately classifying maize leaves into healthy and diseased categories. The comprehensive experiments involved the evaluation of four supervised machine learning algorithms: Decision Trees, Gradient Boosting, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN). Through these experiments, we observed that our proposed approach holds significant potential in enhancing plant disease detection. The decision tree and gradient boosting algorithms, in particular, demonstrated remarkable accuracy rates of 95.05% for GLCM features with 135^o. This underscores the robustness of our feature extraction techniques and the effectiveness of supervised machine learning in addressing this critical agricultural challenge.

Future studies and improvements in the field of plant disease classification will be built on the findings described in this publication. First, the addition of advanced deep learning techniques, such as convolutional neural networks (CNNs), could greatly improve classification accuracy by allowing the model to learn hierarchical features directly from raw visual data. Additionally, expanding the dataset to include a wider variety of maize leaf diseases and different stages of disease progression can make the system more robust and applicable to diverse agricultural scenarios. Furthermore, the development of a real-time monitoring system that can be deployed in the field using edge computing technologies holds the potential to provide continuous surveillance, alerting farmers to emerging disease threats. Lastly, efforts to ensure the scalability and accessibility of this technology to resource-constrained agricultural regions should be a priority, bridging the gap between research and practical implementation for the benefit of global food security and sustainable agriculture.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] Singh, Vijai & Sharma, Namita & Singh, Shikha. (2020). A review of imaging techniques for plant disease detection. Artificial Intelligence in Agriculture. 4. 229-242. 10.1016/j.aiia.2020.10.002.

[2] Gavhale, Ms & Gawande, Ujwalla. (2014). An Overview of the Research on Plant Leaves Disease Detection Using Image Processing Techniques. IOSR Journal of Computer Engineering. 16. 10-16. 10.9790/0661-16151016.

[3] Islam, Rashedul & Rafiqul, Md. (2015). An Image Processing Technique to Calculate Percentage of Disease Affected Pixels of Paddy Leaf. International Journal of Computer Applications. 123. 28-34. 10.5120/ijca2015905495.

[4] Horticultural Statistics at a Glance 2018, Department of Agriculture, Cooperation & Farmers' Welfare Ministry of Agriculture & Farmers' Welfare Government of India. https://agricoop.nic.in/sites/default/files/Horticulture%20Statistics%20at%20a%20Glance-2018.pdf

[5] Savita N. Ghaiwat, Parul Arora, "Detection and Classification of Plant Leaf Diseases Using Image Processing Techniques: A Review," International Journal of Recent Advances in Engineering & Technology, ISSN (Online): 2347 - 2812, Volume-2, Issue - 3, 2014.

[6] Prof. Sanjay B. Dhaygude, Mr.Nitin P. Kumbhar "Agricultural plant Leaf Disease Detection Using Image Processing" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 1, January 2013.

[7] Mrunalini R. Badnakhe and Prashant R. Deshmukh, "An Application of K-Means Clustering and Artificial Intelligence in Pattern Recognition for Crop Diseases," International Conference on Advancements in Information Technology 2011 IPCSIT vol.20 (2011).

[8] S. Arivazhagan, R. Newlin Shebiah, S. Ananthi, S. Vishnu Varthini "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features" Agric Eng Int: CIGR Journal, 15(1): 211¯217 2013.

[9] S. J. Suji Prasad, M. Thangatamilan, M. Suresh, H. Panchal, C. A. Rajan, C. Sagana, et al., "An efficient Lora-based Smart Agriculture Management and Monitoring System using wireless sensor networks", International Journal of Ambient Energy, vol. 43, no. 1, pp. 5447-5450, 2021.

[10] Anand. H. Kulkarni, Ashwin Patil R. K." Applying image processing technique to detect plant diseases," International Journal of Modern Engineering Research, Vol.2, Issue.5, Sep-Oct. 2012 pp-3661-3664.

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023

[11] Sabah Bashir, Navdeep Sharma "Remote Area Plant Disease Detection Using Image Processing," IOSR Journal of Electronics and Communication Engineering, ISSN: 2278-2834 Volume 2, Issue 6 2012, PP 31-34.

[12] Smita Naikwadi, Niket Amoda" ADVANCES IN IMAGE PROCESSING FOR DETECTION OF PLANT DISEASES," International Journal of Application or Innovation in Engineering & Management, Volume 2, Issue 11, November 2013.

[13] S, M., Sharma, A., Singh, S.P. et al. SVM-based compliance discrepancies detection using remote sensing for organic farms. Arab J Geosci 14, 1334 (2021). https://doi.org/10.1007/s12517-021-07700-4

[14] Sanjay B. Patil et al. "LEAF DISEASE SEVERITY MEASUREMENT USING IMAGE PROCESSING," International Journal of Engineering and Technology Vol.3 (5), 2011, 297-301.

[15] Piyush Chaudhary, "Color Transform Based Approach for Disease Spot Detection on Plant Leaf," International Journal of Computer Science and Telecommunications, Volume 3, Issue 6, June 2012.

[16] S. Samanta, A. Sarkar, C. Gupta and A. Sharma, "Machine learning integrated blockchain model for industry 4.0 smart applications", Knowledge engineering for modern information systems, 2021.

[17] Arti N. Rathod, Bhavesh Tanawal, Vatsal Shah" Image Processing Techniques for Detection of Leaf Disease," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 11, November 2012.

[18] M. C. R, S. Sharma, A. Sharma, M. Sunil Kumar, S. Kelkar and S. Vishal Deshmukh, "Cloud Top Management Role in Reducing Mobile Broadband Transmission Hazards and Offering Safety," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 1064-1068, doi: 10.1109/ICACITE57410.2023.10182893.

[19] Panigrahi, Kshyanaprava & Das, Himansu & Sahoo, Abhaya & Moharana, Suresh. (2020). Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms. 10.1007/978-981-15-2414-1_66.

[20] PlantVillage Dataset, https://plantvillage.psu.edu/posts/6948-plantvillage-dataset-download.

[21] Scanlan, Neil W. "Comparative performance analysis of texture characterization models in DIRSIG." (2003).

[22] K. Panigrahi, H. Das, A. Sahoo, and S. Moharana, "Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms," in Proceedings of the International Conference on Intelligent Computing and Applications, Singapore: Springer, 2020, pp. 659–669.

[23] Zhang, Xin, Jintian Cui, Weisheng Wang, and Chao Lin. 2017. "A Study for Texture Feature Extraction of High-Resolution Satellite Images Based on a Direction Measure and Gray Level Co-Occurrence Matrix Fusion Algorithm" Sensors 17, no. 7: 1474.

[24] Zubair, Abdul Rasak & Alo, Seun. (2019). Grey Level Co-occurrence Matrix (GLCM) Based Second Order Statistics for Image Texture Analysis. 93.

[25] Peng Nie, Michele Roccotelli, Maria Pia Fanti, Zhengfeng Ming, Zhiwu Li, "Prediction of home energy consumption based on gradient boosting regression tree," Energy Reports, Vol 7, 2021, pp. 1246-1255

[26] Jafarzadeh, Hamid, Masoud Mahdianpari, Eric Gill, Fariba Mohammadimanesh, and Saeid Homayouni. 2021. "Bagging and Boosting Ensemble Classifiers for Classification of Multispectral, Hyperspectral, and PolSAR Data: A Comparative Evaluation" Remote Sensing 13, no. 21: 4405. https://doi.org/10.3390/rs13214405

Doi: https://doi.org/10.54216/FPA.130201
Received: April 02, 2023 Revised: July 01, 2023 Accepted: September 04, 2023